Carleton University
Google
Microsoft
usenix Association
SIGCHI

SOUPS 2015

Proceedings of the Eleventh Symposium On Usable Privacy and Security

Ottawa, Canada

July 22-24, 2015
http://cups.cs.cmu.edu/soups/

Carnegie Mellon CyLab Usable Privacy and Security Laboratory
Master of Science in Information Technology privacy ENGINEERING
Carnegie Mellon University
Carnegie Mellon University CyLab
NSF

Foreword

The Eleventh Symposium On Usable Privacy and Security featured 22 technical papers, three workshops, a tutorial, 30 posters, a panel, 4 lightning talks, 2 demos, and an invited talk. We thank Carleton University for hosting SOUPS 2015.

This year we received 93 technical paper submissions, the most ever. The program committee provided two rounds of reviews. In the first round papers received at least three reviews. Authors had an opportunity to respond to the reviews their papers received in the first round. In the second round, some papers received additional reviews; in the end, papers received as many as six reviews. After a week of online discussion, the program committee held an in-person one-day meeting, which resulted in 22 papers selected for presentation and publication.

SOUPS 2015 featured an invited talk by Valerie Steeves, an Associate Professor in the Department of Criminology at the University of Ottawa in Ottawa, Canada. She is the lead researcher on MediaSmart's Young Canadian in a Wired World project (YCWW), which has been tracking young people's use of new media since 1999. Her talk was titled, "Online Privacy For Kids: What Works, What Doesn't."

On Thursday evening SOUPS 2015 enjoyed a trip to the Canadian Museum of History, including a banquet dinner at the museum. The closing session on Friday featured a panel titled: "Influenced or Ill-Advised? Ethical Considerations for Persuasive Technology in Usable Security" After a lively discussion, we concluded with the traditional SOUPS ice cream social.

We would like to thank all of the authors and the members of the technical papers committee and organizing committee for helping to produce this program. We are grateful to everyone whose assistance with logistical arrangements made this event possible, especially the faculty, staff, and students from Carleton University and the volunteers from Carnegie Mellon University. We would like to thank the US National Science Foundation, Microsoft, Google, and CyLab for their sponsorship of this event, and USENIX for publishing our proceedings. SOUPS 2015 was held in cooperation with USENIX and ACM SIGCHI.

Lorrie Faith Cranor	Robert Biddle	Sunny Consolvo
General Chair	**Technical Papers Co-Chair**	**Technical Papers Co-Chair**
Carnegie Mellon University	Carleton University	Google

ISBN 978-1-931971-249

SOUPS 2015 Organization

General Chair:	Lorrie Faith Cranor, Carnegie Mellon University, USA
Invited Talks Chair:	Franzi Roesner, University of Washington, USA
Lightning Talks and Demos Chair:	Elizabeth Stobert, Carleton University, Canada
Local Activities Co-Chair:	Sonia Chiasson, Carelton University, Canada
Panels Chair:	Matthew Smith, University of Bonn, Germany
Posters Co-Chairs:	Florian Schaub, Carnegie Mellon University, USA
	Yang Wang, Syracuse University, USA
Technical Papers Co-Chairs:	Robert Biddle, Carleton University, Canada
	Sunny Consolvo, Google, USA
Tutorials and Workshops Chair:	Mike Just, Heriot-Watt University, Scotland
Technical Papers Committee:	Robert Biddle, Carleton University (Co-Chair)
	Sunny Consolvo, Google, USA (Co-Chair)
	Lujo Bauer, Carnegie Mellon University, USA
	Richard Beckwith, Intel, USA
	Konstantin Beznosov, University of British Columbia, Canada
	Joseph Bonneau, Stanford University and EFF, USA
	Sonia Chiasson, Carelton, University, Canada
	Paul Dunphy, Newcastle University, UK
	Serge Egelman, University of California, Berkeley, USA
	Will Enck, North Carolina State University, USA
	Alain Forget, Carnegie Mellon University, USA
	Carrie Gates, Dell Research, USA
	Simson Garfinkel, National Institute of Standards and Technology, USA
	Cormac Herley, Microsoft Research, USA
	Iulia Ion, Google, USA
	Maritza Johnson, Google, USA
	Janne Lindqvist, Rutgers University, USA
	Andrew Patrick, Carelton University, Canada
	Adrienne Porter Felt, Google, USA
	Emilee Rader, Michigan State University, USA
	Rob Reeder, Google, USA
	Michael Reiter, University of North Carolina at Chapel Hill, USA
	Matthew Smith, University of Bonn, Germany
	Frank Stajano, University of Cambridge, UK
	Janice Tsai, Microsoft Research, USA
	Kami Vaniea, Indiana University, USA
	David Wagner, University of California Berkeley, USA
	Rick Wash, Michigan State University, USA
Publicity Chair:	Patrick Gage Kelley, University of New Mexico, USA

SOUPS 2015 Awards

Distinguished Paper Award
A Human Capital Model for Mitigating Security Analyst Burnout
Sathya Chandran Sundaramurthy, Alexandru G. Bardas, Jacob Case, Xinming Ou, and Michael Wesch (Kansas State University), John McHugh (RedJack LLC), and Siva Raj Rajagopalan (Honeywell ACS)

IAPP Privacy Paper Award
"My Data Just Goes Everywhere:" User Mental Models of the Internet and Implications for Privacy and Security
Ruogu Kang (HCII, CMU), Laura Dabbish (HCII & Heinz, CMU), Nathaniel Fruchter (Heinz, CMU), and Sara Kiesler (HCII, CMU)

Distinguished poster Awards

Authentication melee: A usability analysis of seven web authentication systems.
Scott Ruoti, Brent Roberts, Kent Seamons (Brigham Young University)

You Can Do Better — Motivational Statements in Password-Meter Feedback.
David Eargle (University of Pittsburgh), John Godfrey, Hsin Miao, Scott Stevenson, Rich Shay, Blase Ur, Lorrie Cranor (Carnegie Mellon University)

Your Location has been Shared 5,398 Times! A Field Study on Mobile App Privacy Nudging.
Hazim Almuhimedi, Florian Schaub, Norman Sadeh (Carnegie Mellon University), Idris Adjerid (University of Notre Dame), Alessandro Acquisti, Joshua Gluck, Lorrie Cranor, Yuvraj Agarwal (Carnegie Mellon University)

Table of Contents

Papers

Privacy Attitudes and Comprehension

Design and Compliance

Authentication Experience

continues on next page ➔

Authentication Methods

Mobile Privacy and Security

Security Experience

SOUPS 2015 Program

Wednesday, July 22

8 - 9 am: **registration, tea/coffee**

9 am – 10:30 am: **morning workshop sessions part 1**

- Examining Cybercrime through multiple lenses (Tutorial)
- Workshop on Inclusive Privacy and Security (WIPS): Privacy and Security for Everyone, Anytime, Anywhere
- Workshop on Usable Security and Privacy Education
- 2nd Annual Privacy Personas and Segmentation (PPS) Workshop

10:30- 11 am: **break**

11 – 12:30 pm: **morning workshop sessions part 2**

- Examining Cybercrime through multiple lenses (Tutorial)
- Workshop on Inclusive Privacy and Security (WIPS): Privacy and Security for Everyone, Anytime, Anywhere
- Workshop on Usable Security and Privacy Education
- 2nd Annual Privacy Personas and Segmentation (PPS) Workshop

12:30 – 1:30 pm: **lunch**

1:30 – 3:00 pm: **afternoon workshop sessions part 1**

- Examining Cybercrime through multiple lenses (Tutorial)
- Workshop on Inclusive Privacy and Security (WIPS): Privacy and Security for Everyone, Anytime, Anywhere
- Workshop on Usable Security and Privacy Education
- 2nd Annual Privacy Personas and Segmentation (PPS) Workshop

3:00 – 3:30 pm: **break**

3:30 – 5:00 pm: **afternoon workshop sessions part 2**

- Examining Cybercrime through multiple lenses (Tutorial)
- Workshop on Inclusive Privacy and Security (WIPS): Privacy and Security for Everyone, Anytime, Anywhere
- Workshop on Usable Security and Privacy Education
- 2nd Annual Privacy Personas and Segmentation (PPS) Workshop

5:15 – 7 pm: **Poster Session with dinner reception**

Thursday, July 23 – John Karat Day

On Thursday, July 23, SOUPS will honor the memory of John Karat with a short tribute during our opening session. If you have a Hawaiian-style shirt, we ask that you wear it on July 23, as that is what John always wore when he attended SOUPS. John was one of the original SOUPS program committee members and a mentor to many in the SOUPS community. John retired as a Research Staff Member at the IBM TJ Watson Research Center in 2010. While at IBM, John conducted HCI research on a variety of topics including privacy, personalization, and information management. John was co-leader of the IBM Privacy Research Institute, established to advance the importance of privacy issues in IT globally. John passed away in June of pancreatic cancer.

8 - 9 am: **registration, tea/coffee**

9 am - 9:30 am: **Welcome and awards presentation-** Distinguished Poster Awards, Distinguished Paper Award, IAPP SOUPS Privacy Award

9:30-10:30 am: **Keynote Speaker, Valerie Steeves**- B.A., J.D., Ph.D. is an Associate Professor in the Department of Criminology at the University of Ottawa in Ottawa, Canada. She is the lead researcher on MediaSmart's Young Canadian in a Wired World project (YCWW), which has been tracking young people's use of new media since 1999. With Jane Bailey, she co-leads the eGirls Project, an examination of the performance of gender on social media. She is also a co-editor of *Transparent Lives: Surveillance in Canada*, a 2014 multi-disciplinary report that maps out seven main trends in emerging surveillance practices, and the author of a series of award-winning multi-media games designed to teach young people how to protect their human rights online. Professor Steeves received her J.D. from the University of Toronto and was called to the Bar of Ontario in 1984.

10:30 – 11:00 am: **break**

11am – 12:30 pm: **session one: PRIVACY ATTITUDES AND COMPREHENSION**

***Session Chair:* Andrew Patrick**

A Design Space for Effective Privacy Notices
Florian Schaub (Carnegie Mellon University), Rebecca Balebako (RAND Corporation), Adam L. Durity (Google), and Lorrie Faith Cranor (Carnegie Mellon University)

WTH..!?!" Experiences, reactions, and expectations related to online privacy panic situations
Julio Angulo (Karlstad University) and Martin Ortlieb (Google)

"My Data Just Goes Everywhere:" User Mental Models of the Internet and Implications for Privacy and Security
Ruogu Kang (HCII, CMU), Laura Dabbish (HCII & Heinz, CMU), Nathaniel Fruchter (Heinz, CMU), and Sara Kiesler (HCII, CMU)

User Perceptions of Sharing, Advertising, and Tracking
Farah Chanchary and Sonia Chiasson (School of Computer Science, Carleton University)

12:30-1:30 pm: **lunch**

1:30 pm – 3:00 pm: **session two: DESIGN AND COMPLIANCE**

***Session Chair:* *S*imson Garfinkel**

Leading Johnny to Water: Designing for Usability and Trust
Erinn Atwater, Cecylia Bocovich, Urs Hengartner, Ed Lank, and Ian Goldberg (University of Waterloo)

Usability of Augmented Reality for Revealing Secret Messages to Users but Not Their Devices
Sarah J Andrabi, Michael K Reiter, and Cynthia Sturton (University of North Carolina, Chapel Hill)

Unpacking security policy compliance: Exploring motivators and barriers of employees' security behaviors
John M Blythe, Lynne Coventry, and Linda Little (PaCT Lab, Northumbria University)

LIGHTNING TALKS AND DEMOS

3-3:30 pm: **break**

3:30 pm – 5:00 pm: **session three: AUTHENTICATION EXPERIENCE**

***Session Chair:* Serge Egelman**

"I Added '!' At The End To Make It Secure": Observing Password Creation in the Lab
Blase Ur, Fumiko Noma, Jonathan Bees, Sean M. Segreti, Richard Shay, Lujo Bauer, Nicolas Christin, and Lorrie Faith Cranor (Carnegie Mellon University)

Social Media as a Resource of Security Experiences: A Qualitative Analysis of #Password Tweets
Paul Dunphy, Vasilis Vlachokyriakos, and Anja Thieme (Newcastle University), James Nicholson (Northumbria University), John McCarthy (University College Cork), and Patrick Olivier (Newcastle University)

"I'm Stuck!": A Contextual Inquiry of People with Visual Impairment in Authentication
Bryan Dosono, Jordan Hayes, and Yang Wang (Syracuse University)

LIGHTNING TALKS AND DEMOS

5:30 – 8 pm: **Dinner**
The SOUPS dinner will be held at the Canadian Museum of History **historymuseum.ca** Guests will have an opportunity to tour the museum before dinner. SOUPS will provide busses from Carleton to the museum and from the museum back to Carleton and to Albert@Bay Suite Hotel.

Friday, July 24

8 - 9 am: **tea/coffee**

9 - 10:30 am: **session four: AUTHENTICATION METHODS**

Session Chair: **Alain Forget**

Where Have You Been? Using Location-Based Security Questions for Fallback Authentication
Alina Hang (Media Informatics Group, University of Munich (LMU)), Alexander De Luca (Google), Michael Richter (Media Informatics Group, University of Munich (LMU)), Matthew Smith (Usable Security and Privacy Lab, University of Bonn), and Heinrich Hussmann (Media Informatics Group, University of Munich (LMU))

The Impact of Cues and User Interaction on the Memorability of System-Assigned Recognition-Based Graphical Passwords
Mahdi Nasrullah Al-Ameen, Kanis Fatema, Matthew Wright, and Shannon Scielzo (The University of Texas at Arlington)

On the Memorability of System-generated PINs: Can Chunking Help?
Jun Ho Huh (Honeywell ACS Labs), Hyoungshick Kim (Sungkyunkwan University), Rakesh B. Bobba (Oregon State University), Masooda N. Bashir (University of Illinois, Urbana-Champaign), and Konstantin Beznosov (University of British Columbia)

Evaluating the Effectiveness of Using Hints for Autobiographical Authentication: A Real Life Study
Yusuf Albayram and Mohammad Maifi Hasan Khan (Department of Computer Science and Engineering University of Connecticut)

10:30 am - 11 am: **break**

11 am - 12:30 pm: **session five: MOBILE PRIVACY AND SECURITY**

Session Chair: **Sonia Chiasson**

Usability and Security Perceptions of Implicit Authentication: Convenient, Secure, Sometimes Annoying
Hassan Khan, Urs Hengartner, and Daniel Vogel (University of Waterloo)

Understanding the Inconsistencies between Text Descriptions and the Use of Privacy-sensitive Resources of Mobile Apps
Takuya Watanabe (Waseda University), Mitsuaki Akiyama (NTT), and Tetsuya Sakai, Hironori Washizaki, and Tatsuya Mori (Waseda University)

On the Impact of Touch ID on iPhone Passcodes
Ivan Cherapau, Ildar Muslukhov, Nalin Asanka, and Konstantin Beznosov (The University of British Columbia)

Learning Random Secrets for Unlocking Mobile Devices
Stuart Schechter (Microsoft Research) and Joseph Bonneau (Stanford University & EFF)

12:30 – 1:30 pm: **lunch**

1:30 – 3 pm: **SESSION SIX: SECURITY EXPERIENCE**

Session Chair: **Matthew Smith**

Too Much Knowledge? Security Beliefs and Protective Behaviors Among US Internet Users
Rick Wash and Emilee Rader (Michigan State University)

Security Practices for Households Bank Customers in the Kingdom of Saudi Arabia
Deena Alghamdi, Ivan Flechais, and Marina Jirotka (Oxford University)

"...no one can hack my mind": Comparing Expert and Non-Expert Security Practices
Iulia Ion, Rob Reeder, and Sunny Consolvo (Google)

A Human Capital Model for Mitigating Security Analyst Burnout
Sathya Chandran Sundaramurthy, Alexandru G. Bardas, Jacob Case, Xinming Ou, and Michael Wesch (Kansas State University), John McHugh (RedJack LLC), and Siva Raj Rajagopalan (Honeywell ACS)

3-3:15 pm – **break**

3:15-4:30 pm – **panel**

This panel will discuss the ethical aspects of designing and deploying persuasive technology in the area of usable security. In some cases it is clear what the obvious desired behaviours are and as such designing persuasive systems is both desirable and achievable. However, in many cases it is not always clear – or unanimously agreed amongst the community – what the target behaviour should be. Especially in these cases it is questionable what the role of persuasive technology is and whether it should be deployed at all in these situations.

Moderator – James Nicolson (Northumbria University)
Robert Biddle (Carleton University)
Lynne Coventry (Northumbria Univeristy)
Serge Egelman (UC Berkeley)
Stuart Schechter (Microsoft Research)
Rebecca Balebako (RAND Corporation)

4:30 pm – **SOUPS social**

Examining Cybercrime through Multiple Lenses

Scope and Focus

"Examining Cybercrime Through Multiple Lenses" is a one-day multi-disciplinary tutorial. While SOUPS discourse typically focuses on security and privacy, we rarely examine the topic through the lens of the cybercrime. The tutorial will be divided into several sessions, each led by an industry, government, and academic expert in their respective area to provide insight and share first-hand experiences.

9:00-10:30am **Welcome and Introduction and session I: Industry Perspectives**

Targeted Crimeware in the Midst of Indiscriminate Activity

Nart Villeneuve is a Principal Threat Intelligence Analyst at FireEye Inc. where he focuses on cyber-espionage and the criminal underground. Nart's research, including prior work at Trend Micro and the University of Toronto, led to the discovery and documentation of multiple malware-based espionage campaigns, in-depth reports on cybercrime networks and technical analysis of Internet censorship and surveillance regimes.

Real World Cyber Security, How Simplified Models are Making it Worse

Pascal Fortin, MBA, CISA, CRISC, CRMA, is a private sector executive in charge of leading cyber security teams for the last 15 years. His firm, GoSecure, performs technical validation (ethical hacking, social engineering), cyber security architecture and operational support as well as managed services. His teams intervene with over 300 clients in Canada yearly and are responsible for detection, handling and remediating to a wide range of security incident and breaches.

10:30-11:00am **break**

11:00-12:30 **session II: Law Enforcement Perspectives**

Challenges to Policing

Bernie Murphy, Director, Ontario Provincial Police (OPP) Behavioural, Forensic and Electronic Services have been with the OPP for 27 years. He has since worked in a number of specialized investigative roles including physical surveillance, forgery, major frauds and homicide. He has worked in management roles as a Major Case Manager in Criminal Investigation Branch, Director of Anti-Rackets Branch and as Director of Strategic Management at the Alcohol and Gaming Commission of Ontario.

Challenges to Forensics Investigations

Staff Sergeant Vern Crowley, Team Leader of OPP Technological Crimes Unit has been doing digital forensics since 1998 and has been the team leader of TCU for the past year. He is responsible for the supervision of 35 civilians and officers who carry out all manner of digital forensic examinations. His Unit provides expert technical investigations in support of major case investigations (homicide etc), child sexual exploitation investigations and a wide variety of other police matters. His officers have specialized skills in the areas of mobile devices, cloud computing environs, encryption and other complex areas.

12:30-1:30 **Lunch**

1:30-3:00 **session III Legal and Social Perspectives**

From Ottawa to Budapest: How Canadian Legislation and International Treaties Address Security-Related Cybercrime

Nicolas Vermeys CISSP is a professor at the Université de Montréal's Faculty of Law, a researcher at the Centre de recherche en droit public (CRDP), and the associate director of the Cyberjustice Laboratory. He also serves as a legal advisor for the law firm of Legault Joly Thiffault, and serves on the board of the Société québécoise d'information juridique (SOQUIJ), Éducaloi, and the Canadian Center for Court Technology (CCCT).

The Sociology of Hackers and Modes of Regulating Cybercrime

Dr. Benoît Dupont holds the Canada Research Chair for Security, Identity, and Technology, is a Professor at the School of Criminology at the Université de Montréal. He is also the Scientific Director of SERENE-RISC, a Canadian network for knowledge mobilization of cybersecurity. His main research interests include the coevolution of crime and technology, the social organization of hackers and other on-line offenders, and the governance and regulation of online risks.

3:00-3:30 **Break**

3:30-5:30 **Interactive Panel Discussion with Presenters and Closing Remarks**

Organizers:

This tutorial is organized by SERENE-RISC. SERENE-RISC is a Canadian Networks of Centers of Excellence for Knowledge Mobilization on cybersercurity. The multidisciplinary network brings together academics in the fields of computer and social sciences, as well as public and private partners representing key Canadian stakeholders. The aim is to mobilize growing knowledge about online risks, to promote the most effective strategies, and to minimize the consequences of cyber attacks.

Sonia Chiasson, Carleton University
Benoît Dupont, Université de Montréal

Workshop on Inclusive Privacy and Security (WIPS)

("Privacy and Security for Everyone, Anytime, Anywhere")

Scope and Focus

Many privacy and security solutions are designed for and evaluated with a narrow range of users (e.g., technology literate, physically capable, young), and the solutions make assumptions about the environment and the user interaction capabilities (e.g., keyboard, mouse, touch screen, audio, camera). However, these solutions (e.g., authentication, CAPTCHAs, anti-phishing tools) are used by a wide variety of people, and in varied situations. While there are accessible and environment-aware solutions, they are often targeted at specific disability conditions (e.g., vision impairment) or situational deficiency (e.g., text entry on a moving train). In general, marginalized groups and situational impairments are under-represented when designing privacy and security solutions. To make these solutions more inclusive, we need to take into consideration the various disability conditions and situational impairments.

We observe that the effects of situational impairments, when people are unable to perform a task due to environmental or other factors related to the current situation, overlap with groups who have generally been marginalized. Privacy and security techniques that benefit one group may very well assist the other group. The marginalized groups of people and those with situational impairments struggle with privacy and security technology. The result is a greater impact on their ability to focus on their primary tasks (web-based email, online banking, driving, surgery, etc.).

The goals of this workshop are as follows:

1. To learn about the experiences and requirements of marginalized groups (e.g., people with various disability conditions, young people, elderly people, technology neophytes) and situational impairments (e.g., dark rooms, noisy locations, in motion or in vibrating environments, in stressful situations, when performing a task involving cognitive load).
2. To share studies of privacy and security solutions to assist these groups.
3. To explore and develop potential directions towards designing solutions for everyone, anytime and anywhere.

9- 9:05 am **Introduction**

9:05-9:50 am **Invited Talk:**
Overview of Accessibility and Current Trends in Accessibility Research
Karyn Moffatt (McGill University, Canada)

9:50-10:30 am **Presentations-Applications**

Accessible Banking: Experiences and Future Directions
Bela Gor (Business Disability Forum, UK) and David Aspinall (University of Edinburgh, UK)

Scoping Secure Online Shopping for Older People
Maria Klara Wolters and David Aspinall (University of Edinburgh, UK)

Assessing online privacy safeguards among marginalized Internet users in public libraries
Bryan Dosono (Syracuse University, USA)

Improving Children's Mobile Privacy Awareness and Behaviour
Leah Zhang-Kennedy and Sonia Chiasson (Carleton University, Canada)

10:30-11:00 am **break**

11:00 am- 12:30 pm **Group work**

12:30-1:30 pm **lunch**

1:30-1:50 pm **Presentations - Methods**

Privacy for Everyone: Towards an Inclusive Design Approach for Accessible Privacy and Security Technology
Katharina Krombholz (SBA Research, Austria), Christopher Frauenberger (Vienna University of Technology, Austria) and Edgar Weippl (SBA Research, Austria)

Don't Forget About Us: Lessons Learned from My Accessible Security Research Experience
Jordan Hayes (Carnegie Mellon University, USA)

1:50 – 2:30 **Panel Discussion**

2:30 – 3:00 pm **Presentations- Solution Ideas**

A Tap on the Wrist - Security Usability of Wearables
Ann-Marie Horcher (Nova Southeastern University, USA)

Towards Universal Authentication: Ability-Based Design, Crowdsourcing, and Privacy-Preserving Biometrics
Huichuan Xia (Syracuse University, USA)

3:00 – 3:30 pm **break**

3:30 – 4:00 pm **Presentations- Solutions Ideas**

Accessible CAPTCHA for Everyone: Is it Possible?
Sajad Shirali-Shahreza (University of Toronto, Canada)

A Research Framework and Initial Study of Browser Security for the Visually Impaired
Elaine Lau and Zachary Peterson (Cal Poly San Luis Obispo, USA)

4:00 – 4:30 pm **Wrap up**

Organizers

Mike Just, Heriot-Watt University, UK
Yang Wang, Syracuse University, USA
Larry Koved, IBM Research, USA
Karyn Moffatt, McGill University, Canada

Workshop on Usable Security and Privacy Education

Scope and Focus

The past 15 years has seen a dramatic increase in attention to usable security and privacy research, yet the vast majority of computing students are being exposed to very little of this discipline. A variety of usable security courses are being taught, particularly by researchers within the SOUPS community. These courses are often electives taught to senior or graduate students, organized around the common and recent research themes in the field. One potential barrier to expanding the breadth and depth of usable security education across computing programs is the lack of a framework or body of knowledge defining what students could and should know about usable security and privacy. How do we translate the research themes and results in the field into educational topics? What are the knowledge units, skills, and learning objectives for general computing and/or security students? Such a framework could also provide topic consistency and guidance on how to integrate those knowledge units into existing and new courses; and on the types of learning materials that need to be developed based upon current and future research results.

This workshop aims to bring together educators in usable security and privacy who are interested in discussing these issues and contributing to the development of a body of knowledge. The goal of the workshop is to brainstorm and start to organize the topics, knowledge units, and skills as well as learning goals and objectives within usable security and privacy for a variety of computing students. The workshop will consist of lightning talks, discussions, and breakout sessions. The outcomes of these discussions will be documented in detail and made available on a workshop website. This will be iteratively refined by the organizers and the community throughout the following year.

Program

8:00 - 9:00 am – Coffee/tea

9:00 - 9:15 am – **Welcome and State of Usable Security and Privacy Education**

9:15 - 10:15 am – **Lightning Talks**

The Teaching Privacy Curriculum
Serge Egelman, Gerald Friedland, Julia Bernd, Dan Garcia, and Blanca Gordo; International Computer Science Institute and University of California at Berkeley

Courses for understanding the impact of privacy and security choices
Emily McReynolds; University of Washington

Developing a Standardized and Multidisciplinary Curriculum for Digital Forensics
Masooda Bashir and Roy Campbell; University of Illinois at Urbana-Champaign

Recommendations for a Graduate Seminar in Usable Security
Kent Seamons; Brigham Young University

Usable Security and Privacy in Technology Ethics Courses
Patrick Gage Kelley; University of New Mexico

Human Factors in Security and Privacy

Zinaida Benenson; Friedrich-Alexander-University Erlangen-Nuremberg, Germany

Ethnomethodology and Usable Security: The Value of Descriptive Research for Graduate Students
Hervé Saint-Louis; University of Toronto

An interdisciplinary study of phishing and spear-phishing attacks
Robin Gonzalez, Michael Locasto; University of Calgary

10:15 – 10:30 am – Introduction to Body of Knowledge (BoK) & Security and Privacy research topic outline

10:30 - 11:00 am – Break

11:00 - 12:30 pm – **Session 1: Topic Brainstorming and Organization of Bok**

12:30 - 1:30 pm – Lunch

1:30 - 3:00 pm – **Session 2: Learning Objectives for BoK**

3:00 - 3:30 pm – Break

3:30 - 4:00 pm – **Session 3: Group Discussion of Educational Strategies, Methods, Materials and Future of Usable Security and Privacy Education**

4:00 - 4:30 pm – Integration of Usable Security into CS/IT curriculum

4:30 – 5:00 pm – Open Discussion

5:15 - 7:00 pm – Poster Session and Dinner Reception

Workshop Organizers

Heather Richter Lipford, UNC Charlotte
Simson Garfinkel, NIST
Andrew Besmer, Winthrop University
Jason Watson, University of North Alabama

2nd Annual Privacy Personas and Segmentation (PPS) Workshop

The PPS workshop is an opportunity for researchers and practitioners to explore improved methods and tools for understanding privacy concerns, facilitating the construction of privacy personas and/or segmenting users on the basis of their diverse privacy attitudes, concerns, and behaviors. These characterizations may make it possible to better predict online behavior and disclosure, to personalize interfaces, to suggest appropriate default settings, and to measure shifts in public sentiment.

Scope and Focus

Scholars and practitioners have long been interested in understanding and measuring privacy attitudes and concerns, and their relationship with privacy behavior. Over time, an awareness has emerged of the importance of context in privacy concerns, and the complex relationship between concerns and actual behaviors. In recent years, proposals have been made to segment individuals into more granular and detailed categories of privacy concerns or behaviors, and classifying or predicting their privacy types, or personas. Such efforts have been motivated by the goals of better understanding the relationships between privacy attitudes, concerns and behaviors, and of helping end users make better privacy decisions.

The focus of the PPS workshop is to bring together researchers and practitioners interested in privacy personas and segmentation, to encourage a paradigm shift in the measurement, modeling, and characterization of privacy concerns which recognizes the complex interaction of factors influencing it. Those interested in participating should submit a research or position paper on a relevant topic. Topics of interest include, but are not limited to:

- The use of segmentations and personas for representing privacy concerns
- Critiques of existing approaches and explorations of the inherent limitations of privacy segmentations or privacy personas
- New paradigms and instruments for understanding, measuring, and modeling privacy concerns, including machine learning based approaches
- Evaluations and critiques of existing instruments for measuring privacy concerns
- The role of context, personality, experiences, and other traits in influencing privacy concerns
- The relationship between privacy concerns, attitudes, and behavior
- Algorithms and tools for providing personalized privacy recommendations
- Other topics related to measuring, modeling, and characterizing privacy concerns

8:00 - 9:00 am: Coffee/tea

9:00 - 10:30 am: **Session 1 – Multiple Personae** (Chair: Alessandro Acquisti)

Welcome Opening Keynote: A. Michael Froomkin
Privacy Personae – US Legal (and Political) Considerations

Implications of Device Sharing Behaviors for Predicting Privacy Preferences
Anna Turner, Tara Matthews, Kerwell Liao, Marianne Berkovich, and Sunny Consolvo

10:30 - 11:00 am: Break

11:00 - 12:30 pm: **Session 2 – Apps 'n Ads: Awareness, Perceptions, and Behaviors** (Chair: Allison Woodruff)

Privacy and Behavioral Advertising: Towards Meeting Users' Preferences
Pedro Giovanni Leon, Ashwini Rao, Florian Schaub, Abigail Marsh, Lorrie Faith Cranor and Norman Sadeh

Perceived Frequency of Advertising Practices
Sai Teja Peddinti, Allen Collins, Aaron Sedley, Nina Taft, Anna Turner and Allison Woodruff.

Location-Based Applications – Benefits, Risks, and Concerns as Usage Predictors
Maija Poikela, Ina Wechsung and Sebastian Möller.

12:30 - 1:30 pm: Lunch

1:30 - 3:00 pm: **Session 3 – Social and Cultural Antecedents of Privacy** (Chair: Norman Sadeh)

Multiple Facets of Information Privacy: A Socio-Cultural Approach
Hsiao-Ying Huang and Masooda Bashir.

Cross-Cultural Privacy Prediction
Yao Li, Bart P. Knijnenburg, Alfred Kobsa and M-H. Carolyn Nguyen

Balancing Privacy Concerns and Impression Management Strategies on Facebook
Jessica Vitak

3:00 - 3:30 pm: Break

3:30 - 4:00 pm: **Session 4 – Lightning and Learning** (Chair: Allison Woodruff)
Lightning Talks.

Closing Keynote: Norman Sadeh
Learning People's Privacy Preferences: Opportunities and Challenges.

Please note: Authors will be given a maximum of 18 minutes to present their papers, plus time for questions and discussion. Time limits will be rigorously enforced throughout the day by session chairs.

Organizers

Alessandro Acquisti, Carnegie Mellon University
Bart Knijnenburg, University of California, Irvine
Norman Sadeh, Carnegie Mellon University
Allison Woodruff, Google

Posters

Poster: What to do when your cover's been blown: Public perceptions of re-identification attacks
Ester Moher (Children's Hospital of Eastern Ontario), Khaled El Emam (University of Ottawa, Children's Hospital of Eastern Ontario)

Poster: H4Plock: Supporting Mobile User Authentication through Gestural Input and Tactile Output
Abdullah Ali (University of Maryland, Baltimore County), Ravi Kuber (University of Maryland, Baltimore County), Adam J. Aviv (United States Naval Academy)

Poster: Protecting Personal Health Information: The Roles of Context, Framing and Priming in Privacy-Related Choices
Vanessa Boothroyd (Privacy Analytics, Inc.), Ester Moher (University of Ottawa, Children's Hospital of Eastern Ontario), Khaled El Emam (Privacy Analytics, Inc., CHEO)

Poster: Alternative Keyboard Layouts for Improved Password Entry and Creation on Mobile Devices
Ethan Genco, Ryan Kelly, Cody Vernon, Adam J. Aviv (United States Naval Academy)

Poster: Do bigger grids sizes mean better passwords? 3x3 vs. 4x4 Grid Sizes for Android Unlock Patterns
Devon Budzitowski (United States Naval Academy), Adam J. Aviv (United States Naval Academy), Ravi Kuber (University of Maryland, Baltimore County)

Poster: Using Authorization Logic to Capture User Policies in Mobile Ecosystems
Joseph Hallett, David Aspinal (University of Edinburgh)

Poster: How I Learned To Be Secure: Advice Sources and Personality Factors in Cybersecurity
Elissa M. Redmiles, Amelia Malone, Michelle L. Mazurek (University of Maryland)

Poster: User-Generated Free-Form Gestures for Authentication: Security and Memorability
Michael Sherman, Gradeigh D. Clark, Yulong Yang, Shridatt Sugrim (Rutgers University), Arttu Modig (University of Helsinki), Janne Lindqvist (Rutgers University), Antti Oulasvirta (Max Planck Institute for Informatics and Saarland University), Teemu Roos (University of Helsinki)

Poster: Who is behind the Onion? Understanding Tor-Relay Operators
Hsiao-Ying Huang, Masooda Bashir (University of Illinois at Urbana-Champaign)

Poster: Using Signal Detection Theory to Measure Phishing Detection Ability and Behavior
Casey Canfield, Baruch Fischhoff, Alex Davis (Carnegie Mellon University)

Poster: Usability Problems with Password Creation Systems: Results from Expert and User Evaluation
Saja Althubaiti, Helen Petrie (University of York)

Poster: A Framework for Comparative Usability Studies on Secure Device Pairing
Achal Channarasappa, Pranita Ramakrishnan, Joshua Tan, Jeremy Thomas (Carnegie Mellon University)

Poster: Preliminary Investigation on Psychological Traits of Users Prone to be damaged by Cyber-attack
Takeaki Terada, Yoshinori Katayama, Satoru Torii, Hiroshi Tsuda (Fujitsu Limited)

Poster: Your Location has been Shared 5,398 Times! A Field Study on Mobile App Privacy Nudging
Hazim Almuhimedi, Florian Schaub, Norman Sadeh (Carnegie Mellon University), Idris Adjerid (University of Notre Dame), Alessandro Acquisti, Joshua Gluck, Lorrie Faith Cranor, Yuvraj Agarwal (Carnegie Mellon University)

Poster: Digital signature services for users - Improving user experience to support trust among work partners
Lorraine Tosi, Aurélien Bénel, Karine Lan (Université de Technologie de Troyes)

A Design Space for Effective Privacy Notices

Florian Schaub,[1] Rebecca Balebako,[2] * Adam L. Durity,[3] * Lorrie Faith Cranor[1]

[1]Carnegie Mellon University
Pittsburgh, PA, USA
{ fschaub, lorrie }@cmu.edu

[2]RAND Corporation
Pittsburgh, PA, USA
balebako@rand.org

[3]Google
Mountain View, CA, USA
adurity@google.com

ABSTRACT

Notifying users about a system's data practices is supposed to enable users to make informed privacy decisions. Yet, current notice and choice mechanisms, such as privacy policies, are often ineffective because they are neither usable nor useful, and are therefore ignored by users. Constrained interfaces on mobile devices, wearables, and smart home devices connected in an Internet of Things exacerbate the issue. Much research has studied usability issues of privacy notices and many proposals for more usable privacy notices exist. Yet, there is little guidance for designers and developers on the design aspects that can impact the effectiveness of privacy notices. In this paper, we make multiple contributions to remedy this issue. We survey the existing literature on privacy notices and identify challenges, requirements, and best practices for privacy notice design. Further, we map out the design space for privacy notices by identifying relevant dimensions. This provides a taxonomy and consistent terminology of notice approaches to foster understanding and reasoning about notice options available in the context of specific systems. Our systemization of knowledge and the developed design space can help designers, developers, and researchers identify notice and choice requirements and develop a comprehensive notice concept for their system that addresses the needs of different audiences and considers the system's limitations and opportunities for providing notice.

1. INTRODUCTION

The purpose of a privacy notice is to make a system's users or a company's customers aware of data practices involving personal information. Internal practices with regard to the collection, processing, retention, and sharing of personal information should be transparent to users. The privacy notice acts as a public announcement of those practices. Privacy notices can take different shapes and leverage different channels, ranging from a privacy policy document posted on a website, or linked to from mobile app stores or mobile apps, to signs posted in public places to inform about CCTV cameras in operation. Even an LED indicating that a camera or microphone is active and recording constitutes a privacy notice, albeit one with limited information about the data practices associated with the recording. Providing notice about data practices is an essential aspect of data protection frameworks and regulation around the world [57]. While transparency has been emphasized as an important practice for decades, existing privacy notices often fail to help users make informed choices. They can be lengthy or overly complex, discouraging users from reading them.

Smartphones and mobile apps introduce additional privacy issues as they support recording of sensor and behavioral information that enables inference of behavior patterns and profiling of users. Yet, comparatively smaller screens and other device restrictions constrain how users can be given notice about and control over data practices.

The increasing adoption of wearable devices, such as smart watches or fitness trackers, as well as smart home devices, such as smart thermostats, connected light bulbs, or smart meters, represents a trend towards smaller devices that are even more constrained in terms of interaction capabilities, but are also highly connected with each other and the cloud. While providing notice and choice is still considered essential in the "Internet of Things" (IoT) [48, 74], finding appropriate and usable notice and choice mechanisms can be challenging.

The challenges of providing usable privacy notice have been recognized by regulators and researchers. For instance, FTC chairwoman Edith Ramirez [107] stated in the IoT context: "In my mind, the question is not whether consumers should be given a say over unexpected uses of their data; rather, the question is how to provide simplified notice and choice." An extensive body of research has studied usability issues of privacy notices (e.g., [14, 33, 64, 51]) and proposed improved notice interfaces (e.g., [34, 66, 67]), as well as technical means to support them (e.g., [75, 127, 131]). Multi-stakeholder processes have been initiated in the wake of the White House's proposed Consumer Bill of Rights [122] to tackle transparency and control issues of mobile privacy [92] and facial recognition [93]. While such efforts have resulted in guidance for notices in the context of particular systems, they have given little consideration to usability [14].

Existing frameworks and processes for building privacy-friendly systems, such as Privacy by Design [36] or privacy impact assessments [136], focus on the analysis of a system's data practices and less so on the design of notices. Even the OECD report on "making privacy notices simple" [94] basi-

*Rebecca Balebako and Adam Durity performed this work while at Carnegie Mellon University.

Symposium on Usable Privacy and Security (SOUPS) 2015, July 22–24, 2015, Ottawa, Canada.

cally states that one should design a simplified notice, conduct usability tests, and deploy it – the crucial point of *how* to design a simplified notice is not addressed. Common proposals to improve the usability of privacy notices are the use of multi-layered notices [9, 26] or just-in-time notices [47].

Despite the previous work on privacy notices, transparency tools, and privacy mechanisms, a system designer or developer has very little guidance on how to arrive at a privacy notice design suitable and appropriate for their specific system and its respective characteristics. Existing best practices are spread throughout the literature and have not previously been organized into a comprehensive design framework. As a result, privacy notices are often hastily bolted on rather than well-integrated into a system's interaction design. Designers may not be aware of the many alternatives for designing usable privacy notices and therefore do not systematically consider the options. Furthermore, designers and researchers do not yet have a standard vocabulary for describing privacy notice options.

In this paper, we make multiple contributions to ease the design of privacy notices and their integration into a system. The goal is to help developers embed privacy notices and choice options into their system design where relevant, with minimal disruption to the system's interaction flow. First, we identify challenges, requirements, and best practices for the design of privacy notices. Based on a survey of existing literature and privacy notice examples, we develop a design space of privacy notices. This design space and its dimensions provide a systemization of knowledge and a taxonomy to foster understanding and reasoning about opportunities for privacy notices and controls. We demonstrate the utility of our design space by discussing existing privacy notice approaches in different domains.

2. BACKGROUND

The concept of privacy notices is founded on the idea that users of services and systems that collect or process personal information should be informed about what information is collected about them and for which purposes, with whom it is shared, how long it is stored, and their options for controlling or preventing certain data practices [45, 95]. Given such transparency, users should be able to make informed privacy and consent decisions.

2.1 Roles of Privacy Notices

Privacy notices serve different roles depending on a stakeholder's perspective. Consumers, companies, and regulators see privacy notices in different ways.

For *companies*, privacy notices serve multiple purposes, including demonstrating legal compliance and building customer trust. Privacy notices are often primarily a necessity to ensure compliance with legal and regulatory requirements, rather than a tool to create transparency for users. For instance, the European Data Protection directives have strict notice requirements [41, 43]. In the U.S., not providing notice could be interpreted as a deceptive trade practice by the FTC [45] or violate federal, state, or sector-specific privacy legislation, such as CalOPPA [96] or HIPAA [27].

Yet, there are also intrinsic reasons why businesses and system designers should aim to provide privacy notices that are meaningful to users. Being upfront about data practices – especially about those that may be unexpected or could be misinterpreted – provides the opportunity to explain their purpose and intentions in order to gain user acceptance and avoid backlash. Furthermore, companies that provide privacy-friendly and secure systems can leverage privacy notices to make users aware of privacy-friendly data practices. Implementing and highlighting good security and privacy practices can further create a competitive advantage as users may perceive the system as more trustworthy.

Regulators, such as data protection authorities or the FTC, rely on companies' privacy notices – primarily their privacy policies – as an important tool to investigate and enforce regulatory compliance [31]. If a company violates its privacy policy, it provides regulators with a basis to take action; for example, the FTC may treat a violation as an unfair or deceptive trade practice [45, 116]. Further, data protection authorities in Europe and other countries may assess whether the described practices meet more stringent criteria, such as use limitation, proportionality of data practices, and user access options [41, 43].

2.2 Hurdles to Effective Privacy Notices

While privacy notices fulfill many roles for different stakeholders, in practice most privacy notices are ineffective at informing consumers [33, 83]. This ineffectiveness stems from hurdles that can be attributed not only to general shortcomings of the notice and choice concept [25, 33, 116], but also to the challenges in designing effective privacy notices.

Notice complexity. The different roles of privacy notices result in a conflation of requirements. Besides informing users about data practices and their choices, privacy notices serve to demonstrate compliance with (self-)regulation and limit the system provider's liability [23]. As a result, privacy notices often take the shape of long privacy policies or terms of service that are necessarily complex because the respective laws, regulations, and business practices are complex [25]. For instance, website privacy policies are typically long, complex documents laden with legal jargon. Indeed it has been estimated that to read the privacy policies for all the websites an American Internet user visits annually would take about 244 hours per year [83]. Privacy policies also read like contracts because regulators aim to enforce them like contracts [25]. Notices may further be purposefully vague to avoid limiting potential future uses of collected data [116]. The effect is that these notices are difficult for most people to understand [83, 111].

Lack of choices. Many privacy notices inform about data practices but do not offer real choices. Using a website, an app, a wearable device, or a smart home appliance is interpreted as consent to the data practices – regardless of the user having seen or read them. Even if notices are seen by users, they largely describe a system's data practices, with few choices to opt-out of certain practices, such as sharing data for marketing purposes. Thus, users are effectively left with a take-it-or-leave-it choice – give up your privacy or go elsewhere [116]. Users almost always grant consent if it is required to receive the service they want [25]. In the extreme case, privacy notices are turned into mere warnings that do not empower individuals to make informed choices [25] (e.g., “Warning: CCTV in use” signs). Yet, privacy notices can only be effective if they are actionable and offer meaningful choices [33]. Awareness of data practices can enable users to make informed privacy decisions, but privacy controls are needed in order to realize them [113].

Notice fatigue. Notice complexity and the lack of choices mean that most privacy notices are largely meaningless to consumers [25]. Users may feel it is pointless to read them, and most users don't. A recent White House report [106] stated, "Only in some fantasy world do users actually read these notices and understand their implications before clicking to indicate their consent." Furthermore, businesses may change their data practices and notices at any time, which means any effort spent on understanding the notice may have been in vain [116]. Privacy notices and security warnings are often shown at inopportune times when they conflict with the user's primary task [63], therefore they are dismissed or accepted without scrutiny. Frequent exposure to seemingly irrelevant privacy notices results in habituation, i.e., notices are dismissed without even registering their content [6, 56]. Further, a notice's framing, distractions or time delays can reduce the notice's effectiveness [3].

Decoupled notices. Some systems decouple a privacy notice from the actual system or device, for example by providing it on a website or in a manual. Privacy notices are not only relevant for websites, mobile apps, or surveillance cameras, but for the whole gamut of systems and devices that process user information. Designing and providing appropriate notices for novel systems, such as smart home appliances or wearable devices, is challenging [48]. The straightforward approach is to decouple the privacy notice from the system. For example, many manufacturers of fitness tracking devices provide a privacy policy on their websites, while the actual device does not provide any privacy notices [103]. As a result, users are less likely to read the notice and may therefore be surprised when they realize that their mental models do not match the system's actual data practices [48].

These issues paint a somewhat dire picture of the state of privacy notices. However, just abandoning the concept of notice is not a viable option, as the transparency notices should provide is essential for users, businesses, and regulators alike [107]. We argue that many of these issues can be addressed by placing the emphasis on how privacy notices are designed. Instead of providing notice merely to fulfill legal and regulatory requirements, notices should effectively inform users about data practices and provide appropriate choices. Some proposed solutions point in that direction, such as multi-layered privacy notices [9], just-in-time notices [101], and notices focused on unexpected data practices [48, 107]. However, so far, there is little guidance on the actual design and integration of such notices into real-world systems. Next, we identify requirements and best practices for effective and usable privacy notice design.

3. REQUIREMENTS & BEST PRACTICES FOR PRIVACY NOTICE DESIGN

In order to make privacy notices effective and usable, they should not be tacked on after the system has been completed but instead be integrated into a system's design. Privacy notices and choice options can then be designed for specific audiences and their notice requirements, and take into account a system's opportunities and constraints.

In this section, we identify common requirements, necessary considerations, and best practices for privacy notice. These aspects are based on a survey of the usable privacy literature and an analysis of existing privacy design and assessment frameworks, such as Privacy by Design [36], privacy impact assessments [136], and proposals for layered notice design [9, 26, 94].

Together with the design space presented in the next section, the requirements and best practices discussed in this section provide guidelines and a toolbox for system designers and researchers that can aid them in the development of usable and more effective privacy notices for their systems.

3.1 Understand Privacy in the System

The first step in designing effective privacy notices is to understand a system's information flows and data practices in order to determine whether privacy notices are needed, who should be notified, and about what. Such an assessment can be conducted as part of a privacy impact assessment (PIA) [136], which further serves the purpose of identifying privacy risks associated with the system and making recommendations for privacy-friendly systems design. PIAs are becoming an essential – in some countries mandatory – aspect of systems design [134]. They serve the broader goal of ensuring a system's legal and regulatory compliance, as well as informing privacy by design and risk mitigation. A common approach in existing PIA frameworks [135, 136] is to first assess if the system collects or processes privacy-sensitive information to determine if a full PIA is required. The next step is to describe the system in detail, including its information flows and stakeholders. This description is the basis for analyzing the system's privacy implications and risks [36, 37]. A PIA produces a report detailing identified issues and recommendations on how to address them.

The resulting recommendations for privacy improvements may include changing collection practice, or identifying opportunities for data minimization. Data minimization reduces the risk of using data in ways that deviate from users' expectations as well as liability risks associated with data theft and unintended disclosure [48]. As an additional benefit, it also reduces the complexity of data practices that need to be communicated to users in privacy notices. If done early in a system's design process, this may also be an opportunity to consider and improve system constraints related to privacy. For example, recognizing that a video camera is collecting information, the device designers may decide to include a light or other signal indicating when the camera is on. The PIA report and data practices should be updated to reflect any privacy-friendly improvements. This process may involve multiple iterations.

Conducting a PIA informs notice design by helping to determine if notices are necessary in the first place, providing an overview of data practices for which notice should be given, potentially reducing the complexity of data practices, and determining the audiences that need to be considered in notice design. The outcome of a PIA is a deep understanding of a system's privacy characteristics, which can be codified in a comprehensive privacy policy.

A privacy policy describes a system's data practices including all relevant parameters, namely what data is being collected about users (and why), how this information is being used (and why), whether it is shared with third parties and for what purposes, how long information is retained, as well as available choice and access mechanisms [45]. This full privacy policy serves as the definitive (and legally binding) privacy notice. As such, it may be a long and complex document, which primarily serves the company to demonstrate transparency and regulatory compliance. It is therefore mainly relevant for businesses and regulators and less interesting or useful to users. However, a well-defined privacy policy can serve as the basis for designing concise, user-friendly privacy notices as it maps out the different data practices about which users may need to be informed.

3.2 Different Notices for Different Audiences

Privacy impact assessments and the creation of privacy policies are well-known and established concepts, but notice design often stops with the privacy policy. Whereas the full privacy policy may be sufficient for businesses and regulators, the key challenge is to design effective privacy notices for users. Therefore, one needs to understand which audiences have to be addressed by notices [23]. While determining a website's audience may be straightforward (typically the visitors of the website), mobile applications, wearables, smart cars, or smart home appliances expand the audiences and user groups that need to be considered. Such systems may have a primary user, but potentially also multiple users with different privacy preferences. For example, a home lock automation system may collect information about all family or household members, including guests [123]. Wearables, such as Google Glass, may incidentally collect information about bystanders. Social media and mobile applications enable users to share information with and about others, e.g., by tagging someone in a geo-referenced photo.

To determine the different audiences for privacy notices, the set of all data practices specified in the privacy policy needs to be analyzed to determine which data practices affect which audience. Typical audience groups are the *primary user* of a system; *secondary users*, such as household members, having potentially less control over the system; and *incidental users*, such as bystanders, who may not even be aware that information about them is collected by a system. Depending on the system, other or additional audience groups may need to be considered. There may also be regulatory requirements applying to specific audience groups, such as children [85], that have to be considered.

While some audience groups may be affected by the same data practices (e.g., data collection about the primary user and other household members by a smart home system), other groups may only be affected by very specific data practices (e.g., while all of a wearable's data practices affect the primary user, bystanders are only effected if they're incidentally recorded by the device, for instance, when the primary user takes a photo or video with a wearable device).

3.3 Relevant and Actionable Information

To be effective and draw the user's attention, privacy notices must contain relevant information. For each audience, one should identify those data practices that are likely unexpected for this audience in the prevalent transaction or context. Those practices are relevant because they cross contextual boundaries [82] and thus violate contextual integrity [15, 91]. Providing notice and choice for such practices should be prioritized. The FTC notes with respect to the IoT that not every data collection requires choice, but that users should have control over unexpected data practices, such as data sharing with third parties [48]. FTC chairwoman Ramirez explains this rationale as follows [107]: "Consumers know, for instance, that a smart thermostat is gathering information about their heating habits, and that a fitness band is collecting data about their physical activity. But would they expect this information to be shared with data brokers or marketing firms? Probably not." In these cases, users need clear privacy notices.

If possible, one should not only rely on estimations of what may be expected or unexpected. User surveys and experiments can reveal actual privacy expectations. Creating personas [90] that represent different members of a specific audience group can help ensure that less obvious concerns are appropriately considered.

For each data practice, all parameters relevant for creating a notice should be gathered. For instance, for a data collection practice this may include by whom information is collected, why, how it is used, for how long it is retained, and if and how it is eventually deleted. For third-party sharing practices, it is relevant with whom information is shared, why, and whether and how usage is restricted or limited in time. Data protection regulation may also provide specific notice requirements (e.g., [41, 42, 43]).

Regardless of regulatory requirements, additional information should be compiled about data practices – especially unexpected ones – to ensure the effectiveness of notices provided to users. The notice should help the recipient make informed privacy decisions. This can be achieved by identifying reasons or benefits for the practice with regard to a specific audience, determining implications and risks for the respective audience, and identifying remedies or choices available to the respective audience. Providing reasons offers the opportunity to explain the purpose of a potentially unexpected, yet benign data practice [85]. Communicating risks [16], for instance with examples [59], supports an individual's assessment of privacy implications, especially when data practices are complex or abstract. Offering specific choices makes the information actionable.

3.4 System Constraints and Opportunities

A specific system may impose constraints on privacy notices that need to be considered in their design. In general, aspects to consider are the different interfaces provided by a system, including their input and output modalities, as well as their relation to specific audience groups. Specific interfaces may have further constraints, such as limited screen real estate. For instance, the FTC [48] notes that providing notice and choice in the context of the IoT can be challenging due to the ubiquity of devices, persistence of collection, and practical obstacles for providing information if devices lack displays or explicit user interfaces. Similar issues have already been recognized in the context of ubiquitous computing [74]. Designing notices for specific audiences may further be limited by how the respective audience can be reached or how they can communicate their privacy choices [103].

Systems may also provide opportunities that can be leveraged to provide a layered and contextualized notice concept for each audience, and potentially even integrate privacy notices and controls into a user's primary activity [113]. By recognizing the constraints, designers may be able to find creative and perhaps novel ways for giving notice. For instance, the lack of explicit user interfaces on a device can be compensated with privacy dashboards, video tutorials, privacy icons or barcodes on the device, and offering choices at the point of sale or in setup wizards [48]. Identified constraints may also be addressed by considering notice mechanisms as part of the system design, i.e., adjusting system features to accommodate notices and controls.

3.5 Layered and Contextualized Notices

While it may be essential to be transparent about many aspects of a system's data practices, showing everything at once in a single notice is rarely effective. Instead, all but the most simple notices should consist of multiple layers. Multi-layered notices constitute a set of complementary privacy notices that are tailored to the respective audience and the prevalent contexts in which they are presented. The granularity of information provided in a specific notice layer must be appropriate for the respective context. For example, a full privacy policy can be complemented by short and condensed notices summarizing the key data practices [9, 85]. Just-in-time or transactional notices provide notice about a specific data practice when it becomes relevant for the user [47], for example, informing about how contact information is used or whether it is shared with third parties when a user registers on a website.

A multi-layered notice concept combines notices shown at different times, using different modalities and interfaces, and varying in terms of content and granularity in a structured approach. For example, some data practices may not require an immediate notice, particularly those that are consistent with users' expectations [46]. It can be expected that a fitness tracker collects information about the user's physical activities – this is the main purpose of the device – thus this collection does not necessarily require prior notice. Automatically uploading a user's activity data to a server or sharing it with other apps may be less expected, thus appropriate notice and choice should be given [85].

Any specific notice should include only the information and control options most relevant and meaningful to a specific audience at that time. Following the details-on-demand pattern [118], initial notices can either point towards additional information and controls or be complemented with alternative user interfaces to review data practices or privacy settings. Deciding what information to include in an initial short notice is a crucial aspect at this stage, because users are more likely to provide consent to the short notice than click through to a more detailed privacy notice. Thus, if such a short notice does not capture all relevant information it may hide information and impair transparency [84]. This is especially an issue for unexpected data practices. Therefore, the notice concept should structure notice layers hierarchically in such a way that the smallest notice either already captures the main aspects of the data practice or draws attention to more expressive notices. Subsequent layers may add additional characteristics.

Designers further need to be aware of not overwhelming users with privacy notices. While many data practices may warrant a notice, providing too many or repetitive privacy notices can result in habituation – users click notices away without considering their content. After a few repetitions, the content of a warning literally does not register anymore in the user's brain [5, 6]. Finding the appropriate number of notices may require user testing. Polymorphic messages [5] or forcing interaction with the notice [21, 22] can reduce habituation effects. A good practice is to prioritize what and when notices are shown based on privacy risks associated with the respective data practice [49].

An example for multi-layered design is the Microsoft Kinect sensor. This device uses video, depth-cameras, and audio to enable users to interact with games through motion and speech. The Kinect has two LEDs that indicate whether motion detection is active or whether video and audio are being recorded and potentially sent to a server. Users can further access a full privacy notice through the screen to which the Xbox is connected, as well as on the Xbox website [137]. Unfortunately, the LED indicators alone cannot make users aware of what information is being collected or shared for what purposes, whereas the policy will likely be ignored by most users. Thus, additional notice layers could enhance awareness and obtain informed consent from users.

In Section 4 we introduce a design space for privacy notices that supports the development of a layered and contextualized notice concept by exposing the relevant dimensions that can be leveraged in the design of individual notices, namely the *timing*, *channel*, and *modality* of notices, as well as the *control* options a notice may provide.

3.6 User-centered Design and Evaluation

Once a notice concept has been developed for each audience, individual notices can be designed and evaluated in a user-centered design process, or by engaging users in participatory design [130]. When conceptual notices for different audiences overlap in terms of timing, channel, modality and content, they can potentially be combined into a single notice serving multiple audiences, as long as the resulting notice meets the requirements of each audience group.

User testing and usability evaluation of notices can be integrated into a system's overall evaluation and quality assurance processes. One should evaluate the individual notices, as well as their combination and the overall notice concept. Notices should be evaluated in the context of the actual system or system prototypes to ensure that they integrate well into the system's interaction design. The effectiveness of notices and warnings can be evaluated along multiple dimensions, such as user attention, comprehension, and recall [8, 12]. It is also important to evaluate whether notices help users make informed choices, both about using a particular service and about exercising choice options [40, 66, 67].

Typically, notices should be evaluated in rigorous user studies. However, budget and time constraints may not always allow for extensive evaluation. In such cases, expert evaluation with usability heuristics [89] can provide at least some indication of the notices' effectiveness. Crowdsourcing platforms also offer an opportunity for conducting quick and inexpensive evaluations of privacy notice design [14].

The outlined best practices support the development of a comprehensive set of privacy notices tailored to a system's different audiences. In the next section, we describe the design space of privacy notices in detail to effectively support the design of individual notices as well as audience-specific notice concepts.

4. DESIGN SPACE OF PRIVACY NOTICES

The design practices outlined in the previous section help to integrate notice design into a system's development process. The purpose of the design space described in this section is to aid the design of specific notices by supporting system designers and privacy engineers in considering the design dimensions of privacy notices. The design space also provides a taxonomy and vocabulary to compare, categorize, and communicate about different notice designs – within a product team as well as with other involved stakeholders, such as the legal department, responsible for drafting the privacy policy, and management. The design space approach has also been used in other privacy and security research, for example, for the creation of a taxonomy of social network data [112], the investigation of web browser Privacy Enhancing Technologies (PETs) [138], and the examination of interfaces for anti-phishing systems [28].

We constructed our design space according to design science principles [102, 126]. Following Peffers et al.'s research methodology [102], we developed and refined the design space in an iterative process, starting with an extensive literature review and the collection and assessment of multiple existing information systems and their privacy notices. This resulted in an initial privacy notice taxonomy, for which we collected feedback in informal discussions with about 20 privacy experts and professionals in summer 2014 at the Symposium on Usable Privacy and Security [120] and at the Workshop on the Future of Privacy Notice and Choice [30]. In further iterations, we refined the design space by taking the expert feedback into consideration and assessing the applicability and expressiveness of the design space in the context of several scenarios grounded in existing privacy notices.

Figure 1 provides an overview of the design space. Its main dimensions are a notice's *timing* (when it is provided), *channel* (how it is delivered), *modality* (what interaction modes are used), and *control* (how are choices provided). In the following, we describe each dimension in detail. Note, that it often makes sense to consider these dimensions in parallel rather than in sequence, as different dimensions can impact each other. Furthermore, the options for each dimension presented here are not meant to be exclusive. The design space can be extended to accommodate novel systems and interaction methods.

4.1 Timing

Timing has been shown to have a significant impact on the effectiveness of notices [12, 40, 56, 100]. Showing a notice at an inopportune time may result in users ignoring the notice rather than shifting their attention to it [132]. Delays between seeing a notice and making a privacy decision (e.g., caused by distractions) can change the user's perception of the notice [98] and even cancel out a notice's effect [3]. Thus, users may make different decisions at different points in time, depending on what primary task they are engaged in, information provided in a notice, and other contextual factors [2]. A comprehensive notice concept should provide notices at different times tailored to a user's needs in that context. We describe six possible timing opportunities here.

4.1.1 At setup

Notice can be provided when a system is used for the first time [85]. For instance, as part of a software installation process users are shown and have to accept the system's terms of use. Receiving and acknowledging a HIPAA privacy notice [125] when checking into a doctor's office in the U.S. can also be considered a setup notice – even if provided on paper. Typically, privacy notices shown at setup time are complemented by a persistently available privacy policy that can be accessed retrospectively by users on demand.

An advantage of providing notices at setup time is that users can inspect a system's data practices before using or purchasing it. The system developer may also prefer to provide information about data practices before use for liability and transparency reasons. Setup notices can be used to make affirmative privacy statements to gain user trust. For example, a form to sign up for an email newsletter may contain a concise statement that email addresses are not shared with third parties [24]. Setup notices also provide the opportunity to explain unexpected data practices that may have a benign purpose in the context of the system [85]. Such explanations can be integrated into the system's setup wizard or video tutorials. Showing privacy information before a website is visited can even impact purchase decisions. Egelman et al. found that participants were more likely to pay a premium at a privacy-protective website when they saw privacy information in search results, as opposed to on the website after selecting a search result [40].

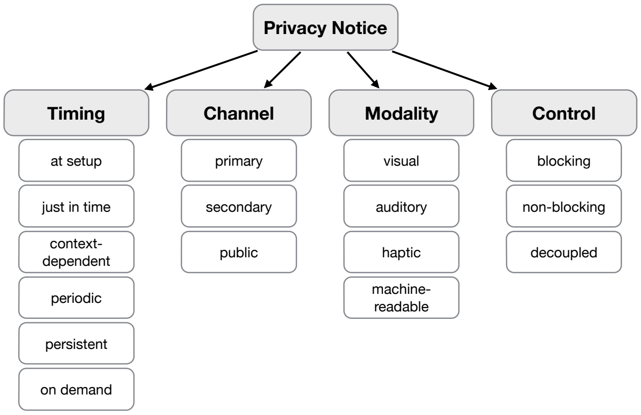


Figure 1: The privacy notice design space.

However, privacy notices at setup also have multiple shortcomings. Users have become largely habituated to install-time notices, such as end-user license agreements, and ignore them [19]. At setup time, users may have difficulty making informed decisions because they have not used the system yet and cannot fully assess its utility or weigh privacy trade-offs. Furthermore, users may be focused on the primary task, namely completing the setup process to be able to use the system, and fail to pay attention to notices [56]. Therefore, privacy notices provided at setup time should be concise and focus on data practices immediately relevant to the primary user rather than presenting extensive terms of service [85]. Integrating privacy information into other materials that explain the functionality of the system may further increase the chance that users do not ignore it.

4.1.2 Just in time

A privacy notice can be shown when a data practice is active, for example when information is being collected, used, or shared. Such notices are referred to as "contextualized" or "just-in-time" notices [13, 68, 85]. Patrick and Kenny [101] first proposed just-in-time click through agreements in order to provide notice and obtain consent with a concise dialog specific to a certain data practice or transactional context. An example of notices triggered by data collection are cookie consent notices shown on websites in Europe [43]. Just-in-time notices can complement or replace setup notices.

Just-in-time notices and obtaining express consent are particularly relevant for data practices considered sensitive or unexpected [48, 85]. For instance, in the case of mobile apps, access to sensitive information such as the user's location, contacts, photos, calendars, or the ability to record audio and video should be accompanied by just-in-time notices [47]. Another example are cars with automatic headlamps that continually sense ambient light conditions; providing notice about this type of data collection might not be necessary. However, privacy expectations may be violated when this information is shared with an insurance company to determine how often the car is driven at night. In such cases, privacy notice as well as choices should be provided. While just-in-time notices enhance transparency and enable users to make privacy decisions in context, users have also been shown to more freely share information if they are given relevant explanations at the time of data collection [68].

Typically, just-in-time notices are shown before data are collected, used, or shared if express user consent is required. On websites, information about how collected data will be used can be presented near input fields in online forms [68]. Just-in-time summary dialogs [7] can show summarized transaction data before it is sent to a service provider. This approach is often used before applications send error or crash reports. Small delays to avoid interrupting the user's primary task may be acceptable [100, 98].

4.1.3 Context-dependent

The user's and system's context can also be considered to show additional notices or controls if deemed necessary [113]. Relevant context may be determined by a change of location, additional users included in or receiving the data, and other situational parameters. Some locations may be particularly sensitive, therefore users may appreciate being reminded that they are sharing their location when they are in a new place, or when they are sharing other information that may be sensitive in a specific context. For example, Wang et al. [129] proposed a notice that provides cues to Facebook users about the audience of their future post to help avoid oversharing. Facebook introduced a privacy checkup message in 2014 that is displayed under certain conditions before posting publicly. It acts as a "nudge" [1, 29] to make users aware that the post will be public and to help them manage who can see their posts (see Figure 2). In sensor-equipped environments, such as smart homes, new users or visitors should also be made aware of what information is being collected and how it is used [74]. Privacy-preserving proximity testing could help determine when the user is near a sensor [10, 75].

Challenges in providing context-dependent notices are detecting relevant situations and context changes. Furthermore, determining whether a context is relevant to an individual's privacy concerns could in itself require access to that person's sensitive data and privacy preferences [113]. However, providing context-specific support may help users make privacy decisions that are more aligned with their desired level of privacy in the respective situation and thus foster trust in the system.

4.1.4 Periodic

Notices can be shown once, the first couple of times a data practice occurs, or every time. The sensitivity of the data practice may determine the appropriate frequency. Additionally, if the notice includes a consent or control option, it may be appropriate to obtain consent on different occasions, depending on the context, user action, or data being collected. However, showing a notice more than once can be overbearing and can lead to notice fatigue [18] and habituation [6, 22]. Thus, repeating notices need to be designed carefully [5] and their frequency needs to be balanced with user needs. Data practices that are reasonably expected as part of the system may require only a single notice, whereas practices falling outside the expected context of use may warrant repeated notices. In general, it is also advisable to show a notice anew if a data practice has changed.

Periodic reminders of data practices can further help users maintain awareness of privacy-sensitive information flows. Reminders are especially appropriate if data practices are largely invisible [10]. For example, in the health domain, patient monitoring devices in the home may remind users on a weekly basis that data is being collected. Those messages make the user aware of the on-going practice and can provide control options. Almuhimedi et al. [4] find that periodic reminders of how often a user's location and other information has been accessed by mobile apps caused participants to adjust and refine their privacy settings. Another example of periodic reminders are the annual privacy notices U.S. financial institutions must provide to customers [35].

A challenge with periodic notices is that they must be relevant to users in order to be not perceived as annoying. Reminders should not be shown too frequently and should focus on data practices about which users may not be aware. If a system has too many data practices requiring reminders, data practices can be prioritized based on their potential privacy impact or a combined notice can remind about multiple data practices. Individual reminders can also be integrated into an overall notification schedule to ensure that users are not overwhelmed. Rotating warnings or changing their look can further reduce habituation effects [5, 132].

4.1.5 Persistent

Persistent notices can provide awareness of ongoing data practices in a less obtrusive manner. A persistent indicator is typically non-blocking and may be shown whenever a data practices is active, for instance when information is being collected continuously or when information is being transmitted [34, 47]. When inactive or not shown, persistent notices also indicate that the respective data practice is currently not active. For instance, Android and iOS display a small icon in the status bar whenever an application accesses the user's location, if the icon is not shown the user's location is not being accessed. Privacy browser plugins, such as Privacy Bird [34] or Ghostery [54], place an icon in the browser's toolbar to inform users about the data practices or third party trackers of the website visited. Recording lights are examples of persistent notices that indicate when a sensor is active. Camcorders, webcams, the Kinect sensor, Google Glass, and other devices feature such indicators.

An issue with such ambient indicators is that they often go unnoticed [105] and that most systems can only accommodate such indicators for a small number of data practices. A system should only provide a small set of persistent indicators to indicate activity of especially critical data practices. Furthermore, persistent indicators should be designed to be noticeable when they are active.

4.1.6 On demand

All previous timing options pertain to the system actively providing notices to users. Users may also actively seek privacy information and request a privacy notice. Therefore, systems should expose opportunities to access privacy notices on demand [85]. A typical example is posting a privacy policy at a persistent location [74] and providing links to it from a website, app, or other privacy notices in the system. A better option are privacy settings interfaces or privacy dashboards within the system that provide information about data practices; controls to manage consent; summary reports of what information has been collected, used, and shared by the system; as well as options to manage or delete collected information. Contact information for a privacy office should be provided to enable users to make written requests.

4.2 Channel

Privacy notices can be delivered through different channels. We distinguish *primary*, *secondary*, and *public* channels. A system may leverage multiple channels to provide different types of notices.

4.2.1 Primary

When a privacy notice is provided on the same platform or device a user interacts with, a primary channel is used for delivering the notice. One example is a privacy notice shown on the user's smartphone that is either provided by the app in use or the operating system. Another example are privacy notices shown on websites. The defining characteristic of a primary channel is that the notice is provided within the user's interaction with the system, i.e., the user is not required to change contexts. Thus, a browser plugin that provides privacy information about a website (e.g., Privacy Bird [34]) would also be considered a primary channel as the notice is provided within the browsing context.

Using a primary channel is typically preferable, because the notice is presented within the context of the system, which supports users in evaluating privacy implications and their privacy preferences [97, 113]. The primary channel is particularly suitable to provide notice to primary users, but can also be used to provide notices to secondary users. For instance, other household members can also be addressed by a smart home appliance's privacy indicators.

4.2.2 Secondary

Some systems may have no or only limited primary channels that can be leveraged for privacy notices and obtaining consent [10]. Wearables, smart home appliances, and IoT devices are examples of systems with constrained interaction capabilities. Such devices may have very small or no displays, which makes it difficult to display notices in an informative way [103]. For instance, privacy policies are more difficult to read on mobile devices [119]. LEDs and other output features could serve as persistent privacy indicators but are often insufficient to communicate relevant aspects of data practices, such as for what purposes data is being collected or with whom it is being shared. Moreover, IoT devices may be installed in remote or less accessible locations. The user may not be near the sensor device when a notice is generated. The user's context may further constrain the use of primary channels for privacy notices. For instance, car owners cannot read detailed privacy notices while driving; users of Internet-connected gym equipment may only want basic information about data sharing while they exercise, but may be interested in learning more about privacy implications when at home.

In such cases, privacy notices can be provided via secondary channels, i.e., outside the respective system or context. A secondary channel leverages out-of-band communication to notify primary and secondary users. For instance, secondary channels can be used to provide setup notices. Rather than showing privacy information on the respective device, choices could be provided at the point of sale (e.g., opt-outs or opt-ins for specific data practices) or as part of video tutorials [48]. Just-in-time, context-dependent, and periodic notices can be delivered as text messages or emails, or any other available communication channel. This requires that the user agrees to receive such notices and provides respective contact information during setup [48]. For instance, the iOS update process gives the option to email oneself the terms of service instead of reading them on the phone.

On-demand notices can be made persistently available at a well-defined location [74], such as posting a (multi-layered) privacy policy on the system's website. Pointers to the privacy policy from the system or device (e.g., using visual markers [10, 48]) can ease access to that privacy notice layer.

An increasingly common approach is to make privacy notices and controls available on a companion device, e.g., on a paired smartphone rather than directly on the wearable or IoT device. Such companion devices provide larger displays and more input and output options to make notices more accessible. Companion devices can also act as privacy proxies [75] for a larger number of constrained devices and systems. Examples are centralized control centers for smart home and IoT devices [48], or privacy and permission managers on mobile devices [4, 47].

4.2.3 Public

Primary and secondary channels are targeted at specific users. However, some systems are not aware of the identity of their users, especially secondary and incidental users. In such cases, public channels can be leveraged to provide notice and potentially choices. Examples of public channel privacy notices are signs posted in public places to inform about video surveillance or a camera's recording indicator.

Public notices can also be supported by technology. IoT devices and other systems may broadcast data practice specifications wirelessly to other devices nearby [10] in so called privacy beacons [75]. For instance, a camera could inform about the purpose of its recordings, how long recordings are retained and who may access them. Such beacons can also inform about available privacy controls [69].

Public channels can also be leveraged by users to communicate their privacy preferences. Markers can be placed on physical objects to control object or face recognition [109]. A privacy beaconing approach can be used to broadcast preferences to others nearby, for instance transmitting the wish to not be photographed to camera phones nearby [70].

4.3 Modality

Different modalities can be used to communicate privacy notices to users. Which modality should be selected depends on what the specific notice strives to achieve, the user's likely attention level, and the system's opportunities and constraints. According to the C-HIP model [32, 132], users process warning messages by switching their attention to them, extracting relevant information from the warning, and comprehending the information; a user's attitudes and beliefs determine if the user acts on the warning. Privacy notices can target each aspect of this process and the choice of modality can increase the effectiveness. For example, if users are engaged in a task that requires visual attention (e.g., driving), using audio to convey privacy information may be more appropriate. Note that not all modalities may be consistently available or effective. Accessibility issues due to physical or visual impairments need to be considered in notice design [128]. Users may also not hear or see a notice due to distractions, blocked line of site, or headphone use. Thus, it is important to evaluate the saliency of different modalities used in notice design [132].

We first discuss visual notices, including text and icons, as they are most common. Auditory and haptic signals can also be used to communicate privacy information. However, they have a lower capacity for conveying information compared to visual notices, which may result in a user preference for visual or textual notices [28]. Quite often, modalities are combined; for example, an audio signal may be used to draw attention to a visual notice displayed on a screen. Finally, machine-readable privacy notices enable the use of different modalities and representations depending on context.

4.3.1 Visual

Visual notices can be provided as text, images, icons, or a combination thereof. Presentation and layout are important aspects in the design of visual notices [132], including colors, fonts, and white space, all of which can impact users' attention and comprehension of the notice [28, 29].

Textual notices can convey complex ideas to users. However, linguistic properties have been shown to influence the perception of warning messages [58]; a notice's framing affects sharing decisions [3]. Specialized terms and jargon may lead to low understanding or the inability to make appropriate privacy decisions [14, 77]. Thus, designers should pay attention to a notice's wording [132], including user testing [14].

While today's website privacy policies are often lengthy [64, 83], privacy notices do not have to be. Relevant information can often be expressed more concisely than in prose. For instance, short notices for smartphone apps have been proposed that convey useful privacy information in the form of risk or expectation scores [53, 79, 80, 88]. Privacy tables and privacy nutrition labels have also been proposed to summarize websites' data practices [66, 84, 87]. Some privacy notice formats have also been standardized by industry or regulators, e.g., financial privacy notices in the U.S. [52]. Standardized notices offer a familiar interface for users, and ease comparison of products [66].

The effectiveness of notices can be increased by personalizing them to the specific user; for instance by including the user's name in the notice [133] or leveraging other user characteristics, such as their demographics or familiarity with a system [132]. An aspect related to personalization is the translation of textual privacy notices into the user's language. Failing to translate may leave international users uninformed about the privacy policy or unable to exercise control over their privacy settings [124].

Images, icons, and LEDs are further options for conveying privacy information visually. Icons can quickly convey privacy settings or currently active data practices. They can be combined with a control switch to activate or deactivate the data practice [48]. However, due to privacy's often abstract nature, images or icons depicting privacy concepts can be difficult to develop. A number of icon sets have been proposed to represent various privacy concepts, both in industry [38, 78, 104, 108] and in research projects [29, 34, 55, 61], with varying levels of success. For example, the AdChoices icon used by the online advertising industry and placed on web ads has been shown to have low user comprehension [78]. Physical indicators, such as LEDs, may use light to visually indicate data practices. LEDs do not have to be binary (on or off) but could leverage colors and blinking patterns to convey different information [60]. Google Glass' display is a transparent glass block that is visibly illuminated if the device is in use, which gives bystanders an indication of whether the device is active.

The meaning of abstract indicators, such as icons or LEDs, often needs to be learned, thus requiring user education. Users may also not notice them [105]. However, when done well pictorial symbols increase the salience and likelihood of a warning being noticed [132], thus, combining icons with textual explanations in privacy notices may improve the effectiveness of the notice, yet, does not require that users learn the exact meaning of the icon.

Visceral notices take an experiential rather than descriptive approach [23]. For example, eyes appearing and growing on a smartphone's home screen relative to how often the user's location has been accessed [114] can leverage strong reactions to anthropomorphic design [23] to provide an ambient sense of exposure.

4.3.2 Auditory

Auditory notices can take at least two forms: spoken word and sounds. Spoken word may be the form of an announcement, pre-recorded or otherwise. One familiar example is the announcement when calling a hotline that the call might be recorded before being connected to a representative.

Sounds can be specialized for the device, or based on well-known sounds in that culture. Calo discusses several examples of visceral notices in which audio signals can "leverage a consumer's familiarity with an old technology" [23]. One example are digital cameras; although some digital cameras and smartphones do not have a physical shutter, they are often configured to emit a shutter sound to make secondary users (i.e., the subjects of the picture) and passersby (incidental users) aware that the device is collecting data by taking a picture. A bill was proposed in the US Congress in 2009 to make such camera shutter sounds mandatory [23]. The bill was not passed, however, some Asian countries have had such requirements for many years.

Auditory warnings can also draw attention to data practices or other notices [132]. For example, the P3P browser plugin Privacy Bird emitted different bird chirping sounds depending on whether the website's privacy policy matched the user's specified privacy preferences [34]. Balebako et al. [13] used sounds to draw attention to occasions when game apps accessed the user's location and other data during game play. Auditory notices face similar challenges as icons – unless familiar [23], their meanings need to be learned. However, they can draw attention to ongoing data practices or privacy notices requiring user attention, especially for systems and devices with constrained interaction capabilities.

4.3.3 Haptic and other

While not widely used for privacy notices yet, haptic feedback provides a potential modality to communicate privacy information. For instance, Balebako et al. [13] combined sound and vibration to notify users about data sharing on smartphones. Similar approaches could be used in wearable devices without displays.

Other modalities taking advantage of human senses, such as smell, wind, ambient lighting, or even taste [71], could be potentially leveraged for privacy notices as well. For instance, olfactory displays [72] could use chemical compounds to generate a pleasant or disgusting smell depending on whether a system or app is privacy-friendly or invasive. Such approaches may warrant further exploration.

4.3.4 *Machine-readable*

The previous modalities directly engage the user's senses. An additional modality offered by technical systems is to encode data practices in a machine-readable format and communicate them to other systems or devices where the information is rendered into a privacy notice. This way, the origin system only needs to specify the data practices and can leave it to the recipient how the information is presented to the user, leveraging that system's input and output capabilities. This also provides the opportunity to present notices in different formats on different devices, or differently for specific audiences. However, there is also a risk that machine-readable data practices are misinterpreted or misrepresented by a device. Transparent documentation, certification, or established guidelines on how the machine-readable format should be interpreted may alleviate this issue [110].

Maganis et al. equipped devices with small displays that show active QR codes, which encode recent data collection history and the device's privacy policy [81]. Privacy beacons [75] have already been mentioned as an approach to transmit machine-readable data practices to other devices.

The Platform for Privacy Preferences (P3P) is a standard machine-readable format for expressing data practices. Websites provide a P3P policy that P3P user agents, such as Privacy Bird [34] or Internet Explorer, can obtain and render. While P3P failed to reach widespread adoption [33], communicating data practices in a machine-readable format may gain acceptance in the IoT context [10, 113]. Smartphone apps or centralized command centers could aggregate privacy information from multiple constrained devices and offer a unified notice format and privacy controls.

4.4 Control

Whenever possible, privacy notices should not only provide information about data practices but also include privacy choices or control options. Choices make the information in privacy notices actionable and enable users to express their consent and their privacy preferences.

The typical choice models are *opt-in*, i.e., the user must explicitly agree to a data practice, and *opt-out*, i.e., the user may advise the system provider to stop a specific practice. However, choices need not be binary. Instead users can be provided with controls to refine purposes for which collected information can be used, specify recipients of information sharing, or vary the granularity of information collected or shared. The goal should be to provide means for users to express preferences globally and selectively [74] instead of a take-it-or-leave-it approach. Controls need to be designed well in order to not overwhelm the user with choices [115].

Furthermore, offering elaborate privacy controls can lead to oversharing over time, either because users feel in control and thus share more [20], or just because of the usability cost of managing the settings [65]. Therefore, default settings need to be carefully considered [85], as they may be kept unchanged out of convenience or because they are interpreted as implicit recommendations [2]. Notice can also give explicit recommendations, for example, as nudges that highlight beneficial choices [1, 4], or with social navigation cues that inform about others' privacy choices [17, 99].

Controls can be directly integrated into the notice, in which case they may be blocking or non-blocking, or they can be decoupled to be used on demand by users. This may be desirable if the control panel is complex or if the notice provides only limited opportunities for integrating control.

4.4.1 *Blocking*

Setup, just-in-time, context-dependent, and periodic notices may include blocking controls. A blocking notice requires the user to make a choice or provide consent based on the information provided in the notice. Until the user provides a choice he or she cannot continue and the respective data practice is blocked. Blocking notices typically constitute opt-in consent, e.g., when terms of service must be accepted in order to use the service, but ideally should provide more meaningful choices. For example, if a smartphone privacy notice states that the camera app can access the device's location, users should be able to selectively allow or deny this access while still being able to use the app.

An issue with such clickthrough agreements [101] is that users may click without reading the provided information. Moving away from just presenting yes and no buttons can increase engagement with the dialog. For instance, Fischer-Hübner et al. propose using a map metaphor on which the user has to drag and drop data onto areas corresponding to their sharing preference [50]. Bravo-Lillo et al. found that forcing users to interact with relevant information in the notice, e.g., by having to move the mouse cursor over a specific text, can effectively address habituation [21, 22].

4.4.2 *Non-blocking*

Blocking controls require engagement, which can be obtrusive. Non-blocking controls can provide control options without forcing user interaction. For instance, Facebook and Google+ provide integrated sharing controls when users create a post. Users who do not interact with these controls have their posts shared according to the same settings as their previous posts. The same dialog can also inform about a post's audience [129]. Privacy notices can also link to a system's privacy settings to ease access to privacy controls without blocking the interaction flow.

4.4.3 *Decoupled*

Some notices may not provide any integrated privacy controls due to system or device constraints. They can be complemented by privacy controls that are decoupled from the specific privacy notice. For instance privacy dashboards and privacy managers enable users to review and change privacy settings when needed [47, 48]. Online companies, like Google, offer such privacy dashboards to control privacy settings across multiple of their services; advertising associations provide websites to allow web users to opt out of targeted advertising for all partners. Apple's iOS provides a settings menu to control privacy settings for installed apps.

Decoupled privacy controls may also take a more holistic approach by attempting to learn users' privacy preferences from their control choices. Those learned preferences could then be applied to other systems, for example, when a new device is connected to the user's smart home system [48].

5. USE CASES

The description of the privacy notice design space highlights the variety of potential privacy notice designs. In this section, we discuss privacy notice approaches in three different domains, how they map onto our design space, and identify potential design alternatives.

5.1 Website & Social Media Privacy Policies

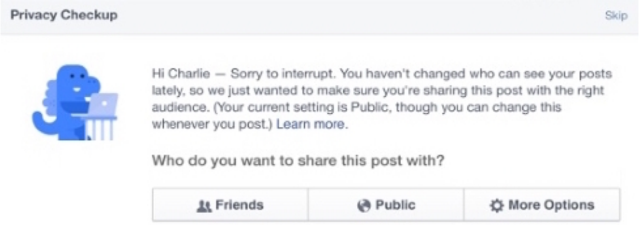
The prevalent approach for providing notice on websites is to post the website's privacy policy on a dedicated page. Audiences of a website's privacy policy are primary and secondary users, as well as regulators. Websites typically provide notices on demand (*timing*), i.e., users need to seek and access the privacy policy if they want to learn about a website's data practices. Website notices typically use the *primary channel*, because websites are not tied to a specific hardware or screen size, and are largely visual (*modality*). Privacy controls are often decoupled from a privacy notice (*control*), i.e., the privacy policy may point to a settings page that allows users to manage privacy, typically by opting-out of certain practices, such as data sharing with advertisers.

The status quo of website privacy notices is a major reason why notice and choice is considered ineffective [25, 33, 116]. However, some websites have developed more effective notice concepts. For instance, Microsoft [86], Facebook [44], and others have implemented multi-layered privacy policies that are also interactive. Explanations are integrated into the privacy page and details are provided on demand rather than showing a long privacy policy. But improving the presentation of the privacy policy is not sufficient if users do not access it. Users require notices integrated into the website in addition to a privacy policy.

One approach is to leverage different timing options, such as just-in-time and contextual notices. Notices can be shown when a data practice occurs for the first time or when the user uses a specific feature for the first time. Notices can be integrated into online forms to make users aware of how their provided data is used and with whom it may be shared. Browsers also provide just-in-time notices for resource access, e.g., when a website wants to use the user's location. Contextual notices can be used to warn about potentially unintended settings. For instance, Facebook introduced a privacy checkup warning when posting publicly, see Figure 2. This blocking notice explains the potential issue and offers integrated privacy controls. The same notice could be realized with non-blocking controls, e.g., as a banner below the post entry field; by blocking the publishing of the post, users are forced to validate their settings. Also note that the dialog does not contain an "OK" button; the user needs to make a specific choice.

Varying a notice's channel is not meaningful for most websites, because the primary channel is well accessible. However, secondary channels, such as email, SMS, and mobile app notifications, are being used by social media sites to provide privacy-relevant notifications, for instance, when one has been tagged in a photo or receives a friend request.

Most privacy controls on websites are decoupled from specific notices. But account registration forms may require users to provide opt-in consent for certain data practices before the account can be created. Cookie consent notices as required by European data protection regulation have been implemented as blocking notices, as well as non-blocking notices, e.g., a banner shown at the top of a page that does not impair use of the website.


Privacy Checkup

Hi Charlie — Sorry to interrupt. You haven't changed who can see your posts lately, so we just wanted to make sure you're sharing this post with the right audience. (Your current setting is Public, though you can change this whenever you post.) Learn more.

Who do you want to share this post with?

Figure 2: Facebook's privacy checkup notice warns the user before posting publicly.

5.2 Smartphone app permissions

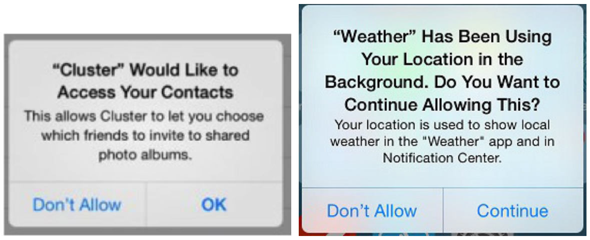
In contrast to websites' privacy policies, privacy and permission management on smartphones employs a more interactive approach. Whereas websites manage privacy on their own, mobile platforms regulate who installed apps access sensors and user resources (e.g., contacts and text messages). Currently, the two major platforms, Android and iOS, take different approaches in terms of how they provide privacy notices and controls to users concerning apps' access to resources. In the following, we discuss how their approaches utilize different parts of the design space. We focus on the current versions of those platforms (iOS 8.x and Android 5.x).

For both iOS and Android the smartphone itself is the *primary channel*. Both systems show privacy notices mainly on the device. Apps can also be installed via a *secondary channel*, namely a Web store for Android and the iTunes application for iOS. While this secondary channel is available via a computer, the notice design is almost identical to app installation directly on the device.

In terms of *modality*, both systems primarily use *visual notices*. Android further requires apps to declare requested permissions in a manifest (*machine-readable*), while iOS apps may specify usage descriptions for access of restricted resources (e.g., location or contacts).

In their app stores, both platforms provide links to an app's privacy policy, which users may access *on demand*. Android further integrates privacy notices into an app's installation process (*at setup*). The user sees a screen that lists the requested app permissions, and the user must either accept all permissions (*blocking*) or not install the app. When an app update changes the requested permissions, a similar notice is shown (*periodic*).

The app installation process on iOS does not include any privacy notices. Instead, iOS shows notices when an app wants to access a resource for the first time, see Figure 3. These notices are *blocking* and ask the user to allow or deny the access request. The notice may contain a developer-specified explanation [121]. In addition, many iOS apps integrate explanations into the application flow before the access request is shown. As of iOS 8, the app developer can also choose to show the authorization notice in advance, but iOS enforces that a resource cannot be accessed without user authorization. In iOS 8, permission requests are not periodic and are only requested once [76]. However, iOS shows peri-



"Cluster" Would Like to Access Your Contacts

This allows Cluster to let you choose which friends to invite to shared photo albums.

"Weather" Has Been Using Your Location in the Background. Do You Want to Continue Allowing This?

Your location is used to show local weather in the "Weather" app and in Notification Center.

Figure 3: iOS's just-in-time notice with purpose explanation (*left*) and periodic reminder (*right*).

odic reminders for apps that have permission to access the user's location in the background, see Figure 3. Both iOS and Android also use a *persistent* location icon in the toolbar indicating that location information is being accessed.

On iOS, users can access a privacy settings menu that facilitates inspection and adjustment of privacy settings for specific resources, globally as well as for specific apps (*on demand, decoupled*). Android provided a similar option (AppOps) in a previous version, but the current version does not allow users to change an app's permissions. Users can inspect an app's permissions in the app store but the only option for privacy control is to uninstall the app.

Thus, the main difference between the platforms is the level of control afforded to the user. iOS users may choose to deny an app access to any specific resource requested, yet continue to use the app. In contrast, Android users must accept all of an app's permissions in order to use it. Android could make better use of different timing options for notices and offer more controls; iOS leverages the options in the privacy notice design space more comprehensively at the time of writing. However, Google announced in May 2015 that Android will also allow users to grant or revoke permissions selectively in the future [39].

5.3 Photo & Video Lifelogging

Lifelogging [117] aims to support memorization and retrieval of everyday events. A common aspect of lifelogging approaches is the recording of photos and videos at frequent intervals, e.g., with GoPro cameras or neck-worn cameras that automatically take a picture every few minutes. A major issue with those technologies is that they not only record the primary user but also bystanders (*incidental users*) [62]. Yet, privacy notices for lifelogging devices, such as the Autographer camera [11] or Google Glass, are mainly targeted at the primary user. They typically provide a privacy policy on the manufacturer's website, privacy settings in a mobile companion app, or a web portal to control sharing and access to the data stream (*secondary channel, on demand, decoupled*). Incidental users neither receive privacy notices nor have options to control being recorded, except for a recording indicator light or a shutter sound on some devices.

Based on the design space, we can consider alternatives to notify and give control to incidental users. Ideally, incidental users should be informed at the moment they are being recorded or soon afterwards (*just-in-time*) and should be given the opportunity to withdraw consent (*control*). Notices on the device (*primary channel*) are likely not effective, as they may not be salient. In order to leverage a secondary channel, e.g., send a notification to the bystander's smartphone, the bystander would need to be identified in order to determine whom to contact, which introduces additional privacy implications. Another option is to use a *public channel*, for instance by wirelessly broadcasting a *machine-readable* notice that a photo has been taken. The incidental user's device could render the notice visually and use sound or vibration to draw attention to the visual notice (*modalities*). A blocking control option does not make much sense in the context of taking photos and videos, as the primary user would have to wait until consent is collected from all bystanders, even though bystanders and the user may be in motion. Thus, incidental users could be given the *non-blocking* option to retroactively opt-out of being photographed. This choice would need to be relayed back to the primary user's device, which could then either delete the photo or detect and remove or blur the incidental user (which poses additional technical challenges that would need to be addressed). While this could provide a viable solution, it also requires bystanders to express their consent any time someone takes a photo nearby, which may become cumbersome in crowded places or at popular tourist spots. An incidental user's preferences could either be stored on their device, which then could automatically respond to such photo notifications, or the incidental user's photo preferences could be broadcast to photographers nearby [70].

6. CONCLUSIONS

We presented a design space that provides a structured approach and vocabulary to discuss and compare different privacy notice designs. This can support the design of privacy notices and controls. The design space should be leveraged as part of a comprehensive design process that focuses on audience-specific privacy notice requirements and considers a system's opportunities and constraints, in order to develop a notice and choice concept that is well integrated with the respective system, rather than bolted on. Notices should be evaluated in user studies.

A key aspect of effective notice design is the realization that a privacy policy, which may be necessary for regulatory compliance, is insufficient and often unsuitable for informing users. Privacy policies need to be accompanied by a notice concept that leverages the options provided in the notice design space to provide information relevant to the targeted audience and to make that information actionable by providing real choices. Actionability is important, because privacy notices without control may leave users feeling helpless [100]. Empowering users with privacy controls increases their trust and may result in increased use and disclosure [20].

Novel technologies and integrated devices, such as wearables or the Internet of Things, pose new challenges for the design of privacy notices and controls. Information collection is continuous and sharing paramount [73]. Public policy, legislation, and technological approaches need to work together to enable users to manage their privacy in such systems. The identified best practices and the proposed design space provide the means to reason about meaningful design options for notice and control in such systems. For instance, by leveraging alternative channels or modalities, and providing notices and control options at different times in the information lifecycle. A future challenge is to develop and provide tools to support the identification of notice requirements, system opportunities, and applicable options in the design space, and explore the (semi-)automated generation of notice and control interfaces.

Acknowledgments

This research was partially funded by NSF grants CNS-1012763, CNS-1330596, and DGE-0903659, as well as by Facebook. The authors would like to thank Aditya Marella and Maritza Johnson for initial contributions to this project; the privacy and usability experts who provided feedback on iterations of the design space; as well as Alessandro Acquisti, Sameer Patil, Yang Wang, our shepherd Andrew Patrick, and our reviewers for feedback on earlier drafts.

7. REFERENCES

[1] A. Acquisti. Nudging privacy: The behavioral economics of personal information. *IEEE Security Privacy*, 7(6):82–85, 2009.

[2] A. Acquisti, L. Brandimarte, and G. Loewenstein. Privacy and human behavior in the age of information. *Science*, 347(6221):509–514, 2015.

[3] I. Adjerid, A. Acquisti, L. Brandimarte, and G. Loewenstein. Sleights of privacy: Framing, disclosures, and the limits of transparency. In *Proc. SOUPS '13*, page 9. ACM, 2013.

[4] H. Almuhimedi, F. Schaub, N. Sadeh, I. Adjerid, A. Acquisti, J. Gluck, L. F. Cranor, and Y. Agarwal. Your location has been shared 5,398 times! a field study on mobile app privacy nudging. In *Proc. CHI '15*. ACM, 2015.

[5] B. Anderson, B. Kirwan, D. Eargle, S. Howard, and A. Vance. How polymorphic warnings reduce habituation in the brain – insights from an fMRI study. In *Proc. CHI '15*. ACM, 2015.

[6] B. Anderson, A. Vance, B. Kirwan, E. D., and S. Howard. Users aren't (necessarily) lazy: Using NeuroIS to explain habituation to security warnings. In *Proc. ICIS '14*, 2014.

[7] J. Angulo, S. Fischer-Hübner, T. Pulls, and U. König. HCI for Policy Display and Administration. In *Privacy and Identity Management for Life*, pages 261–277. Springer, 2011.

[8] J. J. Argo and K. J. Main. Meta-Analyses of the Effectiveness of Warning Labels. *Journal of Public Policy & Marketing*, 23(2):193–208, Oct. 2004.

[9] Article 29 Data Protection Working Party. Opinion 10/2004 on More Harmonised Information Provisions. WP 100, Nov. 2004.

[10] Article 29 Data Protection Working Party. Opinion 8/2014 on the Recent Developments on the Internet of Things. WP 223, Sept. 2014.

[11] Autographer. http://www.autographer.com, 2012. accessed: 2015-06-01.

[12] R. Balebako. *Mitigating the Risks of Smartphone Data Sharing: Identifying Opportunities and Evaluating Notice*. PhD thesis, Engineering and Public Policy, Carnegie Mellon University, 2014.

[13] R. Balebako, J. Jung, W. Lu, L. F. Cranor, and C. Nguyen. Little brothers watching you: Raising awareness of data leaks on smartphones. In *Proc. SOUPS '13*. ACM, 2013.

[14] R. Balebako, R. Shay, and L. F. Cranor. Is your inseam a biometric? a case study on the role of usability studies in developing public policy. In *Proc. USEC '14*, 2014.

[15] L. Barkhuus. The Mismeasurement of Privacy: Using Contextual Integrity to Reconsider Privacy in HCI. In *Proc. CHI '12*. ACM, 2012.

[16] L. Bauer, C. Bravo-Lillo, L. F. Cranor, and E. Fragkaki. Warning design guidelines. Tech. report CMU-CyLab-13-002, CyLab, Carnegie Mellon University, 2013.

[17] A. Besmer, J. Watson, and H. R. Lipford. The impact of social navigation on privacy policy configuration. In *Proc. SOUPS '10*. ACM, 2010.

[18] R. Böhme and J. Grossklags. The security cost of cheap user interaction. In *Proc. Workshop on New Security Paradigms*. ACM, 2011.

[19] R. Böhme and S. Köpsell. Trained to accept?: A field experiment on consent dialogs. In *Proc. CHI '10*. ACM, 2010.

[20] L. Brandimarte, A. Acquisti, and G. Loewenstein. Misplaced confidences privacy and the control paradox. *Social Psychological and Personality Science*, 4(3):340–347, 2013.

[21] C. Bravo-Lillo, L. F. Cranor, S. Komanduri, S. Schechter, and M. Sleeper. Harder to ignore? Revisiting pop-up fatigue and approaches to prevent it. In *Proc. SOUPS '14*, 2014.

[22] C. Bravo-Lillo, S. Komanduri, L. F. Cranor, R. W. Reeder, M. Sleeper, J. Downs, and S. Schechter. Your attention please: Designing security-decision uis to make genuine risks harder to ignore. In *Proc. SOUPS '13*. ACM, 2013.

[23] R. Calo. Against notice skepticism in privacy (and elsewhere). *Notre Dame Law Review*, 87(3):1027–1072, 2012.

[24] J. Cannon. *Privacy in Technology*. IAPP, 2014.

[25] F. Cate. The Limits of Notice and Choice. *IEEE Security Privacy*, 8(2):59–62, Mar. 2010.

[26] Center for Information Policy Leadership. Ten Steps to Develop a Multilayered Privacy Notice. White paper, Mar. 2007.

[27] Centers for Medicare & Medicaid Services. The Health Insurance Portability and Accountability Act of 1996 (HIPAA). http://www.cms.hhs.gov/hipaa/, 1996.

[28] Y. Chen, F. M. Zahedi, and A. Abbasi. Interface design elements for anti-phishing systems. In *Service-Oriented Perspectives in Design Science Research*, pages 253–265. Springer, 2011.

[29] E. Choe, J. Jung, B. Lee, and K. Fisher. Nudging people away from privacy-invasive mobile apps through visual framing. In *Proc. INTERACT '13*. Springer, 2013.

[30] CMU CyLab. Workshop on the future of privacy notice and choice. https://www.cylab.cmu.edu/news_events/events/fopnac/, June 27 2015.

[31] L. Cranor. Giving notice: Why privacy policies and security breach notifications aren't enough. *IEEE Communications Magazine*, 43(8):18–19, Aug. 2005.

[32] L. F. Cranor. A framework for reasoning about the human in the loop. In *Proc. UPSEC '08*. USENIX Assoc., 2008.

[33] L. F. Cranor. Necessary but not sufficient: Standardized mechanisms for privacy notice and

choice. *Journal on Telecommunications and High Technology Law*, 10:273, 2012.

[34] L. F. Cranor, P. Guduru, and M. Arjula. User interfaces for privacy agents. *ACM TOCHI*, 13(2):135–178, 2006.

[35] L. F. Cranor, K. Idouchi, P. G. Leon, M. Sleeper, and B. Ur. Are they actually any different? Comparing thousands of financial institutions' privacy practices. In *Proc. WEIS '13*, 2013.

[36] G. Danezis, J. Domingo-Ferrer, M. Hansen, J.-H. Hoepman, D. Le Métayer, R. Tirtea, and S. Schiffner. Privacy and Data Protection by Design – from policy to engineering. report, ENISA, Dec. 2014.

[37] M. Deng, K. Wuyts, R. Scandariato, B. Preneel, and W. Joosen. A privacy threat analysis framework: supporting the elicitation and fulfillment of privacy requirements. *Requirements Engineering*, 16(1):3–32, Nov. 2010.

[38] Disconnect.me. Privacy policies are too complicated: We've simplified them. https://disconnect.me/icons, Dec. 2014. accessed: 2015-06-01.

[39] J. Eason. Android M developer preview & tools. Android Developers Blog, May 28 2015. http://android-developers.blogspot.com/2015/05/android-m-developer-preview-tools.html, accessed: 2015-06-01.

[40] S. Egelman, J. Tsai, L. F. Cranor, and A. Acquisti. Timing is everything?: the effects of timing and placement of online privacy indicators. In *Proc. CHI '09*. ACM, 2009.

[41] European Parliament and Council. Directive 95/46/EC on the protection of individuals with regard to the processing of personal data and on the free movement of such data. *Official Journal of the European Communities*, (L 281):31–50, 1995.

[42] European Parliament and Council. Directive 2002/58/EC concerning the processing of personal data and the protection of privacy in the electronic communications sector (Directive on privacy and electronic communications). *Official Journal of the European Communities*, (L 201), 2002.

[43] European Parliament and Council. Directive 2009/136/EC. *Official Journal of the European Communities*, (L 337), 2009.

[44] Facebook. Data policy. https://www.facebook.com/privacy/explanation, 2015. accessed: 2015-06-01.

[45] Federal Trade Commission. Privacy online: a report to Congress. FTC report, 1998.

[46] Federal Trade Commission. Protecting consumer privacy in an era of rapid change. FTC report, 2012.

[47] Federal Trade Commission. Mobile privacy disclosures: Building trust through transparency. FTC staff report, Feb. 2013.

[48] Federal Trade Commission. Internet of things: Privacy & security in a connected world. FTC staff report, Jan. 2015.

[49] A. Felt, S. Egelman, M. Finifter, D. Akhawe, and D. Wagner. How to ask for permission. In *Proc. HOTSEC '12*, 2012.

[50] S. Fischer-Hübner, J. S. Pettersson, M. Bergmann, M. Hansen, S. Pearson, and M. C. Mont. HCI Designs for Privacy-Enhancing Identity Management. In *Digital Privacy: Theory, Technologies, and Practices*, pages 229–252. Auerbach Pub., 2008.

[51] H. Fu, Y. Yang, N. Shingte, J. Lindqvist, and M. Gruteser. A field study of run-time location access disclosures on android smartphones. In *Proc. USEC '14*, 2014.

[52] L. Garrison, M. Hastak, J. M. Hogarth, S. Kleimann, and A. S. Levy. Designing Evidence-based Disclosures: A Case Study of Financial Privacy Notices. *Journal of Consumer Affairs*, 46(2):204–234, June 2012.

[53] C. Gates, N. Li, H. Peng, B. Sarma, Y. Qi, R. Potharaju, C. Nita-Rotaru, and I. Molloy. Generating summary risk scores for mobile applications. *IEEE Trans. Dependable and Secure Computing*, 11(3):238–251, May 2014.

[54] Ghostery. https://www.ghostery.com. accessed: 2015-06-01.

[55] J. Gomez, T. Pinnick, and A. Soltani. KnowPrivacy. Final report, UC Berkeley, School of Information, 2009.

[56] N. S. Good, J. Grossklags, D. K. Mulligan, and J. A. Konstan. Noticing notice: a large-scale experiment on the timing of software license agreements. In *Proc. CHI '07*. ACM, 2007.

[57] G. Greenleaf. Sheherezade and the 101 data privacy laws: Origins, significance and global trajectories. *Journal of Law, Information and Science*, 23(1):4–49, 2014.

[58] M. Harbach, S. Fahl, P. Yakovleva, and M. Smith. Sorry, I don't get it: An analysis of warning message texts. In *Proc. USEC '13*. Springer, 2013.

[59] M. Harbach, M. Hettig, S. Weber, and M. Smith. Using personal examples to improve risk communication for security & privacy decisions. In *Proc. CHI '14*. ACM, 2014.

[60] C. Harrison, J. Horstman, G. Hsieh, and S. Hudson. Unlocking the expressivity of point lights. In *Proc. CHI '12*. ACM, 2012.

[61] L.-E. Holtz, H. Zwingelberg, and M. Hansen. Privacy Policy Icons. In *Privacy and Identity Management for Life*, pages 279–285. Springer, 2011.

[62] G. Iachello, K. N. Truong, G. D. Abowd, G. R. Hayes, and M. Stevens. Prototyping and sampling experience to evaluate ubiquitous computing privacy in the real world. In *Proc. CHI '06*. ACM, 2006.

[63] P. G. Inglesant and M. A. Sasse. The True Cost of Unusable Password Policies: Password Use in the Wild. In *Proc. CHI '10*. ACM, 2010.

[64] C. Jensen and C. Potts. Privacy policies as decision-making tools: an evaluation of online privacy notices. In *Proc. CHI '04*. ACM, 2004.

[65] M. J. Keith, C. Maynes, P. B. Lowry, and J. Babb. Privacy fatigue: The effect of privacy control complexity on consumer electronic information disclosure. In *Proc. ICIS '14*. SSRN, 2014.

[66] P. G. Kelley, L. Cesca, J. Bresee, and L. F. Cranor. Standardizing privacy notices: an online study of the nutrition label approach. In *Proc. CHI '10*. ACM,

2010.

[67] P. G. Kelley, L. F. Cranor, and N. Sadeh. Privacy as part of the app decision-making process. In *Proc. CHI '13*. ACM, 2013.

[68] A. Kobsa and M. Teltzrow. Contextualized communication of privacy practices and personalization benefits: Impacts on users' data sharing and purchase behavior. In *Proc. PETS '05*. Springer, 2005.

[69] B. Könings, F. Schaub, and M. Weber. PriFi beacons: piggybacking privacy implications on wifi beacons. In *Ubicomp '13 Adjunct Proceedings*. ACM, 2013.

[70] B. Könings, S. Thoma, F. Schaub, and M. Weber. Pripref broadcaster: Enabling users to broadcast privacy preferences in their physical proximity. In *Proc. MUM '14*. ACM, 2014.

[71] P. Kortum. *HCI beyond the GUI: Design for haptic, speech, olfactory, and other nontraditional interfaces*. Morgan Kaufmann, 2008.

[72] P. Kortum. *HCI beyond the GUI: Design for haptic, speech, olfactory, and other nontraditional interfaces*. Morgan Kaufmann, 2008.

[73] S. Landau. Control use of data to protect privacy. *Science*, 347(6221):504–506, Jan. 2015.

[74] M. Langheinrich. Privacy by design – principles of privacy-aware ubiquitous systems. In *Proc. UbiComp '01*. Springer, 2001.

[75] M. Langheinrich. A Privacy Awareness System for Ubiquitous Computing Environments. In *Proc. UbiComp '02*. Springer, 2002.

[76] M. Lazer-Walker. Core location in ios 8. http://nshipster.com/core-location-in-ios-8/, 2014. accessed: 2015-06-01.

[77] P. Leon, B. Ur, R. Shay, Y. Wang, R. Balebako, and L. Cranor. Why Johnny can't opt out: A usability evaluation of tools to limit online behavioral advertising. In *Proc. CHI '12*. ACM, 2012.

[78] P. G. Leon, J. Cranshaw, L. F. Cranor, J. Graves, M. Hastak, B. Ur, and G. Xu. What do online behavioral advertising privacy disclosures communicate to users? In *Proc. WPES '12*. ACM, 2012.

[79] I. Liccardi, J. Pato, D. J. Weitzner, H. Abelson, and D. De Roure. No technical understanding required: Helping users make informed choices about access to their personal data. In *Proc. MOBIQUITOUS '14*. ICST, 2014.

[80] J. Lin, S. Amini, J. I. Hong, N. Sadeh, J. Lindqvist, and J. Zhang. Expectation and purpose: understanding users' mental models of mobile app privacy through crowdsourcing. In *Proc. UbiComp '12*. ACM, 2012.

[81] G. Maganis, J. Jung, T. Kohno, A. Sheth, and D. Wetherall. Sensor tricorder: What does that sensor know about me? In *Proc. HotMobile '11*. ACM, 2011.

[82] G. Marx. Murky conceptual waters: The public and the private. *Ethics and Information technology*, pages 157–169, 2001.

[83] A. M. McDonald and L. F. Cranor. The Cost of Reading Privacy Policies. *I/S: A Journal of Law and Policy for the Information Society*, 4(3):540–565, 2008.

[84] A. M. Mcdonald, R. W. Reeder, P. G. Kelley, and L. F. Cranor. A comparative study of online privacy policies and formats. In *Proc. PETS '09*. Springer, 2009.

[85] Microsoft. Privacy Guidelines for Developing Software Products and Services. Technical Report version 3.1, 2008.

[86] Microsoft. Microsoft.com privacy statement. https://www.microsoft.com/privacystatement/en-us/core/default.aspx, 2014. accessed: 2015-06-01.

[87] G. R. Milne, M. J. Culnan, and H. Greene. A longitudinal assessment of online privacy notice readability. *Journal of Public Policy & Marketing*, 25(2):238–249, 2006.

[88] A. Mylonas, M. Theoharidou, and D. Gritzalis. Assessing privacy risks in android: A user-centric approach. In *Workshop on Risk Assessment and Risk-Driven Testing*. Springer, 2014.

[89] J. Nielsen and R. Molich. Heuristic evaluation of user interfaces. In *Proc. CHI '90*. ACM, 1990.

[90] L. Nielsen. Personas. In *The Encyclopedia of Human-Computer Interaction*. The Interaction Design Foundation, 2nd ed. edition, 2014. https://www.interaction-design.org/encyclopedia/personas.html.

[91] H. Nissenbaum. A contextual approach to privacy online. *Daedalus*, 140(4):32–48, 2011.

[92] NTIA. Short form notice code of conduct to promote transparency in mobile app practices. Redline draft, July 2013. http://www.ntia.doc.gov/files/ntia/publications/july_25_code_draft.pdf.

[93] NTIA. Privacy multistakeholder process: Facial recognition technology, 2014. http://www.ntia.doc.gov/other-publication/2014/privacy-multistakeholder-process-facial-recognition-technology, accessed: 2015-06-01.

[94] OECD. Making Privacy Notices Simple. Digital Economy Papers 120, July 2006. http://www.oecd-ilibrary.org/science-and-technology/making-privacy-notices-simple_231428216052.

[95] OECD. The OECD Privacy Framework. Report, 2013. http://www.oecd.org/sti/ieconomy/oecd_privacy_framework.pdf.

[96] Official California Legislative Information. The Online Privacy Protection Act of 2003, 2003.

[97] L. Palen and P. Dourish. Unpacking "privacy" for a networked world. In *Proc. CHI '03*. ACM, 2003.

[98] S. Patil, R. Hoyle, R. Schlegel, A. Kapadia, and A. J. Lee. Interrupt now or inform later?: Comparing immediate and delayed privacy feedback. In *Proc. CHI '15*. ACM, 2015.

[99] S. Patil, X. Page, and A. Kobsa. With a little help from my friends: Can social navigation inform interpersonal privacy preferences? In *Proc. CSCW '11*. ACM, 2011.

[100] S. Patil, R. Schlegel, A. Kapadia, and A. J. Lee.

Reflection or action?: How feedback and control affect location sharing decisions. In *Proc. CHI '14*. ACM, 2014.

[101] A. Patrick and S. Kenny. From privacy legislation to interface design: Implementing information privacy in human-computer interactions. In *Proc. PET '03*. Springer, 2003.

[102] K. Peffers, T. Tuunanen, M. A. Rothenberger, and S. Chatterjee. A design science research methodology for information systems research. *Journal of management information systems*, 24(3):45–77, 2007.

[103] S. R. Peppet. Regulating the Internet of Things: First Steps Toward Managing Discrimination, Privacy, Security, and Consent. *Texas Law Review*, 93(85):85–176, 2014.

[104] T. Pinnick. Privacy short notice design. TRUSTe blog, Feb. 2011. http://www.truste.com/blog/2011/02/17/privacy-short-notice-design/, accessed: 2015-06-01.

[105] R. S. Portnoff, L. N. Lee, S. Egelman, P. Mishra, D. Leung, and D. Wagner. Somebody's watching me? assessing the effectiveness of webcam indicator lights. In *Proc. CHI '15*, 2015.

[106] President's Concil of Advisors on Science and Technology. Big data and privacy: A technological perspective. Report to the President, Executive Office of the President, May 2014.

[107] E. Ramirez. Privacy and the IoT: Navigating policy issues. CES Opening Remarks, 2015. FTC public statement.

[108] A. Raskin. Privacy icons: Alpha release. http://www.azarask.in/blog/post/privacy-icons/. accessed: 2015-06-01.

[109] N. Raval, A. Srivastava, K. Lebeck, L. Cox, and A. Machanavajjhala. Markit: Privacy markers for protecting visual secrets. In *UbiComp '14 Adjunct Proceedings*. ACM, 2014.

[110] J. Reidenberg and L. F. Cranor. Can User Agents Accurately Represent Privacy Policies? Available at SSRN: http://papers.ssrn.com/abstract=328860, 2002.

[111] J. R. Reidenberg, T. Breaux, L. F. Cranor, B. French, A. Grannis, J. T. Graves, F. Liu, A. M. McDonald, T. B. Norton, R. Ramanath, N. C. Russell, N. Sadeh, and F. Schaub. Disagreeable privacy policies: Mismatches between meaning and users' understanding. *Berkeley Technology Law Journal*, 30, 2015.

[112] C. Richthammer, M. Netter, M. Riesner, J. Sänger, and G. Pernul. Taxonomy of social network data types. *EURASIP Journal on Information Security*, 2014(1):1–17, 2014.

[113] F. Schaub, B. Könings, and M. Weber. Context-adaptive privacy: Leveraging context awareness to support privacy decision making. *IEEE Pervasive Computing*, 14(1):34–43, 2015.

[114] R. Schlegel, A. Kapadia, and A. J. Lee. Eyeing your exposure: Quantifying and controlling information sharing for improved privacy. In *Proc. SOUPS '11*. ACM, 2011.

[115] B. Schwartz. *The Paradox of Choice: Why More is Less*. HarperCollins Publishers, 2004.

[116] P. M. Schwartz and D. Solove. Notice & Choice. In *The Second NPLAN/BMSG Meeting on Digital Media and Marketing to Children*, 2009.

[117] A. J. Sellen and S. Whittaker. Beyond total capture: A constructive critique of lifelogging. *Commun. ACM*, 53(5):70–77, May 2010.

[118] B. Shneiderman. The eyes have it: A task by data type taxonomy for information visualizations. In *Proc. Symp. on Visual Languages*. IEEE, 1996.

[119] R. I. Singh, M. Sumeeth, and J. Miller. Evaluating the readability of privacy policies in mobile environments. *International Journal of Mobile Human Computer Interaction*, 3(1):55–78, 2011.

[120] SOUPS 2014 organizing committee. Tenth Symposium on Usable Privacy and Security. http://cups.cs.cmu.edu/soups/2014/, July 9–11 2014.

[121] J. Tan, K. Nguyen, M. Theodorides, H. Negrón-Arroyo, C. Thompson, S. Egelman, and D. Wagner. The effect of developer-specified explanations for permission requests on smartphone user behavior. In *Proc. CHI '14*. ACM, 2014.

[122] The White House. Consumer data privacy in a networked world. Technical report, Feb. 2012. http://www.whitehouse.gov/sites/default/files/privacy-final.pdf.

[123] B. Ur, J. Jung, and S. Schechter. Intruders versus intrusiveness: teens' and parents' perspectives on home-entryway surveillance. In *Proc. UbiComp '14*. ACM, 2014.

[124] B. Ur, M. Sleeper, and L. F. Cranor. Privacy policies in social media: Providing translated privacy notice. *I/S: A Journal of Law and Policy for the Information Society*, 9(2), 2013.

[125] U.S. Department of Health & Human Services. Notice of privacy practices for protected health information, April 2003.

[126] R. H. von Alan, S. T. March, J. Park, and S. Ram. Design science in information systems research. *MIS quarterly*, 28(1):75–105, 2004.

[127] W3C. Tracking protection working group. http://www.w3.org/2011/tracking-protection/. accessed: 2015-06-01.

[128] W3C. Web accessibility and usability working together. http://www.w3.org/WAI/intro/usable. accessed: 2015-06-01.

[129] Y. Wang, P. G. Leon, A. Acquisti, L. F. Cranor, A. Forget, and N. Sadeh. A field trial of privacy nudges on facebook. In *Proc. CHI '14*. ACM, 2014.

[130] S. Weber, M. Harbach, and M. Smith. Participatory Design for Security-Related User Interfaces. In *Proc. USEC '15*, 2015.

[131] R. Wenning, M. Schunter, L. Cranor, B. Dobbs, S. Egelman, G. Hogben, J. Humphrey, M. Langheinrich, M. Marchiori, M. Presler-Marshall, J. Reagle, and D. A. Stampley. The Platform for Privacy Preferences 1.1 (P3P 1.1) Specification. http://www.w3.org/TR/P3P11/, 2006. accessed: 2015-06-01.

[132] M. S. Wogalter, V. C. Conzola, and T. L. Smith-Jackson. Research-based guidelines for warning design and evaluation. *Applied Ergonomics*,

33(3):219–230, 2002.

[133] M. S. Wogalter, B. M. Racicot, M. J. Kalsher, and S. Noel Simpson. Personalization of warning signs: The role of perceived relevance on behavioral compliance. *International Journal of Industrial Ergonomics*, 14(3):233–242, Oct. 1994.

[134] D. Wright. Should privacy impact assessments be mandatory? *Communications of the ACM*, 54(8):121–131, Aug. 2011.

[135] D. Wright. Making Privacy Impact Assessment More Effective. *The Information Society*, 29(5):307–315, Oct. 2013.

[136] D. Wright, K. Wadhwa, P. D. Hert, D. Kloza, and D. G. Justice. A Privacy Impact Assessment Framework for data protection and privacy rights. Deliverable September, PIAF project, 2011.

[137] Xbox.com. Kinect and Xbox One privacy FAQ. http://www.xbox.com/en-US/kinect/privacyandonlinesafety.

[138] H. Xu, R. E. Crossler, and F. Bélanger. A value sensitive design investigation of privacy enhancing tools in web browsers. *Decision Support Systems*, 54(1):424–433, 2012.

“WTH..!?!” Experiences, reactions, and expectations related to online privacy panic situations [*]

Julio Angulo
Karlstad University
Karlstad, Sweden
julio.angulo.r@gmail.com

Martin Ortlieb
Google
Zürich, Switzerland
mortlieb@google.com

ABSTRACT

There are moments in which users might find themselves experiencing feelings of panic with the realization that their privacy or personal information on the Internet might be at risk. We present an exploratory study on common experiences of online privacy-related panic and on users' reactions to frequently occurring privacy incidents. By using the metaphor of a *privacy panic button*, we also gather users' expectations on the type of help that they would like to obtain in such situations. Through user interviews ($n = 16$) and a survey ($n = 549$), we identify 18 scenarios of privacy panic situations. We ranked these scenarios according to their frequency of occurrence and to the concerns of users to become victims of these incidents. We explore users' underlying worries of falling pray for these incidents and other contextual factors common to privacy panic experiences. Based on our findings we present implications for the design of a help system for users experiencing privacy panic situations.

1. INTRODUCTION

With so many of our daily activities spent interacting with information technologies and so much of our personal data being stored and handled online, the chances that an unexpected, unwelcome privacy-related incident occurs at some point in our lives are not very unlikely. New privacy- or security-related incidents are regularly reported through various channels, like news, blogs and collaboratively maintained databases. A report from security company Symantec [58] in 2014 described a worrisome increase of attacks on online services over earlier years, which have resulted on breaches to their customers' data records. However, there are other kinds of privacy-related incidents which can affect individual users directly and emotionally during their daily interactions with online services. We are talking about incidents that, if they occur, might lead users to experience physical symptoms similar to a person in distress (i.e., momentary shortness of breath, accelerated heart rate, rushed adrenaline, tensing of the muscles, etc.). We refer to these cases as *online privacy panic* situations. Such situations might lead users to worry about the possible concequences of the breach to their privacy, which may include losing one's job, being financially defrauded, or even endure physical harm or damage to one's property.

Previous research has studied isolated privacy incidents (e.g., [49], [55], [65] and others), however, as far as we know, no one has tried to capture users' previous experiences of a range of common incidents. Our intention in this paper is to unveil and understand common online privacy panic situations. We investigate some of the contextual factors that characterize such situations, as well as strategies that people take to deal with them. Moreover, we use the metaphor of a *panic button* to investigate users' expectations and mental models of suitable help mechanisms that could lead these users towards a solution, calming their distress, and preventing similar episodes from happening in the future.

To this end, we report on a study consisting of semi-structured interviews ($n = 16$) and a survey ($n = 549$). From the obtained data we identified 18 different cases of online privacy panic. Victims' topmost worries included possible harm to their finances or fear of embarrassment, as well as third-parties knowing things that might not be of their business. Among the most memorable self-reported panic stories were cases of account hijacking and ‘leakage’ of personal data, while incidents involving regrets when sharing content online were found to be experienced most frequently. However, scenarios related to the loss of online data, the loss of a mobile device, or falling pray of identity theft also were at the top of users' concerns. Our findings also indicate that, in the case a service provider were to offer a hypothetical *privacy panic button*, users would expect that the help provided is immediate, uncomplicated, actionable, and in-place. From the results of our study, we present implications for the design of a help system for users experiencing online privacy panic situations.

The rest of this paper is structured as follows. First, we present related work on Internet users' privacy concerns in Section 2.1, as well as research studies which look at different aspects of common privacy incidents in Section 2.2. We then introduce in Section 3 the methodologies used to study users' experienced privacy incidents, including the recruitment of study participants and report on the results obtained. From our findings we present in Section 4 implications for the

[*] (Produces the permission block, and copyright information). For use with SIG-ALTERNATE.CLS. Supported by ACM.

Symposium on Usable Privacy and Security (SOUPS) 2015, July 22–24, 2015, Ottawa, Canada.

design of a possible help system for privacy panic situations. We end with concluding remarks in Section 5.

2. RELATED WORK

In this section we present work related to the study of people's privacy concerns on the Internet and on known privacy incidents that might cause people to experience feelings similar to panic, fear or distress while acting online.

2.1 Users' privacy concerns

Plenty of studies have tried to measure users' privacy concerns on the Internet. For instance, just over a decade ago Earp et al. [16] developed an instrument which indicated that people's primary privacy concerns had to do with the way their personal information was transfered, stored and accessed on the Internet. A follow-up study in 2010 revealed that the types of privacy concerns have not varied much over the years but the levels of these concerns have increased considerably [4]. In 2004, Malhotra et al. [36] introduced the Internet Users' Information Privacy Concerns scale, which tries to capture users levels of concerns with regards to the amount of data collection, the control users have over their data and the awareness they have about service providers' privacy practices. Buchanan et al. [7] developed scales based on the Westin index [27] to understand people's level of technical protection and general caution on the Internet, as well as their level of privacy concerns. Alan Westin himself developed the Privacy Segmentation Index [27], which categorizes individuals' responses into privacy fundamentalist, pragmatist and unconcerned. However, other studies have shown that Westin's categories may be inaccurate predictors of people's real privacy attitudes and behaviours [66].

Despite the efforts of trying to find appropriate measurements of privacy concerns, a survey done in 2002 [21] asked open-ended questions related to people's use of the Internet and found that people's concerns with the Internet had less to do with notions of privacy and more with worries about falling victim to crimes like credit card theft [39]. Another study also found that individuals tend to exhibit optimism biases with relation to online privacy risks, believing that they are less vulnerable to these risks than others [10].

Our study departs from previous studies by focusing on exploring users' actual past experiences with privacy related incidents online. We started our investigation by asking the question '*what are common privacy incidents that are likely to trigger feelings of fear or extreme concern on Internet users with regards to their privacy or personal data?*'

It has been noted that having been a victim of a privacy incident can be closely intertwined with a person's reported levels of privacy concerns [14]. It can be argued that it is the multidimensional [13, 23], dynamic [40] and temporal processes of privacy which can create a change in people's privacy concerns as a result of negative past experiences. All of these, may encourage people to modify their behaviour and become more alert about their future actions online. At the same time, experiencing some kind of privacy incident (or hearing rumors about an incident from someone else) might motivate people to take preventive measures, thereby reducing their privacy concerns for the future, paradoxically making them more vulnerable for incidents later on (e.g., believing that setting up a firewall at one point in time will protect their computer from all future viruses).

In 2009, Paul Buta [8] observed that many of people's privacy-related fears are often induced by reports from alarmist media. He suggests steps to diffuse the panic by taking measures to protect one's privacy. However, some of these might be outdated in today's online environments, like the PrivacyBird tool [12]. Buta's work does not explore people's actual lived experiences of privacy-related panic situations. Our investigations attempt to fill this gap.

2.2 Potential panic evoking privacy incidents

Scenarios related to regretting sharing something on online social networks have been studied by Wang et al. [63], who examined some of the causes for regretting sharing posts on Facebook and reported on the repercussions that certain posts can have on people's lives. Simiarly, Sleeper et al. [55] explored users' sharing regrets while using Twitter as opposed to regrets during in-person conversations. They found that most people regretted sending messages on Twitter if the message revealed too much information or if it expressed criticism towards the recipient, and that most of regretted messages were posted at a time when users were in a highly emotional negative state. Other studies have looked at the concerns of sharing something with unintended audiences and its consequences to the users' reputation [6, 29, 57, 61].

Losing control of one's reputation online due to the content posted by others is also a potential trigger for sudden feelings of fear. Appropriately managing one's reputation is a serious challenge not only because content, once posted on the Internet, becomes very hard to remove, but also because harm to one's reputation online can translate into irreversible real-world problems, such as limiting job opportunities or damaging social relationships. Woodruff [65] has looked at the strategies people take when realizing that content about them has appeared online. Participants in her study expressed feelings of disempowerment when trying to remedy their damaged reputation, also stating that managing one's online reputation is a burdensome and unpleasant task, yet it is necessary. Madden & Smith [34] also explore how people have adapted their behaviours to control privacy settings in social networks and manage their online identities as measures to protect their reputation.

Lampinen et al. [29] also recognize the difficulty of managing the divide between privacy and publicness. The authors report on the issue that arise when third-parties turn one's private content public, thereby revealing too much of others' lives, for example, uploading someone else's picture into a social network site and the feelings of powerlessness this can create. These are actions that, although they might not carry long term consequences to the first-party's reputation, can cause embarrassment and shame. When looking at third-party sharing at a bigger scale, Rader [43] has studied people's concerns with regards to behavioural tracking commonly performed by online services, suggesting that (even) people who are aware of so-called *first party* behavioural tracking are less aware of third-party aggregation of their data and hence are more prone to be concerned about unwanted access. Similarly, Sipior et al. [54] summarized users' concerns with regards to web tracking technologies, and found that users are unaware of web tracking and that they feel exposed and traspassed when they find out about it. Moreover, a report on transparency-enhancing tools also stated that emotions of astonishment, surprise and distress were evoked in non-savvy users by the realization that online companies collect information and analyze data about them which is more than what they have explicitly consented to disclose [3].

Data being handled, shared and aggregated at different services may lead to higher probabilities for the occurrence of identity theft, profiling and other attacks. Identity theft is a major cause of concern due to the long term repercussions that can linger in the victims' lives [5], as well as to the costs to society and to the individual. A 2009 survey reported that people in the U.S. worry much more about being victims of identity theft than home burglary, getting their car stolen or terrorism [45]. Anderson [2] has examined the likelihood of becoming a victim of identity theft based on the person's demographic characteristics, concluding that people with higher incomes, women, adults living alone in the household have higher risks of getting their identity stolen. As a way to combat identity theft, Lai and Hsieh [28] propose a framework for studying the factors that influence people in adopting identity protection strategies.

Many cases of identity theft occur after an intruder successfully infiltrates or takes control of the accounts of an individual at one or many online services. Thus, the event of having an account hacked or hijacked can be another important trigger of users' panic. Shay et al. [49] surveyed to understand users' attitudes and experiences with account hijacking. The researchers found that around 30% of their survey participants had been victims of some form of account hijacking. Compromised accounts were usually valuable, and the incident had a deep emotional impact on the victims. Beyond academic studies, narrations from people who got their account hijacked and the consequences of their misfortune are not hard to find in blogs, news and social media. One famous story was told by technology journalist Matt Honan [22] who narrated his moment of panic when he realized that access to his iPhone and iCloud accounts were being blocked. Security flaws on Amazon and AppleID systems allowed a hacker to get access to his Twitter account, with the purpose of controlling Mat's convenient Twitter user name (*@mat*). Matt's documents, family photos, emails were deleted in the process.

Matt's story is connected to the findings reported by Ion et al. [24] which indicate that people are still sceptical about moving their personal data into the cloud for fear that their files would be leaked, compromised or inaccessible, and because of their uncertainty on how their files are being accessed and used by other parties. Similarly, Clark et al. [11] showed the mismatch between users expectations about what they are sharing in the cloud and the real disclosures they make.

Losing valuable data located on the cloud because of its inaccessibility or malfunction of the service provider can also be a source of annoyance and/or fear, similar to losing one's laptop or mobile device. Surveys reveal that around 30% of people in Canada, the US and the U.K., have experienced the theft or loss of a mobile device, and owners of lost devices often do not have a locking mechanism on their device [17, 58]. Even with recent apps that help users locate their missing smartphones, such as 'Find my iPhone', 'Android Lost' or 'Android Device Manager', large number of devices are reported lost or stolen. Tu Z. & Yuan Y. [60] suggest a behavioural study on the risks and users' coping measures in the event of losing their mobile devices or getting them stolen. They claim that adopting concepts from Protection Motivation Theory can help understand the users' active or passive reactions in case of falling victims to this event.

Some privacy-related news as portrayed by the media can also be a source of panic. Security incidents reported in recent years, such as the Heartbleed bug [15] or Shellshock [19], created a state of momentary fear and concern specially among users who might not understand all the consequences to their privacy and who might get alarmed by the way the media portrays the incident. The revelations made by Edward Snowden starting in 2013 about the governments' surveillance initiatives, were a reason for making people worried about their government spying on them [32].

News about corporations being victims of hacking attacks are also frequently being reported by the media. Each attack can carry risks of leaking the services' customer personal data to the public, such as home addresses, credit card information, personal habits and more. As an example, the attack in 2014 to the messaging service Snapchat and to Apple's cloud service iCloud, leaked a great amount of sensitive personal images on the Internet [38]. The increasing number of attacks to corporations are documented and advertised in online databases, such as `http://datalossdb.org/`, `http://osvdb.org/` and `http://www.privacyrights.org/data-breach`, which tend to be collaboratively maintained by groups of privacy and security advocates.

Surely, many other types of incidents exist than the ones summarized in this section. Taking some of this related work as the base of our study, our intention is to unveil incidents – and the context of these incidents – that are commonly experienced by Internet users and which potentially evoke sudden feelings related to panic and fear.

3. METHODOLOGY

In this paper we report on the results from two data collection activities. First, we carried out a series of user interviews with 16 participants with the purpose of collecting stories of privacy-related incidents, as well as their concerns to become victims to other similar incidents. Then, we validated and complemented the findings from the interviews by analyzing the submissions of 549 respondents to a survey. The following subsections describe the goals, design and administration of these activities.

3.1 Interviews

We started our investigations by laying out a list of common scenarios that can potentially trigger privacy-related panic. These scenarios were identified though a combination of our own experience working with various privacy challenges, discussions with colleagues experts in the field of privacy, common incidents reported by the media, and the available research literature, some of which was presented in Section 2.2. Table 1 presents descriptions of these 12 initially identified scenarios, which ranged from misfortunes such of identity theft, online stalking or threatening, or losing one's mobile device, to more common cases of sharing something online with the wrong audiences or regret sharing it, as well as others. Complete descriptions of these 12 cases can be seen in Appendix B.1.

3.1.1 Screener

Based on the identified scenarios and other questions about users' concerns and previous studies on privacy incidents, we designed a set of questions with the purpose of screening for

Table 1: Twelve initially identified triggers of privacy panic

Code	Panic scenario	Description
CSC	Changes in my social context	Realizing that someone who I used to be closed with, but whom I no longer trust (e.g., ex-partner, previous employer, old friend) still has access to my accounts or my personal information
DLK	My data was leaked online	Finding out that my personal data has being leaked or obtained by someone who I do not approve of
DLO	I deleted my data or I am not able to access it	Deleting or not being able to access my online data or data in an account that is valuable to me, such as documents, pictures, or other online files
HIJ	My account was hijacked or hacked	Finding out that someone has hacked my account(s) or accessed it without my permission or knowledge
IDT	My identity was stolen or misused	Finding out that someone else is using my identity and personal information to pretend to be me on the Internet
LMD	My mobile device was lost or stolen	Losing a mobile device (like a smartphone or tablet) or getting it stolen
MED	I saw an alert on the news or media	Finding out through the news and media that my privacy or personal data can be at risk
MSR	I regret having shared something online	Sharing something on the Internet and regret sharing it once it's too late
MSV	I shared content with the wrong people	Sharing something on the Internet and realizing that it can be seen by the wrong person or group of people
REP	My reputation was damaged	Someone else posting things or spreading rumours about me on the Internet which may damage my reputation privately or professionally
STK	I was being stalked, threatened or bullied	Feeling uncomfortable because someone seems to be stalking me, threatening, bullying or bothering me on the Internet
TPS	Third-parties shared data about me	Finding out that another person or a company has shared my personal information with others or posted information about me on the Internet without consulting me first

participants that could be invited for a one-on-one interview. After asking interested respondents for demographic information, the screener used 5-point Likert-scale statements to measure respondent's general privacy concerns and their concerns to fall victims for the initially identified panic scenarios.

The screener was distributed to a pool of about 600 people from approximately 15 different country locations who are registered on a platform that was built by the company where one of the authors works. The platform allows for people to voluntarily sign-up to participate in user studies. In the period of one week, we received a total of 128 responses. From these, 29 were female (1 didn't state gender), 78 had some form of employment, 35 were students, and 15 were not employed or retired. The majority (79) were between the ages of 24 and 40 years old.

Respondents to the screener also were encouraged to share similar stories of privacy-related panic that they had experienced. This allowed us to corroborate some of our initially identified scenarios. For instance, one respondent told how his reputation was at risk when "*someone was defaming my wife and me anonymously to my employer and friends*" (TP073), which can be seen as damage to his reputation (REP). Another respondent wrote that "*While I was in a relation with a previous partner, she had access to all my information and accounts as we had trust within each other. Eventually when the relationship ended I didn't feel it was appropriate to ask her to forget that information or delete it, as I thought it was common sense. Weeks later I noticed that someone besides myself was logging in to my accounts, changing information, reading my messages and so on*" (TP034), which is an example of the previously identified scenario of the person changing his social circumstances (CSC).

Table 2: Interview participants demographics

ID	G	Age	Location	Nationality	Profession	Privacy
TP008	M	18-23	Switzerland	Mexico	Engineering student	3.83
TP026	F	24-30	Germany	Poland	Sales representative	3.83
TP027	M	24-30	Slovakia	Slovakia	Financial advisor	2.33
TP030	F	24-30	UK	Malaysia	Medical doctor	4.00
TP053	F	31-40	Switzerland	USA	Business consultant	1.17
TP065	F	31-40	Switzerland	Switzerland	Secretary	1.67
TP069	M	31-40	Germany	Pakistan	Media enterpreneur	4.00
TP076	F	41-50	Portugal	Portugal	Web content creator	4.33
TP082	M	41-50	Switzerland	Switzerland	Security officer	1.83
TP085	F	41-50	UK	Caribbean	eCommerce director	4.00
TP089	M	51-60	Canada	Canada	Law enforcement	2.33
TP091	M	51-60	Switzerland	Switzerland	Civil engineer	3.00
TP093	M	51-60	Switzerland	Switzerland	Airport host	2.00
TP096	M	60+	Switzerland	Switzerland	Software support	3.17
TP097	M	24-30	Switzerland	India	Neuroscience student	4.17
TP105	M	24-30	UK	Italy	Retail store manager	5.00
TP056	M	31-40	Switzerland	India	Store assistant	2.67

3.1.2 Participants

We analyzed all responses of the screener trying to identify a group of about 15 to 20 participants for an additional one-on-one interview. We wanted to select a balanced group of ages, genders, location, familiarity with technology, and in particular people who we thought had a privacy panic story to tell. At the end, 17 out of the 128 respondents to the screener were invited to participate in an interview. The demographics of the selected participants along with their privacy concern score are shown in Table 2. Eight of the interviews were conducted remotely using a video conferencing system with screen-sharing capabilities. To 9 participants who were in reasonable distance to our location in Switzerland we extended invitations for a face-to-face interview. With one no-show (TP056), we had a total of 16 interviews.

Each interview session lasted around one hour and participants were compensated with the equivalent of a 50€ gift card for their time. In order to make use of the time more effectively, and as a way to refresh their memory with a privacy panic incident that they had recently experienced, each selected participant was asked to answer a short pre-questionnaire one or two days before the interview session took place. Questions asked about the context of their reported panic situation, the details of which were expanded upon during the interview.

An interview protocol (Appendix A.1) was prepared to maintain consistency across all sessions. The sessions were audio recorded after obtaining consent from participants, and recordings where later transcribed. To analyze the collected data, we used an inductive approach [59], which allows for simple and quick analysis of raw text data in order to extract relevant themes from the transcribed responses. Each test session was divided into two parts; the first part focused on participant's narrations of privacy incidents that they had experienced, and in the second part they gave their opinion about initial metaphors for a plausible "privacy panic button" that could help them in similar panic situations. Details of these are presented in the following sections.

3.1.3 Part 1 - Narrations of privacy incidents

In the first part, after building rapport with participants and introducing them to the study, they were asked to narrate any privacy incident that came to their mind which had caused them to experience feelings of discomfort. In particular, we started by asking the open question "*Have you ever been in a situation where you have felt a sudden feeling of panic, anxiety or distress for something related to your personal information or privacy on the Internet?*"

Participants reported different incidents and concerns that they either experienced in the past or that constantly preoccupied them in their daily Internet activities. Their responses broadly confirmed the panic scenarios that we had identified in the beginning. This was not surprising since we had primed them with similar situations. However, the collected responses helped us to understand each of the scenarios better, and provided a more complete angle to our initial views of these identified incidents. For instance, participant (TP065) told us that she took rigorous precautions to control the information that she or her family uploaded to the Internet, and recalled her feeling of panic when she suddenly realized that pictures of her daughter had been uploaded into a popular social network site by her own mother: "*I was shocked, because I am trying so hard not to post anything about her [my daughter], and then I see my mother [posting them]... It is just the feeling that you don't have total control of your data, [or] your family ...*". The participant in this case expressed her concern about the scenario in which family or friends may put her own privacy at risk without her been able to control it (TPS). Initially, we had considered third-parties to be service providers which shared the users' information, but the story from this participant helped us realize that third-parties could also refer to other people, and not only to online services. Table 7 of Appendix C lists the panic scenarios mapped to relevant quotes from interview participants who had experienced them.

We identified at least three important themes from other stories from the participant who we interviewed. For one, when we asked participants about the approaches they took at solving their situation, their stories suggested that many technical solutions to prevent or remedy privacy incidents might already exist and may be offered by service providers. However, there is no systematic and straight-forward way for lay users to find and appropriately utilize such available solutions in their particular situation.

Second, similar stories also suggested that approaches taken to solve the problem are very dependent on the type of incident, the context in which it occurs and the persons' familiarity with technology. For instance, participant TP008, who was tech-savvy, scored medium-high in privacy concern, and who had experienced the scenario of having his mobile device stolen, took a number of steps to protect his privacy afterwards: "*The first thing I did was to block the phone, disconnect from Whatsapp ... I went into the device manager section in Google and took that device from my list of devices. I assumed that the account would be disconnected [from the phone] and no synching would occur*". Additionally he setup a PIN code remotely using the AndroidLost app, and was even able to fetch pictures of the thief and his location using this same app. He went to the police with this information, and he cancelled his credit cards. On the other hand, a less tech-savvy participant (TP082) who scored low on privacy concern and whose email account had been compromised, recounted how he, while in panic, went on to search on the Internet for the symptoms of his problem, decided to buy Norton antivirus, stopped using his email account and opened a new one, and at the end hired – what he called – a 'PC doctor' who, after charging him tons of money talked him into throwing his computer away and buying a new one.

Third, when participants were asked about the root of their concerns that emerged from the incident ("*Why was this a big concern?*"), their responses indicated that the focus of their worries had to do with financial risks – "*The biggest concern are money and bank account access...*" (TP027), "*A lot of the bank stuff comes through your email, [I'm] afraid that money would go missing*" (TP030) – and possible damages to their reputation – "*[I am afraid that] something embarrassing would go out to my friends or my colleagues*" (TP082). In a minor scale, some participants also mentioned the burden of going through unnecessary or unexpected efforts – "*You don't only realize what [mistake] you have done, but also what you need to do after that*" (TP076) – or the consequences for their future – "*to be honest I'm afraid of what is on the Internet when I look for jobs*" (TP027).

Furthermore, a common reaction to a privacy threat was for participants to change their passwords for one or several of their accounts. This was seen as a solution that was easy to perform, understood by all, and had an immediate effect. Although changing an account password might not be a solution that fits all panic scenarios (i.e., if an account got hacked because hardware was compromised with a keylogger), the act of changing the password was perceived as a security-enhancing behaviour, which although it can make the person calmer and might take her out of panic mode, it can also give a false sense of security in some situations.

3.1.4 Part 2 - Metaphors towards a "panic button" help system

In the second part of the interview, participants were presented with the hypothetical scenario of an online service provider offering a feature analogous to a physical "panic button", but which would tender to privacy or security emergencies online. Specifically, we asked interviewees to tell us 1) what would they expect such a panic button to do in their panic situation and 2) where would they expect it to be located.

Table 3 summarizes the coded responses from these two questions (multiple answers were possible). Table 8 of Appendix C presents sample quotes extracted for the first question. Regarding their expectations of what the panic but-

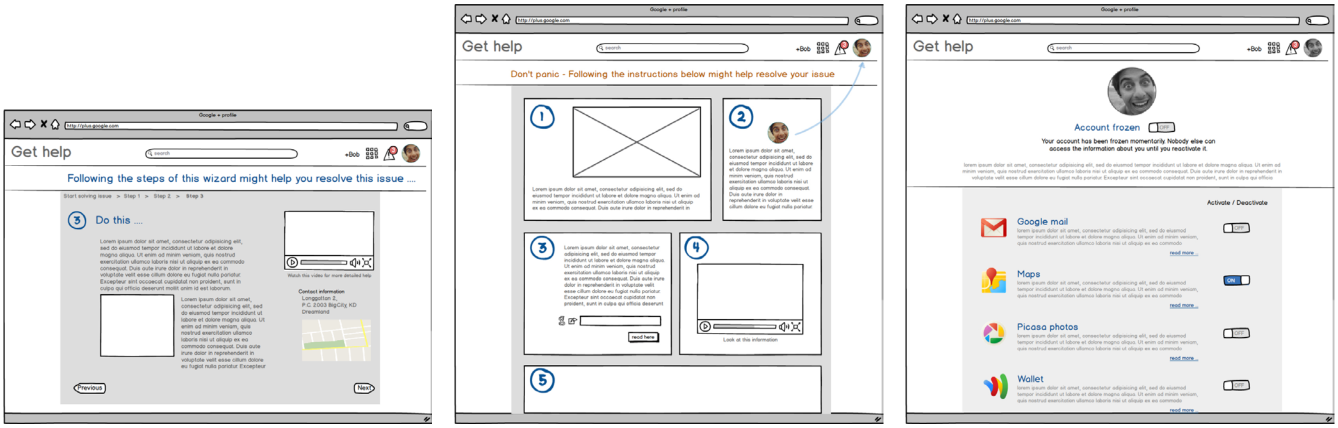
Table 3: Coded results of the expectations of what a panic button should do and how would it be accessed

What would it do?	
Personalized chat / Immediate	6
Give me instructions	3
Freeze or block accounts	3
Lead me through steps	2
Determine my problem	2
Assess the consequences	2
Verify my identity	2
Get me out of panic	1
Educate me (information)	1

(a)

How would you access it?	
Profile	4
Settings	4
Contextual	4
Search	2
Privacy / ToS	1
Mobile device	1

(b)


Google + profile
http://plus.google.com
Get help
search
+Bob
Following the steps of this wizard might help you resolve this issue ....
Start solving issue > Step 1 > Step 2 > Step 3
3 Do this ....
Watch this video for more detailed help
Contact information
Previous
Next
Don't panic - Following the instructions below might help resolve your issue
1
2
3
4
5
read here
Look at this information
Account frozen
OFF
Your account has been frozen momentarily. Nobody else can access the information about you until you reactivate it.
Activate / Deactivate
Google mail
Maps
Picasa photos
Wallet
read more ...
ON

(a) Wizard (b) Emergency card (c) Account freezing

Figure 1: Three alternative sketches demonstrating possible solutions once the 'panic button' is pressed

ton should do, many participants expressed that, once they have pressed it, they would like to receive some kind of personalized contact with someone, in part because this might provide them an *immediate* response to their problem, but also because it would make them "*feel more safe*" (TP082). Some participants thought of the option of freezing or blocking their accounts, as a *breathing moment* where they could be certain that no more harm can be done to their accounts or their privacy, and others mentioned the possibility of getting instructions or a list of steps that carry them towards a solution.

From their responses to the second question, we learned that many participants expected the button to be located somewhere within their profile pages or under the services' settings. Also, many implied that a help button should be available or connected to the factor that caused the panic in the first place, in other words, it should be contextual. For instance, if the user is notified of an attempt to access her account from another country, then let her launch the help process (i.e., press the panic button) from the notification itself. Although only mentioned by two participants, we believe the possibility to start the help process from the search results as a valuable idea, since many people start looking for help by 'Googling' for their problem (e.g., a search engine could detect that the user might have a privacy issue, and offer actionable help within the search results pages).

Participants were then shown sketched ideas of three metaphors that we had considered during the initial design process. The sketches, shown in Figure 1, were not at all representations of existing or planned designs, but rather just instruments to encourage discussions with the interviewed participants. The three metaphor ideas we used for providing help consisted of 1) a wizard (Figure 1a), showing a series of steps displayed one at the time which leads users towards possible solutions, 2) an emergency card (Figure 1b), similar to the cards found in an airplane's seat, listing the proper measures to take in case of an emergency, and 3) account freezing (Figure 1c), a mechanism to 'freeze' or block an account, so that no further actions are allowed until it becomes unfrozen.

We presented each of these, one at a time but in random order between participants, and asked them whether they thought that the idea was a good approach towards a solution in their case. Their extracted responses from the transcribed scripts are shown in Table 9 of Appendix C. For the wizard approach, participants suggested that it should consist of a series of simple questions and visual elements, it shouldn't contain too many steps and should have a few words at each step, since "*when you are in panic the last thing you want is to read*" (TP030). About the emergency card approach, participants appreciated having a complete overview of the possibles steps they ought to follow to reach a solution, noting also that the instructions given need to be simple, clear and straight forward. Regarding the possibility to freeze the account momentarily, participants recognized that the feature is analogous to the security offered by their banks when, for example, suspicious activity is detected in the account or when a credit card is reported lost. In general, participants found this feature reassuring or comforting, commenting things like "*it would make me feel a little better*" (TP093) and "*it makes me feel safe*" (TP026). Some participants also made observations about the considerations that need to be made before freezing someone's account, such as secure ways to unfreeze it, and the social implications of having their account blocked, "*Imagine that I am waiting for this urgent mail from a client [and I cannot access my email anymore]*" (TP076).

After looking at the three options, participants were asked their opinion on which of the three metaphors would be most suitable for helping people in similar panic situations. Their opinions were not mutually exclusive, and 14 out of the 17 interviewed participants favoured the possibility to freeze their account(s), 5 liked the idea of an emergecy card approach and only 2 said that they would prefer to be led through steps in a wizard-like fashion. Other ideas that came out at this stage of the interview included things like getting help from trustworthy or knowledgeable people, providing chat support, providing information about the scope of the problem, using media (images, videos) to instruct on possible solutions, and letting the user undo certain actions.

After obtaining this feedback from participants and making our own reflections, we concluded that a combination of the different approaches would be suitable for the different stages of the panic help process. For instance, as an immediate first step, users can be given the choice of momentarily ‘freezing’ their account, making them feel at ease and experiencing that something has been done to stop further harm. Then, a wizard with few simple multiple choice questions could be used to help determine the problem that the user is experiencing. Lastly, a series of possible tailored solutions or actionable steps could be suggested using an overview layout inspired by in-flight emergency cards.

3.2 Verification survey

The results obtained from the interviews gave us a good initial idea of the reasons and contexts for privacy panic, and the approaches that people take towards trying to solve their problems. In order to confirm and, if necessary, complement those findings we created and launched an electronic survey to collect additional privacy panic experiences from a different cohort of participants. In this section we describe the process of constructing the survey, its distribution and the obtained results.

3.2.1 Survey design and considerations

In the survey we wanted users to provide us with their first memorable experience of privacy panic[1], without us hinting in any of the privacy panic scenarios that we had previously identified, shown in Table 1. Similarly, we didn't want to force all respondents to tell us about a privacy incident if they could not think of any bad experience that had happened to them with regards to their privacy on the Internet. We thus started the survey by asking the same open question as we did at the beginning of the interviews: *Have you ever experienced sudden feelings of concern, anxiety and/or stress about something that happened to you on the Internet, related to your privacy or your personal information?*

If respondents indicated that they had been in such a situation they would proceed to the first section of the survey, where they were first asked to tell their story of the privacy incident and other questions around it (all questions are presented in Appendix B). Then they would continue to the second section of the survey, which asked questions surrounding the different incidents that we identified earlier (Table 1), as well as their concerns to become a victim of each of these incidents and other privacy concerns in general. Respondents who didn't have a privacy panic experience in mind were taken directly to the second section of the survey.

Three of the questions in the first section were open-text questions. In these, participants briefly narrated their story of the privacy incident, the way they found out that something was wrong or out of the ordinary, and the approach they took to try to fix their problem. For these three questions, two independent coders categorized each answer, using the findings from interviews as the bases for the category buckets, but adding additional categories if needed. If there were discrepancies in the category chosen by the two coders, the opinion of a third coder acted as a tie-breaker. The entry was considered a mismatch if there was disagreement between the opinions of all three coders. Cohen's kappa inter-rater reliability for the three questions was calculated, all of them suggested substantial agreement between coders ($kappa = 0.647, \rho < 0.001$, $kappa = 0.782, \rho < 0.001$, $kappa = 0.775, \rho < 0.001$, correspondingly). When categorizing the responses, the coders were instructed to read between the lines to extract the essence of the reason for panic or concern. For instance respondent SP923 narrated the following story: “*I once accidentally clicked on a spam ad that then downloaded spyware on my computer without my knowledge. After running a security sweep a few weeks later the software was detected and deleted however, I am still concerned that my personal information may have been stolen.*” Although the description narrates the infection of the respondent's computer, the underlying concern had to do with the possibility of her data being stolen or leaked.

Table 4: Survey respondents' demographics

		($n = 549$)
Gender	Male	72.1%
	Female	27.0%
	Rather not say	0.9%
Age	18 - 24	38.6%
	25-34	39.0
	35- 45	15.8%
	45-55	1.1%
	55-65	5.1%
	65+	0.4%
Occupation	Employed	41.7%
	Student	26.4%
	Not employed	9.1%
	Self-employed	8.6%
	Retired	0.4%
	Other	13.8%
Industry	Not at all technical	27.0%
	Not too technical	23.0%
	Somewhat technical	13.5%
	Very technical	28.2%
	Missing	8.4%
Crowdsourcing	Microworkers.com	79.8%
	CrowdFlower	15.0%
	ProlificAcademic	5.3%

At the end of the survey we collected some demographics from respondents and information about the crowdsourcing platform that referred them to the survey.

3.2.2 Survey administration and participants

Before launching the survey, two rounds of pilot sessions were carried out where we obtained feedback from a total of 16 participants. The final version of the survey was distributed using three different crowdsourcing platforms: Microworkers.com ($n = 438$), CrowdFlower ($n = 82$) and ProlificAcademic ($n = 29$). We received a total of 830 responses to the survey from these different recruitment platforms, from which 549 of these responses were kept after rigorously disqualifying entries which seemed rushed, incomplete, irrelevant, inappropriate or with very poor language. Responses came from different parts of the world, but the majority of respondents were located in India (31.1%), the United States (11.7%) and the United Kingdom (7.1%). Table 4 shows some additional demographic characteristics of our sample of respondents. For their participation, respondents were paid between \$0.50 and \$1.90 depending on the crowdsourcing platform they used. The survey took an average of 20 minutes for respondents who narrated a story and 14 for those who didn't.

3.2.3 Privacy panic scenarios

From the 549 valid responses to the survey 313 (57%) indicated that they had experienced sudden feelings of concern, anxiety or stress with regards to their privacy or personal information on the Internet. The responses from 5 partici-

[1] We avoided using the word *panic* throughout the survey, since the word in itself might sound too alarming. Instead we talked about *concerns* or *incidents*, as can be seen in Appendix B

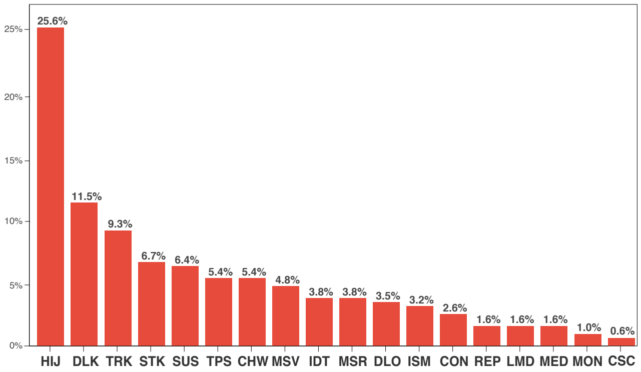


Figure 2: Percentages of self-reported panic stories according to the categorization of survey responses

pants were coded as ambiguous or unclear, and we were left with 308 privacy panic stories. We noted that slightly more than half of the respondents remembered, right on the spot, a personal incident that provoked privacy panic and were willing to tell it. Thus, we considered these as *memorable* cases.

Analysis of the responses from people who indicated to have a privacy incident to tell, yielded for 6 additional scenarios of privacy panic on top of the 12 identified at the beginning. Table 5 shows the newly identified six panic scenarios that resulted from coding the panic stories. These include cases of attempts to be tricked (TRK, e.g., falling for phishing attacks, Nigerian prince emails, buy/sell scams, etc.), realizing that a device has been infected by malware (CHW), realizing that your account is being monitored (MON, e.g., checks on network traffic, someone monitoring the activity in my computer), not having appropriate security measures (ISM, e.g., forgetting to log off from a public computer), getting stressed because of misunderstandings of technology (CON, e.g., confusions with data flows in the cloud and among multiple devices), or becoming aware of suspicious activity in one's account (SUS).

Appendix B.1 lists the characteristics of all 18 identified scenarios. In this list we present the proportion of people who experienced each scenario, their common approaches to solve their issue, their concerns about falling victims for the scenarios, as well as academic literature that has studied related privacy incidents and example quotes extracted from our data.

According to the coding of the self-narrated privacy panic stories, it can be seen that the event of having an account hacked or hijacked was by far one of the most memorable of all cases (HIJ, 25.56%), followed by stories of personal data being leaked (DLK, 11.5%) and attempts to be tricked (TRK, 9.27%). Figure 3.2.3 shows a bar graph of the 18 identified incidents ordered by the frequency of the coded stories of privacy panic. A chi-square analysis comparing stories from Indian and U.S. participants for the three most prominent cases yielded no statistical significant difference between these groups.

Many people who suffered a privacy panic incident indicated that a social network service (48.4%), a payment service (15.8%) and/or a messaging service (15.0%) were involved in the incident. Also, most incidents occurred when using either a desktop (50%), a laptop computer (41.4%) and/or a mobile device (31.2%). Few incidents happened with gaming consoles, wearables or other (5.3%).

Table 5: Six added panic scenarios after the classification made from the responses to the survey

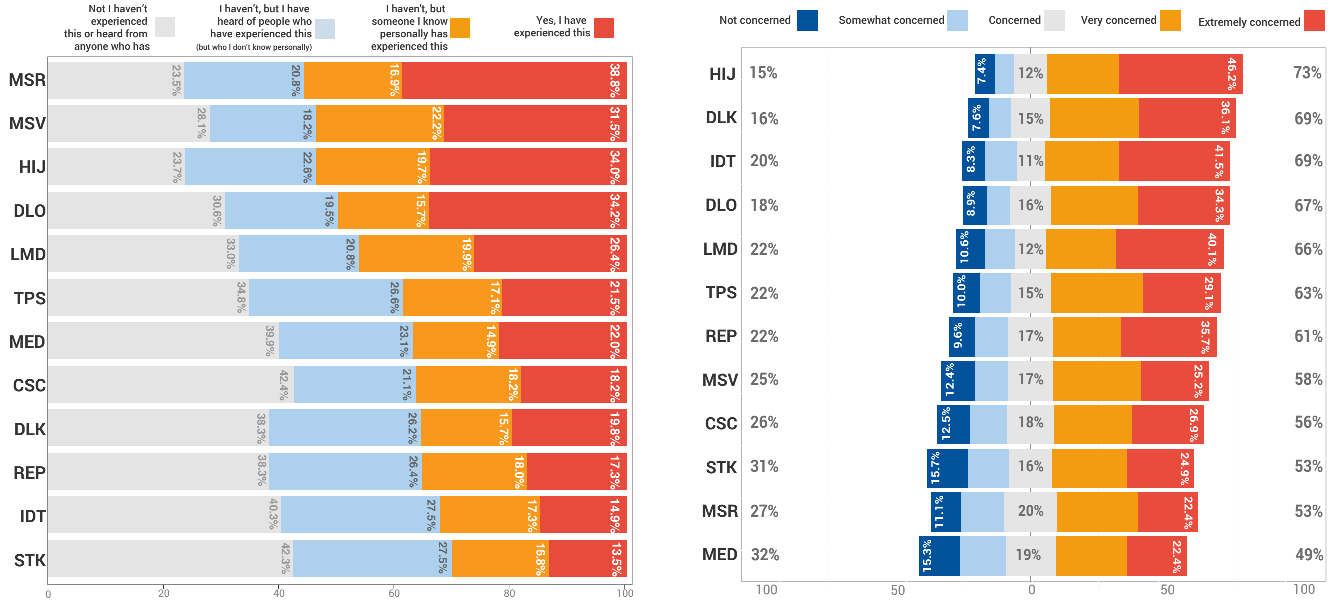
Code	Panic scenario	Description
CHW	My device became infected or compromised	My device (mobile, laptop, desktop, web camera, etc.) has been infected with a virus or malware
CON	Managing all my data and connected devices is stressful	Realizing that my identity or private information is at risk because I find it hard to understand and keep track of all the data exchanges between all my connected devices or Internet services
ISM	I didn't take appropriate measures to secure my account	Realizing that I neglected to take appropriate measures to protect my account or my personal data, which resulted in a breach which could have been avoided
MON	Someone else is monitoring my account	Being suspicious or realizing that someone else is monitoring my account or devices, or looking at my Internet activity
SUS	There was some suspicious activity in my account	Being notified that there was an attempt to access my accounts(s) or to obtain my personal information, or that unusual suspicious activity that I do not recognize has been happening in my account(s)
TRK	An attempt to trick me or defraud me was made	Nearly becoming a victim of fraud, someone trying to trick me or making me believe that a service was secure when it really wasn't

Respondents also reflected on how the incident that they experienced might have also indirectly affected their close friends and family. In other words, there might be occasions in which the effects of a privacy incident are not contained within the main victim. One respondent who had his account hacked commented that the event had repercussions on his family, since they couldn't spent time with him due to the extra work hours he had to put to solve the issue.

In the next section of the survey all respondents, including those who did not tell their panic stories, got to indicate if they, or someone they know, had experienced one of the twelve panic scenarios identified earlier, described in Table 1. This question was asked to encourage respondents who did not tell a specific story to remember and reflect over possible privacy incidents from the list that they might have experienced. The aim was to get an idea of frequency in which these scenarios are experienced by users. Figure 3a shows that many participants indicated having directly experienced the cases of regretting to share content online (MSR, 38.80%), sharing content by mistake with the wrong person or group of people (MSV, 31.50%) as well as having their accounts hacked or hijacked (HIJ, 34.06%) or losing access to their data (DLO, 34.24%). However, when it comes to *rumors of privacy panic* (i.e., hearing panic stories that happened to others), the cases of stalking or threatening (STK, 27.50%), identity theft (IDT, 27.50%) and third-party sharing (TPS, 26.20%) came at the top.

To test for reliability of our results, we looked back at the answers obtained in the initial screener to the same questions. Except for the case of stalking or threatening, a series of chi-square tests for all the other initially identified panic scenarios revealed no statistically significant differences between the screener and the survey samples with regards to the proportions of respondents who had personally experienced each of the initial privacy panic scenarios ($\rho \geq 0.05, \alpha = 95\%$, for all cases except STK, $\rho = 0.045$). Again, a Mann-Whitney U test between Indian and U.S. revealed no significant differences between the cultures.

3.2.4 Respondents' concerns



(a) Experiences with the identified panic scenarios

(b) Indicated levels of concern about the identified panic scenarios

Figure 3: Results from survey respondents

All respondents in our survey were asked to rate their level of concern about falling victims for each of the identified incidents, as well as their concerns of other privacy statements. We adapted two different instruments to measure privacy concerns as suggested by Anton et al. [4] and Buchanan et al. [7], explained in Section 2.1. From the latter, we took only the top four questions with higher factor loading. Additionally, we included our own set of six questions which we believed to be more relevant for our study and to today's technologies, as seen in Appendix B. Results from a factor analysis revealed that our set of questions fit together with the four selected questions from the scale suggested in [7]. Nine out of ten of these questions correlated at least with a factor of .380, a Kaiser-Meyer-Olkin measure of sampling adequacy was .918, and the Barltlett's test of sphericity was found to be significant ($\chi^2(105) = 2735.117, \rho < 0.001$). Hence an average privacy concern score was calculated from the values of these nine questions ($\mu = 3.26, std = 0.924$). A test of normality of this score revealed that our sample was slightly skewed towards privacy concerned respondents ($Kolmogorov - Smirnov = 0.046, Shapiro - Wilk = 0.978; \rho < 0.001$).

A Mann-Whitney U test of the privacy concerns score between people who told a story of privacy panic and those who stated that they had not experienced any such situation, shows that there was a significant difference in the level of stated privacy concern between these two groups ($U = 28295.0, \rho < 0.001$). One possible reason for this difference is that having been a victim of a privacy incident can have an impact on people's privacy concerns online, possibly modifying their behaviours to become more privacy aware and cautious. This is consistent with the arguments in [14] and also supported by many of the responses from our interview participants, which indicate that they tended to use online services in a different way, stopped using certain services or became more weary about the risks that can be encountered on the Internet. It was also noted from the collected responses that there exists a very small but significantly positive increase in the respondents' privacy concerns depending on the amount of privacy panic scenarios that they had directly experienced ($r = 0.182, \rho < 0.001$). Moreover, a nonparametric Spearman's rank order test revealed that there is a significant positive relationship between the reported level of concerns about the privacy statements and the respondents' concerns of becoming victims of the privacy scenarios presented to them ($\rho_s = 0.681, n = 544, \rho < 0.0001$), meaning that individuals who are more concerned about their privacy in general will tend to be more concerned about experiencing the identified privacy panic scenarios.

From the 12 scenarios presented, respondents in our sample indicated that they were *very* or *extremely concerned* about having their account hacked or hijacked (HIJ, 73%), realizing that their data has been leaked online (DLK, 69%), having their identity stolen (IDT, 69%), or losing their online data (DLO, 67%), while they were least concerned about hearing something from the news or media (MED, 32%), being stalked or threaten (STK, 31%) or regretting sharing something online (MSR, 27%). Curiously, these last case also appeared to be the one that most people mentioned having experienced personally at some point of their lives.

Analyzing the sample from our initial screener ($n = 128$) shows that the cases of MSR and MSV where at the bottom of the participants' concerns, yet they occurred most frequently. One possible reason for this seemingly paradoxical attitude could originate from the individuals' perception on how much control they have to remedy or reverse the situation. In other words, the event of sharing something publicly by mistake or having shared something that they later regret can be something that people perceive as "un-

Table 6: Ultimate concerns or reasons for worrying

Reasons for panic	n	%
People knowing things that are not of their business	81	25.88%
Embarrassment or damage to my reputation	73	23.32%
Money going missing or financial harm	68	21.73%
Emotional harm to me or someone close to me	51	16.29%
Physical harm to me or someone close to me	12	3.83%
Possible loss of physical property or something valuable	11	3.51%
Possible loss of my employment	6	1.92%
Other	11	3.52%

doable" or easily correctable, whereas the misfortune of having their identity stolen, having their information leaked or their account hacked, are cases that are seen outside their reachable control, thus raising their levels of concern about such situations.

Similar to our interviews, we wanted to get an understanding about why people worry when experiencing, or being near experiencing, a breach to their privacy. In other words, what do users see as the consequences of a privacy incident and what is the impact on their lives. To find out, we asked the following question to survey respondents who had a panic story to tell: *In the situation you described, which of the following options best represents your ultimate concern or reason for worrying?* Respondents chose one out of seven options which reflected common concerns identified in the interviews, and the proportions of their responses are presented in Table 6. Consistent with the results from the interviews, financial harm and embarrassment or damage to their reputation were among the top of their worries. Above all, survey respondents' ultimate concern had to do with third parties knowing things about them which are not of these third parties' business.

From the table, it can be noted that around 65.5% of people are concerned by things that are 'softer' types of harm that deal with concepts that are hard to quantify, such as emotional harm, reputation or nosiness of others. On the other hand, 31% expressed concerns related to more concrete and measurable worries, such as losing of money, one's employment or material valuables. This suggests that users in a panic situation could be informed through a user interface about a quantified estimate of the impact of the incident in terms of value, which can motivate them to take steps to resolve the issue and enhance the protection of their privacy.

3.3 Limitations

We are aware that the methods of data collection we employed in our study are only recollections of previous privacy incidents. People might forget other important privacy panic experiences or not recall every instance of what really happened and how they went about solving it. Nevertheless our approach gave us a good starting point for understanding such situations, which can later be studied in-depth with methods that capture users' everyday experiences.

Interview participants were recruited based on a convenience sample of people who voluntarily signed up to participate in user studies. This limitation was one of the reasons that drove us to verify our results with an additional survey. Since due to our location, we were not able to employ the services from Amazon's Mechanical Turk (MTurk), the recruitment of the survey participants was done through three other crowdsourcing platforms: Microworkers, Crowdflower and ProlificAcademic. Although we did not find specific studies about the quality of the workers in the platforms we used, we employed three different ones based on findings from a recent study which indicate that different sample providers might provide differences in their variances with regards to privacy measurements [47]. From our sample Microworkers' participants were slightly older and had higher proportions of Southeast Asians, whereas the proportion of female participants in Crowdflower was slightly higher. However, a Kruskal-Wallis H test showed that there is no statistical significant difference of the levels of privacy concerns among the three platforms $\chi(2) = 2.145, \rho = 0.342$, ($\rho > 0.5$, for all cases).

4. IMPLICATIONS FOR THE DESIGN OF A PRIVACY PANIC HELP SYSTEM

After obtaining a better understanding of possible reasons of privacy panic in users, we can now describe a list of implications for the design of a system of that could try to help users in these situations.

Our findings suggest that some types of panic scenarios that occur most often are not necessarily the same as the ones that concern users the most. However, the case of having one's account hacked or hijacked (HIJ) was one that appeared at the top of users concerns and also was frequently experienced personally, which indicates the need for providing users with more education on how to protect their accounts against this type of attacks, and for prioritizing solutions and remediations for this panic scenario. When presenting users with possible solutions to their privacy panic moment, an intelligent help system could try to detect the type of scenario that the user is experiencing and investigate the users' main concerns. Our investigations showed that many users' concerns have to do with their finances, their reputation or other people knowing things that are not of their business. Thus informing users about the consequences of the incident could help them understand what is at risk, hopefully relieving some of their panic.

From the analysis of the stories told by interview participants, it was observed that people with different levels of familiarity with technology tend to approach a privacy panic problem in different ways. Tech-savvy users are often aware of possible solutions to their particular problem and they just need to have those solutions more accessible, without the need for lengthy instructions or the feeling that they are being patronized. On the other hand, non-tech savvy users could feel lost on where or how to start looking for a solution. An interface for a privacy panic help system should cater for these different users, offering expert users with direct calls for action and convenience on the steps to quickly reach a solution, whereas non-experts could be directed with easy-to-follow steps to solve specific problems, and information about how to protect oneself against similar scenarios in the future. Further studies can help determine the specific needs of different types of users in panic situations.

Regarding the type of help that users expect in a panic situation, users would like to have the possibility to contact someone directly, this is specially the case for non-expert users. Contacting someone could imply that the system facilitates users with communication to their acquaintances, other more expert users, or customer support. Ideas for *crowdsourced* help have been presented in [52]. Many users also like the idea of freezing their accounts to stop further damage and alleviate the feelings of panic. However, explanations on what does freezing mean and the mechanisms to recover the account have to be explicit and clear. From a technical perspective, a potential feature to 'freeze' or block an account momentarily could be very difficult to achieve with todays technologies, possibly carrying many security implications and potential for abuse. Although technology should try to address these challenges to meet users' expectations, detailed further investigations of such feature are necessary before it should be implemented.

Users expect to find help within the context in which they realized that a problem existed. For instance, changing the visibility properties of a regretted post on a social network should be possible from within the same view that is displaying the post. Nowadays, many users employ search engines as a way to look for their problem, and we value the possibility of displaying help actions directly within the view showing the search results. For instance, Google's Knowledge Graph [53], displays cards on a side panel whenever a concept is identified in the search query, and similar panels could be presented for searches related to privacy panic scenarios. However, further research is required to identify appropriate search keywords. The contributions of Chilana et al. [9], which propose methods for improving the ways users find appropriate help, could also be considered for these type of scenarios.

In general, users experiencing privacy panic expressed their need for a system that provides them with actions, and not complicated instructions that are lengthy and without assurance that a solution will be reached. A help system has to give users the feeling that something is being done to protect their privacy, and that every step has a purpose.

With these considerations in mind, we identify five characteristics of a system that provides help to users in privacy panic situations:

- **Actionable**: Do not redirect users to other pages where they might (or might not) find help. Let them instead perform in-place, situated actions that are perceived as effective steps towards protecting their data and privacy. For instance, if an effective solution for their case is to change their password, the system should allow them to perform that action right where they are, and do not redirect them to another page.

- **Immediate**: Users in panic expect help quickly, not only because the attack or ongoing harm should be stopped as soon as possible, but also because users can perceive that an efficient and trustworthy service provider should be able to provide them with quick and effective help.

- **Adaptive**: A help system should cater to the different type of users (e.g., experts, vs. non-experts), and adapt to the various types of contexts of these users, as well as the different types of panic scenarios that they might be experiencing.

- **Reassuring**: Depending on the situation and the concerns of users, some users might experience more or less panic than the situation calls for. Providing users with possible scope of the consequences to their privacy, as well as with statements of comfort and reassurance might help users understand the problem better and alleviate their panic more effectively.

- **Preventive:** Users that have experienced a privacy incident should understand why it happened and ways to avoid it from happening again in the future. The system should not only try to help users resolve their problem, but also educate them, facilitate steps to secure their account and encourage them to continue adopting secure behaviours.

While more evidence might be needed to determine concrete design suggestions for a possible help system, the results from this study suggest that the help process to aid users experiencing privacy panic should follow the following guidelines: first, let users take as immediate and easily-applicable actions to protect their account(s) as possible, to stop further harm or blocks access to their disclosures; second, try to identify the user's problem or narrow it down by asking a series of straight-forward questions; and third, present users with actionable, in-place solutions that will try to return things to normal, or even improve the protection of the users data and/or privacy, while at the same time educating users on secure behaviours and on ways to prevent similar events in the future.

5. FINAL REMARKS

We presented our exploratory study into moments of online privacy panic. We identified and ranked 18 situations in which users might experience feelings of concern, distress or panic with regards to their privacy or their personal information online. We presented contextual factors around these situations and we also explored, with the use of a panic button metaphor, users' expectations of a possible help system for these kinds of panic situations. At the end, we introduced implications for the design of such a system.

Although our findings unveil 18 panic scenarios that are valid in today's online ecosystem, we see the need to continue investigating and discovering similar situations of privacy panic that may arise with people's evolving privacy attitudes and concerns and with the emergence of newer technologies, such as wearable devices, ubiquitous sensors surrounding the users' environments, intelligent machines, and others.

Acknowledgments

We want to deeply thank Manya Sleeper and Jonathan Aroner for their contributions to ideas and design of the study. Also, Erik Wästlud and Henrik Anderson at Karlstad University for their help with the analysis of the data. Special thanks to the Privacy UX team and other colleagues at Google for their support and feedback.

6. REFERENCES

[1] H. Almuhimedi, A. P. Felt, R. W. Reeder, and S. Consolvo. Your reputation precedes you: History, reputation, and the chrome malware warning. In *Proceedings of the Symposium on Usable Privacy and Security*, SOUPS '14, pages 113–128, Menlo Park, CA, USA,, July 2014. ACM.

[2] K. B. Anderson. Who are the victims of identity theft? The effect of demographics. *Journal of Public Policy & Marketing*, 25(2):160–171, 2006.

[3] J. Angulo, S. Fischer-Hübner, J. S. Pettersson, and M. T. Jessica Edbom. D:C-7.3 Report on end-user perceptions of privacy-enhancing transparency and accountability. Project deliverable D:C-7.3, A4Cloud Project, September 2014.

[4] A. I. Antón, J. B. Earp, and J. D. Young. How internet users' privacy concerns have evolved since 2002. *Security & Privacy*, 8(1):21–27, 2010.

[5] H. Berghel. Identity theft, social security numbers, and the web. *Communications of the ACM*, 43(2):17–21, 2000.

[6] A. Besmer and H. Lipford. Tagged photos: Concerns, perceptions, and protections. In *Extended Abstracts on Human Factors in Computing Systems*, CHI '09, pages 4585–4590. ACM, 2009.

[7] T. Buchanan, C. Paine, A. N. Joinson, and U.-D. Reips. Development of measures of online privacy concern and protection for use on the internet. *Journal of the American Society for Information Science and Technology*, 58(2):157–165, 2007.

[8] P. Buta. *Privacy Panic – How to Avoid Identity Theft, Stop Spam and Take Control of Your Personal Privacy*. Hillcrest Publishing Group, 2009.

[9] P. K. Chilana, A. J. Ko, and J. O. Wobbrock. Lemonaid: Selection-based crowdsourced contextual help for web applicationselection-based crowdsourced contextual help for web applications. In *Proceedings of the Conference on Human Factors in Computing Systems*, CHI '12, pages 1549–1558. ACM, 2012.

[10] H. Cho, J.-S. Lee, and S. Chung. Optimistic bias about online privacy risks: Testing the moderating effects of perceived controllability and prior experience. *Computers in Human Behavior*, 26(5):987–995, 2010.

[11] J. W. Clark, P. Snyder, D. McCoy, and C. Kanich. I saw images I didn't even know I had: Understanding user perceptions of cloud storage privacy. In *Proceedings of the Conference on Human Factors in Computing Systems*, CHI '15, pages 1641–1644. ACM, 2015.

[12] L. F. Cranor, P. Guduru, and M. Arjula. User interfaces for privacy agents. *ACM Transactions on Computer-Human Interaction*, 13(2):135–178, June 2006.

[13] J. W. DeCew. *In pursuit of privacy: Law, ethics, and the rise of technology*. Cornell University Press, 1997.

[14] T. Donaldson and T. W. Dunfee. Toward a unified conception of business ethics: Integrative social contracts theory. *Academy of management review*, 19(2):252–284, 1994.

[15] Z. Durumeric, J. Kasten, D. Adrian, J. A. Halderman, M. Bailey, F. Li, N. Weaver, J. Amann, J. Beekman, M. Payer, et al. The matter of heartbleed. In *Proceedings of the Conference on Internet Measurement*, CIM '14, pages 475–488. ACM, 2014.

[16] J. B. Earp, A. I. Antón, L. Aiman-Smith, and W. H. Stufflebeam. Examining internet privacy policies within the context of user privacy values. *IEEE Transactions on Engineering Management*, 52(2):227–237, 2005.

[17] S. Egelman, S. Jain, R. S. Portnoff, K. Liao, S. Consolvo, and D. Wagner. Are you ready to lock? In *Proceedings of the Conference on Computer and Communications Security*, pages 750–761, 2014.

[18] S. Egelman, A. Oates, and S. Krishnamurthi. Oops, I did it again: mitigating repeated access control errors on facebookdid it again: mitigating repeated access control errors on facebook. In *Proceedings of the Conference on Human Factors in Computing Systems*, CHI '11, pages 2295–2304. ACM, 2011.

[19] T. Fox-Brewster. What is the Shellshock bug? Is it worse than Heartbleed? *The Guardian*, September 25 2014.

[20] T. Halevi, N. Memon, and O. Nov. Spear-phishing in the wild: A real-world study of personality, phishing self-efficacy and vulnerability to spear-phishing attacks. *Social Science Research Network*, January 2015.

[21] Harris Interactive. First major post-9/11 privacy survey finds consumers demanding companies do more to protect privacy, February 2002.

[22] M. Honan. How Apple and Amazon security flaws led to my epic hacking. *wired.com*, 6, August 2012.

[23] W. Hong and J. Y. Thong. Internet privacy concerns: an integrated conceptualization and four empirical studies. *MIS Quarterly*, 37(1):275–298, 2013.

[24] I. Ion, N. Sachdeva, P. Kumaraguru, and S. Čapkun. Home is safer than the cloud! Privacy concerns for consumer cloud storage. In *Proceedings of the Symposium on Usable Privacy and Security*, SOUPS '11, pages 13:1–13:20, Pittsburgh, PA, USA, 2011. ACM.

[25] P. Klasnja, S. Consolvo, J. Jung, B. M. Greenstein, L. LeGrand, P. Powledge, and D. Wetherall. When I am on wi-fi, I am fearless: Privacy concerns and practices in everyday wi-fi use. In *Proceedings of the Conference on Human Factors in Computing Systems*, CHI '09, pages 1993–2002. ACM, 2009.

[26] R. M. Kowalski, S. Limber, S. P. Limber, and P. W. Agatston. *Cyberbullying: Bullying in the digital age*. John Wiley & Sons, 2012.

[27] P. Kumaraguru and L. F. Cranor. Privacy indexes: A survey of westin's studies. *Institute for Software Research International*, 2005.

[28] F. Lai, D. Li, and C.-T. Hsieh. Fighting identity theft: The coping perspective. *Decision Support Systems*, 52(2):353–363, 2012.

[29] A. Lampinen, V. Lehtinen, A. Lehmuskallio, and S. Tamminen. We're in it together: Interpersonal management of disclosure in social network services. In *Proceedings of the Conference on Human Factors in Computing Systems*, CHI '11, pages 3217–3226. ACM, 2011.

[30] E. Litt, E. Spottswood, J. Birnholtz, J. T. Hancock, M. E. Smith, and L. Reynolds. Awkward encounters of an "other" kind: Collective self-presentation and face threat on Facebook. In *Proceedings of the Conference on Computer Supported Cooperative Work & Social Computing*, CSCW '14, pages 449–460. ACM, 2014.

[31] Y. Liu, K. P. Gummadi, B. Krishnamurthy, and A. Mislove. Analyzing facebook privacy settings: user expectations vs. reality. In *Proceedings of the Conference on Internet Measurement*, CIM '11, pages 61–70. ACM, 2011.

[32] D. Lyon. Surveillance, Snowden, and big data: capacities, consequences, critique. *Big Data & Society*, 1(2), 2014.

[33] S. Machida, T. Kajiyama, S. Shigeru, and I. Echizen. Analysis of facebook friends using disclosure level. In *Proceedings of International Conference on Intelligent Information Hiding and Multimedia Signal Processing*, IIH-MSP '14, pages 471–474. IEEE, 2014.

[34] M. Madden and A. Smith. Reputation management and social media. 2010.

[35] M. Madejski, M. Johnson, and S. M. Bellovin. A study of privacy settings errors in an online social network. In *International Conference on Pervasive Computing and Communications Workshops*, PERCOM '12, pages 340–345. IEEE, 2012.

[36] N. K. Malhotra, S. S. Kim, and J. Agarwal. Internet users' information privacy concerns (IUIPC): The construct, the scale, and a causal model. *Information Systems Research*, 15(4):336–355, 2004.

[37] H. Mao, X. Shuai, and A. Kapadia. Loose tweets: an analysis of privacy leaks on twitter. In *Proceedings of the Workshop on Privacy in the Electronic Society*, WPES '11, pages 1–12. ACM, 2011.

[38] L. O'Connor. Celebrity nude photo leak: Just one more reminder that privacy does not exist online and legally, there's not much we can do about it. *Digital Commons: The Legal Scholarship Repository @ Golden Gate University School of Law*, October 2014.

[39] C. Paine, U.-D. Reips, S. Stieger, A. Joinson, and T. Buchanan. Internet users' perceptions of 'privacy concerns' and 'privacy actions'. *International Journal of Human-Computer Studies*, 65(6):526–536, 2007.

[40] L. Palen and P. Dourish. Unpacking privacy for a networked world. In *Proceedings of the Conference on Human Factors in Computing Systems*, CHI '03, pages 129–136. ACM, 2003.

[41] F. Parry. True crime online: Shocking stories of scamming, stalking, murder and mayhem. *The Electronic Library*, 32(2):279–280, 2014.

[42] R. M. Peters. So you've been notified, now what? the problem with current data-breach notification laws. *Arizona Law Review*, 56:4, 2014.

[43] E. Rader. Awareness of behavioral tracking and information privacy concern in facebook and google. In *Proceedings of the Symposium on Usable Privacy and Security*, SOUPS '14, Menlo Park, CA, USA, July 2014. ACM.

[44] M. Ryan. Cloud computing privacy concerns on our doorstep. *Communications of the ACM*, 54(1), 2011.

[45] L. Saad. Two in three americans worry about identity theft. *Gallup Poll Briefing*, 1, 2009.

[46] M. B. Schmidt and K. P. Arnett. Spyware: a little knowledge is a wonderful thing. *Communications of the ACM*, 48(8):67–70, 2005.

[47] S. Schnorf, A. Sedley, M. Ortlieb, and A. Woodruff. A comparison of six sample providers regarding online privacy benchmarks. In *Proceedings of the Workshop on Privacy Personas and Segmentation (PPS). Symposium On Usable Privacy and Security*, SOUPS '14, Menlo Park, CA, USA, July 2014. ACM.

[48] U. Shankar and C. Karlof. Doppelganger: Better browser privacy without the bother. In *Proceedings of the Conference on Computer and Communications Security*, CCS '06, pages 154–167. ACM, 2006.

[49] R. Shay, I. Ion, R. W. Reeder, and S. Consolvo. My religious aunt asked why i was trying to sell her viagra: experiences with account hijacking. In *Proceedings of the Conference on Human Factors in Computing Systems*, CHI '14, pages 2657–2666. ACM, 2014.

[50] S. Sheng, B. Magnien, P. Kumaraguru, A. Acquisti, L. F. Cranor, J. Hong, and E. Nunge. Anti-phishing phil: the design and evaluation of a game that teaches people not to fall for phish. In *Proceedings of the Symposium on Usable Privacy and Security*, SOUPS '07, pages 88–99, Pittsburgh, PA, USA, July 2007. ACM.

[51] I. Shklovski, S. D. Mainwaring, H. H. Skúladóttir, and H. Borgthorsson. Leakiness and creepiness in app space: Perceptions of privacy and mobile app use. In *Proceedings of the Conference on Human Factors in Computing Systems*, CHI '14, pages 2347–2356. ACM, 2014.

[52] V. Singh, M. B. Twidale, and D. Rathi. Open source technical support: A look at peer help-giving. In *Proceedings of the Hawaii International Conference on System Sciences*, volume 6 of *HICSS '06*, pages 118c–118c. IEEE, January 2006.

[53] A. Singhal. Introducing the knowledge graph: things, not strings. *Official Google Blog*, May 2012.

[54] J. C. Sipior, B. T. Ward, and R. A. Mendoza. Online privacy concerns associated with cookies, flash cookies, and web beacons. *Journal of Internet Commerce*, 10(1):1–16, 2011.

[55] M. Sleeper, J. Cranshaw, P. G. Kelley, B. Ur, A. Acquisti, L. F. Cranor, and N. Sadeh. I read my Twitter the next morning and was astonished: A conversational perspective on twitter regrets. In *Proceedings of the Conference on Human Factors in Computing Systems*, CHI '13, pages 3277–3286, Paris, France, July 2013. ACM.

[56] D. K. Smetters and N. Good. How users use access control. In *Proceedings of the Symposium on Usable Privacy and Security*, SOUPS '09, pages 1–12, Mountain View, CA, USA, 2009. ACM.

[57] F. Stutzman and J. Kramer-Duffield. Friends only: Examining a privacy-enhancing behavior in Facebook. In *Proceedings of the Conference on Human Factors in Computing Systems*, CHI '10, pages 1553–1562. ACM, 2010.

[58] Symantec Corporation. Internet security threat report. 19, April 2014.

[59] D. R. Thomas. A general inductive approach for analyzing qualitative evaluation data. *American journal of evaluation*, 27(2):237–246, 2006.

[60] Z. Tu and Y. Yuan. Understanding user's behaviors in coping with security threat of mobile devices loss and theft. In *Proceedings of the Hawaii International Conference on System Sciences*, HICSS '12, pages 1393–1402. IEEE, 2012.

[61] M. van Der Velden and K. El Emam. "Not all my friends need to know": A qualitative study of teenage patients, privacy, and social media. *Journal of the*

American Medical Informatics Association, 20(1):16–24, 2013.

[62] K. E. Vaniea, E. Rader, and R. Wash. Betrayed by updates: How negative experiences affect future security. In *Proceedings of the 32nd annual ACM conference on Human factors in computing systems*, CHI '14, pages 2671–2674, 2014.

[63] Y. Wang, G. Norcie, S. Komanduri, A. Acquisti, P. G. Leon, and L. F. Cranor. I regretted the minute i pressed share: A qualitative study of regrets on facebook. In *Proceedings of the Symposium on Usable Privacy and Security*, SOUPS '10, Redmon, WA, USA, 2011. ACM.

[64] J. Watson, A. Besmer, and H. R. Lipford. +Your circles: sharing behavior on Google+. In *Proceedings of the Symposium on Usable Privacy and Security*, SOUPS '12, pages 1–9, Washington, DC, USA, July 2012. ACM.

[65] A. Woodruff. Necessary, unpleasant, and disempowering: Reputation management in the internet age. In *Proceedings of the Conference on Human Factors in Computing Systems*, CHI '14, pages 149–158, Toronto, ON, Canada, 2014. ACM.

[66] A. Woodruff, V. Pihur, S. Consolvo, L. Schmidt, L. Brandimarte, and A. Acquisti. Would a privacy fundamentalist sell their DNA for $1000... if nothing bad happened as a result? The Westin categories, behavioral intentions, and consequences. In *Proceedings of the Symposium on Usable Privacy and Security*, SOUPS '14, Menlo Park, CA, USA, July 2014. ACM.

[67] E. Zangerle and G. Specht. Sorry, I was hacked: A classification of compromised twitter accounts. In *Proceedings of the Symposium on Applied Computing*, SAC '14, pages 587–593. ACM, 2014.

APPENDIX

A. INTERVIEWS

A.1 Interview protocol

Introduction to test session

"*Hi, my name is ... The reason we invited you to participate in this study is because users have told us in previous studies that there might be moments in which they find themselves in distressful situations regarding the data about them that is on the Internet. We want to explore the way people react and handle such situations in order to be able to conceive a easy-to-use service that might help them resolve some of those issues. In order to do that in the best way, we would like to hear your stories about previous experiences when you might have felt panic, stress or anxiety about situations involving your personal data.*

When you were invited to participate in this interview, you were asked a series of questions, we might repeat some of those questions now so you can tell us more about those situations.

If it is ok with you we will record this session, mainly to have a record of what is said. There might be other colleagues observing our conversation that will help me take some notes or that are interested on what we have to say. [Start recording at this point].

What will happen is that I will ask you some questions and we will mostly have a conversation revolving those questions. It is important to keep in mind that we are not testing you or your knowledge. There are no right or wrong answers, and we just want you to narrate your stories as accurately and honestly as possible."

A.1.1 Open-ended questions and dialogues

1. **(10 minutes)** Have you ever been in a situation where you have felt a sudden feeling of panic, anxiety or distress for something related to your personal information or data on the Internet?
   (a) How long ago did this happen?
   (b) How did you find out that the issue was going on?
   (c) Do you remember how did you feel at that moment?
   (d) Why was this a big concern?
   (e) What do you think led to the situation in the first place?
   (f) Did you resolve the issue? How?
   (g) Did you try to seek help some how?
2. **(10 minutes)** From your responses to the screener and pre-questionnaire, you mentioned that [AN INCIDENT] happened to you or to someone you know. Can you tell me more about [AN INCIDENT]?
3. **(10–15 minutes)** We are investigating ways to design functionality similar to a so-called "*panic button*", which can help users solve issues similar to the ones you just described.
   □ Imagine that it is you who discover that there's something not normal, where would you click on a service to access such button?
   □ Imagine that a service provider detects that there is something not normal with your account. For instance it can recognize that your account is being access from another part of the world. What would be the best way to notify you?
4. **(5 - 10 minutes)** Imagine that a service provider is offering a panic button and that you are in a situation similar to the one you described earlier. In your opinion, what should happen when you press such a panic button?
5. **(10 - 15 minutes)** We have been thinking of some possible solutions to offer help once you have clicked on the panic button, which of these different options (Figure 1) do you think would be more helpful for you at solving your problem during a panic situation?
6. **(5 min)** Do you think that you would have used it [panic button] in the situation when the incident you described happened?

B. VERIFICATION SURVEY QUESTIONS

(Page 1) Introduction and consent

"*Thank you for your interest in completing this survey.*

The results of this survey will be used for a project at Karlstad University in Sweden. At no time your name or any other information that directly identifies you will be shared with anyone outside the University. Your answers, along with the answers of many other people, will be analysed and might be reported in academic conferences and scientific venues. The results of our research might be used by our research partner to improve their products and provide their users with a better experience. You have the right to contact us and request that your responses will not be considered in scientific publications as long as your request is received before the results are made public.

Please remember that there are no wrong or right answers to the questions in this survey, we just want your honest opinions and answers to the questions asked. We will not collect your name, email address or other specific information that identifies you directly, so that you feel comfortable answering the questions honestly.

In order to start the survey you need to agree to the terms listed here. This survey consists of around 25 - 30 questions and will take approximately 10 - 20 minutes. You can use a computer or a mobile device to complete the survey. At the end of the survey you can give your opinion on things that were not clear or hard to understand."

(Page 2) Your story about a privacy incidents

There are moments when you might have become concerned that your online privacy might be at risk or when you realised that something was not quite right with your online personal information.
We want to know your story about one of those moments!
Please try to remember such a situation as well as you can and answer the following questions in as much detail as possible.

1. **Have you ever experienced sudden feelings of concern, anxiety and/or stress about something that happened to you on the Internet, related to your privacy or your personal information?**
   ○ Yes, I have been in a situation where I experienced such feelings about my privacy or personal information on the Internet
   ○ Probably, but I do not remember any such situation right now
   ○ No, I haven't experienced any such situation

(Page 3) Your story about a privacy incident

2. **Please briefly describe the incident that made you experience feelings of concern, anxiety and/or stress about your privacy or your personal information on the Internet**

3. **Approximately how long ago did this happen?**

   ○ Less than a week ago

   ○ More than a week ago but less than a month ago

   ○ More than a month ago but less than six month ago

   ○ More than six months ago but less than a year ago

   ○ More than a year ago

   ○ I don't remember

4. **How did you realize that something was wrong or out of the ordinary?**

5. **Which type of technology device(s) were involved in the incident that you described?**

   □ Laptop computer □ Tablet □ Wearable technology □ Smartphone □ Desktop computer □ Mobile phone (non-smartphone) □ TV or gaming console □ Other

6. **What did you do to try to repair the problem(s) caused by the incident that you described?**

7. **Were there any Internet services, mobile apps or companies that were involved in the incident?**

   □ Transport □ Messaging □ Photo □ Entertainment □ Dating □ Media streaming □ Government □ Location □ Payment □ Cloud storage □ Social networks □ Online bank □ Travel □ Other

8. **As far as you are aware, was someone else also affected by the incident that you described?**

9. **In the incident that you described, which of the following best represents your ultimate concern or reason for worrying?**

   ○ Possible loss of physical property or something valuable

   ○ Embarrassment

   ○ Damage of reputation

   ○ Possible loss of my employment

   ○ Physical harm to me or someone close to me

   ○ Emotional harm to me or someone close to me

   ○ People knowing things about me that are not of their business

   ○ Money going missing

   ○ Other

(Page 4) About your story

10. **Does the incident you described in the previous page relate to any of the situations listed below?**

    ○ Finding out that another person or a company has shared my personal information or posted information about me on the Internet

    ○ Feeling uncomfortable because someone seems to be stalking me, threatening or bothering me onthe Internet

    ○ Finding out that someone has hacked my account(s) or accessed it without my permission or knowledge

    ○ Finding out that my personal data has being leaked or obtained by someone who I do not approve of

    ○ Finding out that someone else is using my personal information to pretend to be me

    ○ Finding out through the news and media that my privacy or personal data can be at risk

    ○ Deleting or not being able to access my online data that is valuable to me, such as documents, pictures, or other files

    ○ Losing a mobile device or getting it stolen (such as a smartphone or tablet)

    ○ Someone who I no longer trust (e.g., ex-partner, previous employer, old friend) still has access to my accounts or my personal information

    ○ Someone posting things about me online which damage my reputation

    ○ Sharing something in a Social Network by mistake or regret sharing it

    ○ Sharing something in a Social Network and realizing that it can be seen by the wrong person or group of people

    ○ Other

11. **If you chose 'Other', how would you describe in ONE sentence the reason for which you experienced feelings of concern, anxiety or stress**

(Page 5) About your story

12. **Have you ever experienced any of the following situations?**

    Same list of options as previous question presented in randomized order, with the following options for each case:

    ○ Yes, I have experienced this

    ○ I haven't, but someone I know personally has experienced this

    ○ I haven't, but I have heard of people who have experienced this (but who I don't know personally)

    ○ No, I haven't experienced this or heard from anyone who has

(Page 6) About your concerns

13. **How concerned are you about the following situations happening to you?**

    Same list of options as previous question presented in randomized order, with the following options for each case:

    Not concerned ○—○—○—○—○ Extremely concerned

14. **Rate how much do you agree or disagree with the following statements**

    Based on the instrument on privacy concerns suggested in [4]. Order was randomized.

    Strongly disagree ○—○—○—○—○ Strongly agree

    ○ I want a web site to tell me how my personal information will be used

    ○ I am concerned about unauthorized employees getting access to my personal information

    ○ I mind when a web site monitors my purchasing patters

    ○ I mind when a website I visit collects information about my browsing patterns

    ○ I mind when my personal information is shared or sold to third parties

15. **Rate how concerned you are about the following statements**

    Based on the instrument on privacy concerns suggested in [7] plus questions specific to this survey. Order was randomized.

    Not concerned ○—○—○—○—○ Extremely concerned

    ○ Someone looking at the contents of my mobile device

    ○ Strangers looking at the things I post on the Internet

    ○ If you use your credit card to buy something on the internet your credit card number will obtained or intercepted by someone else

    ○ In general, how concerned are you about your privacy while you are using the Internet?

    ○ Online services not being who they claim they are

    ○ People online not being who they say they are

    ○ Advertisers using information about me to advertise to me

    ○ Companies sharing my personal information without my permission

    ○ The level of encryption of my data when I submit it to an Internet service

    ○ The amount of information on the Internet available about me

(Page 7) About you

16. **Gender**

17. **Age range**

18. **Nationality**

19. **Occupation**

20. **Industry**

21. **How would you rate yourself in this scale**

    I often ask others for help with technology (computer, phones, etc.)
    ○—○—○—○—○
    Others often ask me for help with technology (computer, phones, etc.)

22. **From which of the following services did you hear about this survey?**

    ○ Microworkers ○ Crowdflower ○ ProlificAcademic

B.1 Final identified categories of panic

My account was hijacked or hacked HIJ

Occurrences:

interviews	survey
3	80

Description: Finding out that someone has hacked my account(s) or accessed it without my permission or knowledge

Literature: [49] [22]

Sample cases:

SP229 "My e-mail account was hacked and embarrassing e-mails were sent to various people close to me including my employer"

SP324 "My email account was hacked and I have lost it because the hacker change all measure to change the account password"

SP508 "When I wasn't able to log into my Facebook account with an error message saying your account was logged in Burkina Faso a West African country"

Account hijacking (HIJ) happened to be the most prevalent reason for panic among survey respondents. From all the coded incidents submitted by survey participants ($n = 313$), 80 of them (25.60%) narrated a scenario related to the hacking or hijacking of one of their accounts. The victims of this incident found out that their accounts had been accessed by someone else mainly because of a warning given by the service provider (22%) or because they were suddenly unable to access their account (26%). For 20% of them it was their own realization or suspicious that led them to discover that their account had been hacked, for instance cases similar to "*I found out that someone accessed my email account outside of my country where I belong to*" (SP337) were often reported. To resolve this issue (50.2%) of the victims changed their password, (24.1%) contacted the corresponding service provider, and other few tried to update their security or privacy settings in some way (6.3%) or took the steps to recover their account (6.3%). The proportion of all respondents from our survey who stated having personally experienced a case of account hijacking ($33.8\%, n = 549$) significantly agrees with the proportion reported in the study of Shay et al. [49] ($30\%, n = 294$) ($z = 2.51, \rho < 0.01$).

My data was leaked online DLK

Occurrences:

interviews	survey
2	36

Description: Finding out that my personal data has being leaked or obtained by someone who I do not approve of

Literature: [24] [38] [51]

Sample cases:

SP717 "Few years ago I joined Internet casino because of the some bonus program. After that my e-mail is on spam attack almost every day"

SP425 "When Target had a breach of security, credit card numbers were stolen. Mine was among them"

Answers related to the scenario of **data leaked (DLK)** included the user realizing that her personal information has being stolen, finding out that someone else got a hold of the user's personal information by suspicious means, the user finding her personal information on the Internet when using a search engine, and other similar situations. In our sample, 36 (11.5%) out of 313 respondents declared having found out that their data had been leaked on the Internet. Most of them stated that they found out because they noted something on a website or service that awaken their suspicion. For instance, SP861 wrote that he noticed that his phone number had been leaked when he started receiving unsolicited calls. The majority of the victims of this case (12) mentioned that their reason for worrying was the possibility that someone would steal their money. As attempts to resolve their issue, these victims tended to contact the service provider directly or to change their password.

An attempt to trick me or defraud me was made TRK

Occurrences:

interviews	survey
n/a	29

Description: Nearly becoming a victim of fraud, someone trying to trick me or making me believe that a service was secure when it really wasn't

Literature: [50] [41] [20]

Sample cases:

SP246 "I was paying for something from a site. I already put in my credit card details and push submit when I realised that it was a fake site"

SP79 "I was navigating a website, can't remember what website, when my firewall flashed that someone was trying to gain access to my computer"

Many respondents to our survey reported becoming afraid, stressed and concerned when experiencing an **attempt to be tricked, defrauded or scammed (TRK)**. Even though the attacker or fraudster might not succeed in his attempt, such cases have the power to encourage victims to take action with regards to securing their data and possibly promoting behavioural change. Many of the stories in this category had to do with phishing attempts, specially targeted to financial online services or social networks. The majority of participants in this case stated that they tried to mitigate such attempt by closing abruptly their browser (SP093), turning off their computer at once, not using a particular service, or blocking the website that originated the attempt. For instance, SP697 mentioned that he hastily "*deleted that account*" when he was requested to enter his PayPal details in an unknown service. Most victims of this case (57.1%) mentioned that their biggest worry had to do with money going missing. Attempts to be tricked is the one of the incidents that the participants in our sample experienced most recently, since 17 out of the 29 respondents indicated that it occurred to them in the last 6 months.

I was being stalked, threatened or bullied STK

Occurrences:

interviews	survey
1	21

Description: Feeling uncomfortable because someone seems to be stalking me, threatening, bullying or bothering me on the Internet

Literature: [26] [41]

Sample cases:

SP554 "Few months ago I made a friend from South Africa on Facebook... One day she told me to send her my nude pics and I did it. After that she started blackmailing and demanding money and [threatening] me to post them on Facebook ... I unfriended her and decided to not make any close friend online"

SP380 "Someone got my info online and was telling me they will find me"

SP69 "Once I posted an article in the main daily papers and my e-mail address was also posted. After that, I started receiving threats into my inbox and was feeling bad and completely concerned about it"

This category comprises serious cases of **stalking or threatening (STK)** or recurring offenses of cyber-bullying that made users feel very uncomfortable or fearful that some greater, more tangible, harm could be done to them. These are often cases which may trigger many emotional reactions on their victims. One of our interview participants (TP053) who had been stalked by her ex-husband not long ago before the interview, and felt severely emotionally paranoid, was of the opinion that some of the security approaches adopted by many online services were an enabler for her bad experience. In our survey sample there were 21 of these cases of stalking, threatening or bullying online. Most of them tried to resolve the situation by either blocking a contact on their email or a social network or contacting a service provider for help. For instance, SP124 mention how she "*didn't respond to the man who threatened me... I deleted him from my Facebook and blocked his account.*" The majority of these victims indicated that their biggest concerned was that their reputation might be ruined due to blackmailing, or that some sort of emotional harm will be done to them or someone close to them.

There was some suspicious activity in my account SUS

Occurrences:

interviews	survey
n/a	20

Description: Being notified that there was an attempt to access my accounts(s) or to obtain my personal information, or that unusual suspicious activity that I do not recognize has been happening in my account(s)

Literature: [42]

Sample cases:

SP539 "I logged in to my Gmail account and there was a message written in red that my account had been accessed from an IP-address most likely located in China"

SP691 "My email provider informed me of 179 failed login attempts at an email address I had not logged into in over a month"

Suspicious activities (SUS) include, among others, attempts to be hacked which were actually stopped by the service provider. However, the sole fact that a notification was send to the user about an attempt to infiltrate her account can be a trigger of panic and a call for action about improving secure behaviours. Many of the respondents who reported having received a notification did actually changed their passwords or looked at their account history for signs of recurring suspicious activity. At the same time, a recent legal investigation into privacy breaches presented in [42] argues that a more actionable notification scheme should be considered for data breaching events. The responses of 20 people in our survey sample were categorized under this panic scenario, most of them fearing that a stranger would know things about them that were not of their business. Half of them were warned by the service provider or found out by the tools the service offers (e.g., account login history) about suspicious activity in their account.

My device became infected or compromised CHW

Occurrences:

interviews	survey
n/a	17

Description: My device (mobile, laptop, desktop, web camera, etc.) has been infected with a virus or malware

Literature: [46] [62]

Sample cases:

SP977 "I browsed some sites for adults and downloaded something - I don't know what. Unexpectedly, my tablet turned off, and I couldn't turn on them again"

SP164 "Downloaded a file which turned out to be a virus and my information was compromised"

SP493 "I was watching a video and my webcam took a picture of me and then put up a full page warning that unless I pay an amount of money out before the timer runs out the Police will be called"

Cases of **compromised hardware (CHW)** are an old common reason for panic in desktop computers, although recently other hardware devices are being infected, such as mobile phones. 17 people in our sample reported cases in which their devices got infected usually due to visiting a malicious website or downloading a suspicious file. At least 2 people reported having their mobile device infected, and 2 other people mentioned attacks via their compromised web cameras. Their biggest ultimate concern was that information or pictures about them were going to be used to embarrass them, but also matters related to money or strangers knowing things that are not of their business was a big concern. The majority of victims, tried to resolve their issue by restarting, formatting or scanning their device.

Third-parties shared data about me TPS

Occurrences:

interviews	survey
3	17

Description: Finding out that another person or a company has shared my personal information with others or posted information about me on the Internet without consulting me first

Literature: [29] [43] [54]

Sample cases:

SP60 "When people's morphed photos were uploaded in social networking sites ... including mine"

Cases of **third-party sharing (TPS)** are usually predecessors of reputation damage (REP), since people that panic because of someone else sharing data about them online is usually because the things being shared are embarrassing, untruthful or make the first party uncomfortable. In our sample 16 respondents told stories about other people or companies shared their data on their behalf. Six of them said that they tried to contact the third party to get the content taken down. Their concern had to do with they, or someone close to them, being harmed emotionally. One third of respondents for this scenario tried to either contact the third-party to try to resolve the issue directly or contact the service provider where the data was found. Others indicated a change on their behaviour, for instance, one participant explained how he found morphed photos of himself being uploaded to a social network site, and as a consequence he has never uploaded one more photo to a social network again.

I shared content with the wrong people MSV

Occurrences:

interviews	survey
1	15

Description: Sharing something on the Internet and realizing that it can be seen by the wrong person or group of people

Literature: [11] [18] [33] [35] [56] [64]

Sample cases:

SP202 "I once saved a sex story in Facebook under notes. Unfortunately, I posted it public instead of selecting the privacy option 'Only me'. I deleted it soon after but few of my friends asked me about that. I said my account got hacked."

SP450 "I had accidentally posted a personal status update on facebook without realising that the default audience was set to public instead of my usual strict filtering"

Sharing something with the wrong person or group of people (MSV) is one of the panic scenarios that many of our respondents reported having experienced (see Figure 3a), but only 15 of the respondents told the story of such a case. Eight of them claimed that they found out on their own about posting something with the wrong audience, for instance by receiving strange comments from other people. The big majority resolve their issue by changing their privacy settings on the service where they uploaded the content or by taking the content down. One participant reported that he stopped using Facebook groups in order to stop personal information from being revealed to those groups.

My identity was stolen or misused IDT

Occurrences:

interviews	survey
1	12

Description: Finding out that someone else is using my identity and personal information to pretend to be me on the Internet

Literature: [5] [45] [2] [28]

Sample cases:

SP313 "Someone copied my details from my social media account, along with some pictures, and posing as me."

SP59 "I found out that there existed a Facebook profile using my name and my profile picture"

SP704 "Sometime last year, a few people have told me that they've already accepted my invite through email to a private site, and they asked what it was all about, because there were only a few words in it and it seemed half finished. I've never made such a website and I panicked about someone accessing my account to send the invites to those people I knew."

There were 12 reported stories of **identity theft (IDT)** in our sample, which included attackers creating email accounts under the victims name, filling credit card applications with the use of the victim's information, and finding out that accounts have been created using the victim's pictures and other information. Five of them indicated that they found out that their identity had been stolen or misused because someone else told them about it, and six realized on their own due to weird activity in their accounts on online services or banks. Half of the victims of this case, contacted the service provider to report the abuse of their data or the observed strange activity, and three of them took steps to have fake accounts blocked. Their biggest concern had to do with emotional harm, money going missing or being embarrassed.

I regret having shared something online MSR

Occurrences:

interviews	survey
3	12

Description: Sharing something on the Internet and regret sharing it once it's too late

Literature: [55] [63]

Sample cases:

SP809	“When I was younger I took silly pictures of myself. I had a hard time purging them all from the internet”
SP1019	“I posted a picture of my cat on my Facebook page. It was only afterwards when I realized that I had left a marijuana pipe on the desk AND my eBay password was visible on a piece of paper in the picture”
SP1032	“I submitted a photo that had geotagging data on it to an anonymous website”

Out of all panic scenarios, **regretting sharing something online (MSR)** was the scenario that most respondents of our exploratory questionnaire and our survey admitted to had personally experienced (as seen in Figure 3a). However, it is also the case which people were concerned the least about being victims of. Logically, most people who regretted sharing something try to fix their problem by removing the content that was shared. From the 12 reported cases of regret, 7 stated that embarrassment was their biggest worry. One mentioned that he was worried about the lost of physical property, since he posted an item for sale online and was worried that others would get to know his phone, name, address and the valuable item that was at his home. Other cases included people uploading pictures of themselves that they later regretted, making public comments about the government, or people being teased due to the content of their public posts. One participant described the consequences of his regret when, after having uploaded a photo containing embedded location metadata to an anonymous website, other people started ordering pizza to his home address.

I deleted my data or I am not able to access it DLO

Occurrences:

interviews	survey
2	11

Description: Deleting or not being able to access my online data or data in an account that is valuable to me, such as documents, pictures, or other online files

Literature: [24]

Sample cases:

SP435	“My Gmail account was lost and that day I cried... Many personal information was included in that mail”
SP76	“Sometimes I feel stress when can't find an important information”

Cases of **losing access to data or an account (DLO)** included stories related to accounts being blocked, forgetting PIN codes or passwords, and not being able to find data that they presumably deleted by mistake. Many people with this problem tried to contact the service provider for help, or try to update their settings once they recover access to their account.

I didn't take appropriate measures to secure my account ISM

Occurrences:

interviews	survey
n/a	10

Description: Realizing that I neglected to take appropriate measures to protect my account or my personal data, which resulted in a breach which could have been avoided

Literature: [62]

Sample cases:

SP868	“I have logged in a recharging website from another person's computer. After that two days I did not get time to work in computer. I felt afraid that person may hack my password and take my money”
SP478	“I forgot to log out my Facebook account in a computer laboratory, then someone used my Facebook status to inform me that I forgot to log out”
SP39	“I made Paypal account and share my paypal address on the Internet and my password was too short and simple”

Failing to take **appropriate security measures (ISM)** can also result in panic. Examples of stories in this category include users forgetting to log off from an account and others taking the opportunity to post things in the users' behalf (i.e., faceraping), or realizing that their password is too weak or that the security in their account too is vulnerable. This category was made different from identity theft (IDT) or stalking, threatening or bullying (STK) in that the stories describe much lesser offenses, often inducted by friends or family who try to tease the victim but don't mean any great harm. Six out of the ten people who experienced this case stated that they changed their password after the incident. The majority mentioned that they worry about money going missing or public embarrassment.

Managing all my data and connected devices is stressful CON

Occurrences:

interviews	survey
n/a	8

Description: Realizing that my identity or private information is at risk because I find it hard to understand and keep track of all the data exchanges between all my connected devices or Internet services

Literature: [11] [24] [25] [31] [37] [44] [51]

Sample cases:

SP730	“[I panicked] when I added my personal number on the Internet such as Facebook because they need it to verify your account”
SP611	“I realized that Google for instance has everything about me connected... Having all of this out there in the hands of databases of companies created some sort of anxiety. I'm more careful being anonymous on the internet nowadays than I was a few years ago.”
SP828	“In Facebook, I've been watching some publicity about my sexual preferences that I'd prefer to keep private. So I don't know how they get that information”

We refer to difficulties of **managing connected devices and services (CON)** to situations when users find themselves doing something to breach their own privacy or experiencing feelings of stress simply because their lack of understanding on how technology works. For instance, respondents in this category reported being scared at the presence of tailored ads or the realization that service providers can infer information about them through big data analysis. When asked how did they find out that something was wrong or out of the ordinary, one respondent wrote that he “*created an account on Google plus using fake info for anonymity purposes, but then received friend requests from people I know from Facebook*” (SP611). He went on to mention that this unclear coupling of personal information across services keeps him concerned about using social network services. The victims of this use case didn't have any consistent approach to calm their panic. Six of them stated that their main concern lied in others knowing things about them that were not of their business and the remaining two were afraid of some physical harm happening to them.

My reputation was damaged REP

Occurrences:

interviews	survey
2	5

Description: Someone else posting things or spreading rumours about me on the Internet which may damage my reputation privately or professionally

Literature: [1] [6] [29] [30] [33] [34] [57] [61] [65]

Sample cases:

SP105	“Inappropriate photos were posted on some my accounts and fake links too”
SP631	“One of my friends had posted an embarrassing picture that featured me drinking. My family being extremely conservative objected to this vehemently as they did not like the party I was at”

Contrary to the case of third-party sharing, this category represents events where the victims' main concern is not on the fact that data about them has been shared online, but rather that their reputation can be severely damaged. In her study Woodruff [65] recounts the story of a manager who discovered that bad reviews about her were written by her colleague in an online service. Similarly, a participant of our interviews, told us how she panicked when a bad review was made in a popular travel website about an aspect of her business (TP085). Some of the respondents who experienced this panic scenario, try to solve it by contacting the person who uploaded content or trying to take it down themselves. One participant tried to solve his problem by restarting or scanning his computer for malware, since he explained how he was looking for adult content online, when “*in one second some window showed up, and started popping up again and again, saying that I'm sharing my searches with* *people on my Google+ and my Facebook account, where I was logged in... that window was asking for my personal information in order to stop sharing that stuff. I was very wared, and concerned, scared about idea that someone else or specially my friends and relatives, will see stuff that I was looking on the Internet*" (SP927).

I saw an alert on the news or media MED

Occurrences:	interviews	survey
	3	5

Description: Finding out through the news and media that my privacy or personal data can be at risk

Literature: [15] [32]

Sample cases:

SP214 "With all these companies being hacked I fear about my information being stolen"

SP680 "There was a thing going around on the Internet ... It was called the heart bleed virus or something, but I was terrified that all of my personal information was going to be hacked or spread. It was horrible. I stayed off the internet for like a week"

Every now and then **media scandals (MED)** can create a state of panic in their followers. Popular examples include Snowden's revelations about the goverments' surveillance, news about serious bugs, such as Heartbeat and Shellshock, or major data leaks, like the leakage of celebrity pictures through Snapchat or the hacked suffered by Sony in 2014. In our sample only 5 people reported a story dealing with such scandals. Besides finding out through the news, three of these cases indicated that they were warned about the incident by the service provider.

My mobile device was lost or stolen LMD

Occurrences:	interviews	survey
	2	5

Description: Losing a mobile device (like a smartphone or tablet) or getting it stolen

Literature: [17] [60]

Sample cases:

SP42 "I was casual and I never knew that I had lost my phone, later when I found out I was stunned, speechless. I realized that I lost one of my priced possession which I brought out of my pocket money"

SP860 "I lost my phone which contained personal information like passwords to email accounts, bank details and contacts. Plus, my browsers had saved passwords to various password protected sites"

The **lost of a mobile device (LMD)** has also become a big concern, given that plenty of personal information and data is stored in these devices and that they become a portal to our private information and many of our online accounts, in which we are perpetually logged in. Only very few respondents submitted stories related to the loss of their mobile device. However, this case was one of the cases at the top of the users' concerns. Recent approaches to secure mobile devices offered by the manufacturers and other apps, also have lower the frequency of this type of incident. For instance, Android offers the 'Android Device Manager' service and Apple has the 'Find my phone' feature, which makes it much harder for thieves to target these devices, and easier for the owners to reclaim them.

Someone else is monitoring my account MON

Occurrences:	interviews	survey
	n/a	3

Description: Being suspicious or realizing that someone else is monitoring my account or devices, or looking at my Internet activity

Literature: [48]

Sample cases:

SP22 "Got a feeling that someone is checking my browsing history and all"

SP342 "I am using my Facebook account for past 5 years .. when I came to know that my father is secretly monitoring my account I was very angry and I thought that I don't have online privacy"

Scenarios of **account monitoring (MON)** refer to cases in which the respondents might feel that the activity in their accounts or their online communications might be monitored by a third party. The four people in our sample who fell into this category told stories about finding out that family members are monitoring their activity, or simply getting the feeling that someone else is intercepting their online actions. Three out of the four people who experienced this case stated that mere suspicion made them find out that something might be going wrong.

Changes in my social context CSC

Occurrences:	interviews	survey
	1	2

Description: Realizing that someone who I used to be closed with, but whom I no longer trust (e.g., ex-partner, previous employer, old friend) still has access to my accounts or my personal information

Literature: [67]

Sample cases:

SP59 "my ex boyfriend published on Facebook some photos of us together while I was in another relationship"

SP609 "I used to date a girl, she was into games, like I am and she knew about all my usernames/passwords. When we decided to break up, it took a good time to make them all safe and I even got as far as losing some of them to her"

We refer to **changes in social contexts (CSC)** to the cases in which a person's social circumstances have changed and when the person realizes that some personal data or sensitive information was shared with other people who are no longer trustworthy, reliable or close. This can commonly occur when a romantic relation ends, when changing employers, or moving to different cities. In our survey there were only two such stories put under this category, and one interviewee (TP065) also recounted an episode of panic when she realized that her Google calendar was been shared with a person who was no longer her friend.

C. RESULTS FROM INTERVIEWS

Table 7: Identified reasons of privacy panic supported by quotes from interview participants

Code	Sample quotes	
CSC	TP065	"Someone else [an exfriend] having access to my calendar that I shared with him. I was a bit shocked that I forgot to remove him... It would be perfect for a stalker"
DLK	TP027	"Even though they assure me it is secure, there could be a little bug, that will leak out my data, and they can get my card number, address, phone and everything..."
	TP105	"I had my data on the Internet. All the details, they got all details, links to bank accounts..."
DLO	TP076	"I deleted my bookmarks on my old computer long. When I changed computers I exported bookmarks, but then I threw them away [deleted them]. I am concerned about deleting something unintentionally. So I safe my stuff in the Cloud. I trust more storing in the cloud"
	TP097	"Once my computer broke and I lost all my data... I was on the verge of freaking out"
HIJ	TP069	"When login from another country you are asked for the recovery of your email... The first time I got this notification, I thought 'eh?, what is happening'... when you are not familiar with something you will get panic. I thought my account is blocked or hijacked. After that I got a notification that we are trying to prevent an unauthorised login to your account...."
	TP082	"Someone sending emails on my behalf [from my email account]... I was really afraid and I didn't know what to do, so I called one of these PC doctors, really expensive. And at the end nobody could really help me."
IDT	TP093	"I was working as a journalist. Someone contacted me asking me if I wanted to be a member of this thing... I got suspicious... later I found out that someone used my information to get a journalist pass... I'm not sure how he found my information, but I think he found what I was writing"
LMD	TP008	"I got my phone stolen in the train ... I fell asleep and in a lapse of 10 minutes someone snatch my phone from my hand..."
	TP105	"I was vacationing in Gand Canaria and I lost my phone... It is a lot of work when something like that happens"
MED	TP093	"When heart bleed came out I changed all my passwords. I was worried mostly about my credit card.... I was very very worried about my amazon account leaked my bank account data.."
	TP065	"[Facebook] makes studies about people without telling them... I don't want them to do anything with my data without me knowing.... it was in the media some time ago"

(Table 7 continued...)

Code	Sample quotes	
MSR	TP027	"Once I posted a picture of a 5 star hotel were I was staying for a business trip and my colleagues got jealous... Rumours spread some weeks later, which affected me"
	TP069	"First day that Google plus is introduced suddenly my photos got uploaded"
MSV	TP085	"There was something that my daughter put on social media, I think she didn't do well on an exam ... I came across the post later and I had to explain to her why you shouldn't do that... because kids tend to have teachers as friends and everybody... I explained the fact that this is public and that the record stays there forever..."
REP	TP093	"Quite long ago my friends got into my account. I had a strong password, but an easy security question. This made me realize that security is important"
	TP085	"Gay couple experienced discrimination in a reservation they did through one of our hotels. They posted bad reviews about the hotel about the company being homophobic ... there was blind panic within the company. What do you do? How do you react to it?"
STK	TP053	"I used to trust my husband, but then we got divorced. He started stalking me Through bank transactions he knew what I was doing and where I was going."
TPS	TP089	"linkage [of my identity] exists and that it is outside my control, I find that worrying... even if you go into the settings and try to stop all that stuff off"
	TP026	"the connection between searching for something and in a couple of minutes receiving something in my mailbox, I was not feeling comfortable with that"

Table 8: Expectations of a privacy panic button

Expectation	Sample quotes	Codes

Personalized chat / Immediate help	"I would like to have an online chat with someone" (TP027)
	"I need somebody to talk to, it makes me feel more safe" (TP082)
	"A chat window.. more personal... I hate FAQs" (TP065)
	"The best case scenario should offer a real time living person telling you what to do..." (TP097)
	"I would expect some kind of help immediately.... with someone somehow direct,... like a Whatsapp conversation, like a chat" (TP093)
	"You should find an answer and a phone number to talk to someone in some cases" (TP096)
		DLK;
		DLO;
		HIJ;
		CSC;
		MSR

Give me instructions	"It will give me instructions on how to recover my files" (TP076)
	"After posting pictures - let me know how to remove it, or adjust the people who can see the picture" (TP027)
	"I should get a list of reasons why i panic, the list should cover all possible panic situations, it should be a short list" (TP096)
		DLO;
		MSR;
		DLK

Freeze or block my accounts	"The panic should be link to all my details, so that when i press it, it would block the accounts... only lock that device [that I selected]" (TP105) "Freeze the activity of your account...It should be easy to unfreeze if i need it" (TP082) "Freeze it [my account], then inform me, with an SMS, about it so i can do something" (TP030)	LMD; STK
Ask me questions to determine the problem	"Maybe a couple of questions, and not more, of what kind of problem is it" (TP026) "Give me a general list of options of why i am in panic... Ask me why very general and go to more specific questions" (TP008)	DLK; LMD
Lead me through steps (Wizard)	"Start a dialogue with me... [it would ask me] why are you concerned?" (TP089) "It will give me instruction on how to recover my files" (TP076)	TPS; DLO
Assess the consequences	"Show me some bullet points with information on about why is it a bad idea to post this" (TP027) "There has to be someone who can measure the seriousness of the situation..." (TP085)	MSR; MSV
Verify my identity	"Only allow someone from this current IP address to access my account... There should be a link where i can authenticate myself, and correlate my answers with the information that the service knows about me" (TP030) "Take me to a screen.... ask me some security questions... ask me information about myself, like my cellphone" (TP053)	HIJ; STK
Get me out of panic	"The user might think that there is some kind of superman coming, but that's not the case" (TP069)	HIJ
Educate me (give information)	"Suggestions of how to alter my actions so that i dont have bad consequences. Then I might learn in the future." (TP027)	MSR

Table 9: Participants' opinions on three metaphors that were shown to them

Wizard
"It cannot be too long or too complicated, because I am in panic" (TP093)
"I don't like it. It looks too Microsoft thingy... always gets me confused, because you need to look, push next, previous, check" (TP026)
"I is useful to have written information, but if there is a visual way to accompany the written information then it is better... in my case, I don't like lots of text. If I'm trying to solve a problem, I want things to be synthesised..." (TP076)
"Depends on how the panic situation is... When someone needs help, he is trying to look for someone that helps, not that asks many questions" (TP069)
"Better if there are not too much words, because when you are in panic the last thing you want is to read... better to do it with yes and no questions" (TP030)
"Good... as long as you are doing something you are not panicking any more... or get the feeling that you are getting somewhere" (TP065)
"You are use to them during installations... it cannot be too many words... videos and pictures would be good" (TP085)
Emergency card
"You have an overview of the steps to reach a solution, which is better than the step by step" (TP093)
"The emergency card in the airplane, should be reviewed before the plane takes off... [too late to look at it when the plane has already crashed]" (TP089)
"Here you can find the help stages at a glance. He can see, these five steps are related to my problem or not... "(TP069)
" Instructions must be easy. People who understand a lot they know what to do, but for people who do not understand that much it must be easy" (TP082)
"Too much text is not a good idea because I am panicking and I want to do something... you are not thinking rational when you are panicking so it has to be easy and fast..." (TP065)
Account freezing
"It would make me feel a little better" (TP093)
"The question is how do you continue with this?!... the problem is that 'the account is frozen' and then how are we going to continue?" (TP093)
"I like it, I like it very much...This is good thing. I can lock the account and look for a solution. It doesn't give the solution like the [wizard or the emergency card], but it helps me feel safe..." (TP026)
"This is what I was thinking of!!... like block my account down for 30 minutes or something... But then again my ex-husband knows my password, so he could go in and unfreeze it.... so that's where this 2-factor would be nice" (TP053)
"If the account is about to be frozen, i would think "ok, fine, but I wanna have a world on it" (TP076)
"I think this solution is also very useful for the user..." (TP069)
"This of course is very very good... then I know for the moment I am safe, what happened happened, damage is done, but for the moment no more damage" (TP08)
"I would definitely use it... you can control access... it is quite good actually" (TP030)
"Oh, that's great! ... For a first step where you don't know what to do is great when you know that there's no more damage done.. where you can activate that and [breathe] 'now i can think" (TP065)
"Like when you lose your credit card!! It can be very helpful in the panic cases... like loosing your phone and someone going to your email... Why haven't someone else thought about this already!!" (TP027)
"It needs to happen quietly, without other people realising that there has been a big problem" (TP085)
"If i suspect that somebody else is accessing my account then yes, but if it has to do with a file that is not recoverable then not, probably not... (TP097)
Other ideas or comments
Friends helping / comforting / supporting other friends
Freeze account first, then provide me with chat support
Inform users about the limitations of the attack
Provide information through media (YouTube, Images, etc.)
Let users undo certain actions

“My Data Just Goes Everywhere:” User Mental Models of the Internet and Implications for Privacy and Security

Ruogu Kang[1], Laura Dabbish[1,2], Nathaniel Fruchter[1], Sara Kiesler[1]

Human-Computer Interaction Institute[1], Heinz College[2]

Carnegie Mellon University

Pittsburgh, PA

{ruoguk, dabbish, nhf, kiesler}@andrew.cmu.edu

ABSTRACT

Many people use the Internet every day yet know little about how it really works. Prior literature diverges on how people’s Internet knowledge affects their privacy and security decisions. We undertook a qualitative study to understand what people do and do not know about the Internet and how that knowledge affects their responses to privacy and security risks. Lay people, as compared to those with computer science or related backgrounds, had simpler mental models that omitted Internet levels, organizations, and entities. People with more articulated technical models perceived more privacy threats, possibly driven by their more accurate understanding of where specific risks could occur in the network. Despite these differences, we did not find a direct relationship between people’s technical background and the actions they took to control their privacy or increase their security online. Consistent with other work on user knowledge and experience, our study suggests a greater emphasis on policies and systems that protect privacy and security without relying too much on users’ security practices.

1. INTRODUCTION

Today, the Internet is a ubiquitous vehicle for information, communication, and data transportation, and central to many lives. Most who use the Internet, however, have a limited understanding of its technical underpinnings (e.g., [24,30]) and how their personal information is used and accessed online ([28], [39]). Here, we argue that we need to understand what people know and do not know about how the Internet works for two reasons. First, understanding how users think will enable us to design more effective privacy and security controls that match user perceptions. Second, understanding how users think the Internet works will help us develop more effective educational programs so that users, as citizens, can be better informed about privacy policies and other aspects of Internet governance. Towards these goals, we examined users’ mental models of how the Internet works.

Part of the challenge in understanding the Internet is its rapid evolution. The Internet is now massive and embedded into many contexts. It connects billions of individuals around the world through many different types of devices [46]. Many entities are involved in transmitting data and tracking user behavior including third party caching services, first and second level ISPs, cellular network providers, web services, search engines, and ad networks. More personal data than ever is transmitted via the Internet as mobile access proliferates [9] and service providers expand their tracking, creating privacy and security challenges far beyond the ability of end users to manage [38]. Network security tools are not widely used and do not help users understand why or how well they work.

The Internet is not an automated device that works in a simple way to accomplish simple goals. Users have to make decisions that affect their privacy and security, ranging from whether to access public Wi-Fi at an airport to how to share a file with a colleague to constructing a new password for a shopping site. We don’t know the influence of users’ understanding of the Internet on their daily privacy and security practices on the Internet. Does technical knowledge about the Internet help people make good privacy-protecting decisions?

Some previous work has explored user mental models of networking, but has mainly focused on specific domains such as home networking [30,42] and wireless Internet access [24], or specific privacy mechanisms such as firewalls [32]. This previous work does not describe users’ overall mental model of the Internet across domains and its implications for how they think about and take action to protect their privacy on the Internet.

We conducted a qualitative study in which we asked users to describe and explain how the Internet works, both in general and while they did different common, Internet-based tasks. We sampled users with and without computer science or related technical or computational backgrounds. We identified patterns in their conceptual models of the network and awareness of network-related security and privacy issues. A mental models approach, in contrast to surveys or other methods, revealed subtle differences in people’s knowledge of the Internet. Our results suggest that user perceptions do vary as a function of their personal experiences and technical education level. Users’ technical knowledge partly influences their perception of how their data flows on the Internet. However their technical knowledge does not seem to directly correlate with behaving more securely online.

2. RELATED WORK

Why do end users need to understand the Internet? Early literature [12] suggests that Internet literacy is associated with inequality and political participation, as well as economic, legal, and policy decisions. During everyday Internet use, most people may not need to understand technical details such as how a webpage got delivered to their desktop, how caching works, and where their data is being sent; however, when problems occur (e.g., the SSL Heartbleed bug [1] or Target data breach[2]), a better

Symposium on Usable Privacy and Security (SOUPS) 2015, July 22-24, 2015, Ottawa, Canada.

understanding of how the system works can help users understand problems, picture the potential consequences to their personal privacy decisions, and protect themselves from future invasions, such as by obeying recommendations to increase password strength. A clear picture of how users think about the Internet can also help system designers develop technologies that meet users' expectations and help policy makers communicate in ways that are easily understood by lay people [34].

A commonly used method in psychology to elicit users' understanding about a problem is mental models, which are "psychological representations of real, hypothetical, or imaginary situations." [22] Mental models describe how a user thinks about a problem or system; it is the model in the person's mind of how things work. These models are used to make decisions by supporting mental simulation of the likely effect of an action [16]. Mental models of a system can be useful in informing interface design or educational materials because they suggest natural ways to visualize complex system components or user interactions with them.

2.1 Users' mental models of the Internet

A number of researchers have adopted the mental models approach to understand users' perceptions of the Internet [24,30] [30] and Internet-related systems or technologies, such as home computer security [42], firewalls [44], and web security [16].

Diagramming exercises are considered a good way of capturing mental models in addition to traditional verbal reports [22], and this method is frequently used in user-centered Internet research. Poole et al. [30] used a sketching task in order to understand laypersons' knowledge of home networks. Their results suggest that most users, even those who are technically sophisticated, have a poor understanding of home networking structures. Klasnja et al. [24] also used a diagraming task when studying how users understand and use Wi-Fi. Their study revealed that users had an incomplete understanding of important privacy risks when they were connected to Wi-Fi, such as malicious access points, and did not protect themselves against threats, such as seeking SSL encryption. Four out of the eleven participants they observed were aware that other people could possibly access their information being transmitted over Wi-Fi, but this understanding did not raise concerns.

Having a deficient mental model may indicate a lack of awareness of the security risks surrounding Internet activities. Some prior work specifically examined users' perceptions of security systems. Wash [42] interviewed people about how they understood security threats to their home computer and summarized different folk models about home computer security including models centered on viruses and models centered on hackers. Friedman et al. [16] also addressed security risks, interviewing 72 participants and asking them to do a drawing task to illustrate their understanding of web security. They found that the majority of participants relied on simple visual cues like the presence of HTTPS and a lock icon to identify secure connections. Raja et al. [44] studied users' mental models of personal firewalls on Windows Vista using a structured diagramming task. They gave participants images of a computer, firewall, and the Internet depicted as a cloud, and asked participants to connect those pictures with arrows. They then improved understanding of firewalls by showing participants an interface prototype with contextual information.

Many studies show that more technically advanced users have a different understanding of the Internet and computer systems compared to more novice users. Bravo-Lillo and colleagues [8] compared advanced and novice users' differences in their mental models about computer security warnings, finding that advanced users had much more complex models than novice users. Vaniea et al. [41] interviewed people about their experiences with a specific application, Windows Update. They found that a lack of understanding might prevent people from installing important security updates for their computers, thus increasing security risks. Their study suggests that a reasonable level of technical knowledge is essential to guide correct user decisions. Similarly, Zhang-Kennedy et al. [45]'s found that a correct understanding of a system can guide more secure behavior. Their study showed users had a limited understanding of passwords and did not fully understand how password attacks worked. They found users created stronger passwords after using educational infographics about how password attacks work.

Besides privacy-specific research, we can also draw from literatures about people's general understandings of complex systems. Researchers in cognitive psychology argue that complex systems often include multiple levels of organization and complex relationships. Hmelo-Silver and Pfeffer [19] compared experts' and novices' conceptualization of a complex system and found that novices' understanding focuses more on "perceptually available" (concrete) components, whereas experts mention more "functional and behavioral" (conceptually abstract) components. A few other studies [20,33] found that people often assume centralized control and single causality, especially domain novices, whereas experts think about decentralized control and multiple causes when asked to describe a complex system.

The previous work on Internet mental models provides some insight into the nature of users' understanding of the Internet and its anchoring in personal experience. Much of this work, however, is task-specific or focuses on a specific security tool or application. A number of other researchers have conducted interviews or surveys to study users' general or privacy-related Internet knowledge.

2.2 Users' knowledge of the Internet

Various attempts have been made to measure users' knowledge of the Internet. Page & Uncles [27] categorized Internet knowledge into two categories: the knowledge of facts, terms or attributes about the Internet (declarative knowledge), and the knowledge of how to take actions or complete tasks on the Internet (procedural knowledge). Following this argument, Potosky [31] developed an Internet knowledge measure (iKnow) that asks people to rate their agreement as to whether or not they understand terms related to the Internet (e.g., "I know what a browser is"), and whether or not they are able to perform Internet-related tasks (such as "I know how to create a website"). An important question researchers have asked is what impact these two kinds of knowledge have on user security and privacy behavior.

Park [28] measured user knowledge in three dimensions: technical familiarity, awareness of institutional practices, and policy understandings. He found higher user knowledge correlated with online privacy control behavior. Other studies emphasize the role of user skills. Das et al. [11] proposed three factors influence the adoption of security and privacy tools: awareness of security threats and tools, motivation to use security tools, and the knowledge of how to use security tools. Litt [25] found that higher

Internet skills were positively associated with more content generation online and managing one's online presence. boyd and Hargarttai [6] found that users with more Internet skills were more likely to modify their privacy settings on Facebook. Hargittai and Litt [18] developed a scale to specifically measure privacy-related skills. They asked people to evaluate their level of understanding of privacy-related Internet terms such as "privacy settings," "tagging," and email "bcc." Their survey showed that higher privacy-related knowledge was positively associated with better privacy management of social media profiles.

Having more declarative knowledge or skill has not always been shown to predict more secure online behaviors. Dommeyer and Gross [13] found that consumers are aware of privacy-protective strategies, but do not use them. In a study by Nguyen and colleagues [26], some participants expressed uncertainty about how store loyalty cards would be used, but they did not take any protective actions to protect their personal information. Furnell et al. [15] studied how people manage security threats to home PC systems and found advanced technical users did not use more effective security practices than novice users.

The Internet today is much different than what it was 10 years ago, so people may perceive or use it very differently today, especially in managing their privacy. In 2003, the majority of American Internet users expressed strong concern about information used by governments and corporations, but they had little knowledge of how their data flows among companies [39]. A more recent 2011 review of the literature suggests that people's awareness of organizations collecting their personal information increases their privacy concerns [35], but there remains little understanding of how people think the Internet works. In late 2014, Pew Research Center conducted a national U.S. sample survey to test Internet users' knowledge of the Web by asking 17 questions about Internet terms (e.g., "URL"), famous technology celebrities (e.g., identifying Bill Gates' photo), and the underlying structure of the Internet (e.g., explanation of Moore's law) [29]. Their survey indicated that the majority of Internet users recognize everyday Internet usage terms, but very few are familiar with the technical details of the Internet and most do not understand Internet-related policies.

In sum, there is mixed and indirect evidence of whether or not an accurate mental model and more advanced Internet knowledge are associated with more secure online behavior. In light of the new data privacy and security challenges associated with the Internet's evolution, we wanted to assess how people currently understand the Internet, their perceptions of how their data flows on the Internet, and what they are currently doing to protect their privacy or data security. Our work aims to examine the relationship between people's knowledge and their privacy and security behavior in today's Internet environment, and to move towards a better understanding of the kinds of Internet knowledge users need to have.

3. METHOD

We conducted semi-structured interviews with twenty-eight participants about their mental models of the Internet. A list of all the participants is shown in Table 1. In addition, after completing the interviews with technical and nontechnical participants, we invited 5 domain experts (faculty members in computer networking or computer security domain at a research university) to review and evaluate several mental model drawings generated by technical and nontechnical participants. Here, we first introduce the method and results of the interviews with participants. Then, we discuss the implications of our results and incorporate experts' comments into the discussion and implication section.

3.1 Participants

We did three rounds of data collection and recruited a total of 28 participants. Each participant was paid $10 for a 30-45 minute interview session.

The first two rounds of participants were recruited through flyers, personal contacts, and an online participant pool at a US east coast research university. At the outset of this study, we used educational level and college major as a proxy for technical knowledge (used for N01-N09, T01-T03). For other technical participants recruited in the second round (T04-T10), we developed a screening survey for technical knowledge, only accepting participants who scored 5 or higher in an 8-item survey as technical participants (Appendix A.) Those who scored lower than 5 counted as non-technical participants (N10, N11). These nontechnical and technical participants included people from the local area, university staff members, and students pursuing all levels of degree study. Non-technical participants had a mix of backgrounds. Technical participants all had computer-related college majors.

Table 1. Study participants (Total = 28; N = non-technical participants; C = community participants; T = technical participants; *T11 was recruited with the community sample).

Identifier	Gender	Age	Education background
Lay participants (N = 17)			
N01	M	19	Finance
N02	M	22	Finance
N03	M	22	Biomedical Engineering
N04	F	18	Geology
N05	F	22	English
N06	M	22	Law
N07	F	21	Cognitive science
N08	F	19	Statistics; psychology
N09	F	22	Legal studies
N10	M	30	Music; foreign languages
N11	F	18	Neuroscience
C01	M	64	Engineering; public health
C02	M	32	Culinary arts
C03	M	62	Communication arts; religion
C04	M	49	Psychology
C05	F	58	MBA
C06	F	30	Foreign policy
Technical participants (N = 11)			
T01	F	19	Computer science
T02	F	21	Computer science
T03	F	27	Computer science & HCI
T04	M	25	Information technology
T05	F	24	Electrical/CS engineering
T06	M	26	Computer science
T07	M	25	Information technology
T08	M	23	Computer science
T09	M	27	Software engineering
T10	M	24	Software engineering
T11*	M	32	Computer science

Because our initial two samples were similar in age and university education, we also recruited a third group of participants from the local community by posting an advertisement on craigslist with the inclusion criteria of age 30 or older (C01-C06). One of these participants (T11) had a computer science background, so was treated as part of the technical sample. Both the nontechnical and community participants had non-computer science related education backgrounds, so we refer to them together as "lay participants" in the following sections. Participants who had had formal computer science or computing education are referred to as "technical participants."

3.2 Procedure

In the interview study, participants were brought into a room equipped with pen, paper, and a desktop computer. After an overview of the study, participants completed a short survey regarding Internet experience, smartphone literacy and computer knowledge. They were also asked about the number and types of devices they owned.

After completing the survey, participants were guided through the main drawing tasks. Every participant was first prompted to explain how the Internet works, and asked to draw a general diagram of it in whatever form they chose on a large sheet of paper in front of them. Participants were instructed to verbalize their thought process as they drew, consistent with traditional think aloud protocols [14]. A video camera captured participants' drawings and voices. All recordings were labeled using anonymous identifiers. No personally identifiable information was collected or recorded.

Each participant was then asked to draw several diagrams about specific tasks they did on the Internet following the same procedure. The tasks used were a subset of the following: *watching a YouTube video, sending an email, making a payment online, receiving an online advertisement* and *browsing a webpage.* After each model drawing was completed, participants were asked several follow-up questions, clarifying drawings and explanations as needed. Additionally, participants were asked to draw a separate diagram for each task if they thought it worked differently on mobile devices. The interview script is attached in Appendix B.

After the drawing tasks, participants filled out a post-task survey with demographic questions, as well as a series of Internet knowledge questions (attached in Appendix C). The knowledge questions included self-rated familiarity with nine technical terms on a 5-point scale (IP address, cookie, encryption, proxy servers, SSL, Tor, VPNs, privacy settings, and private browsing), and seven true/false questions about security and privacy knowledge (e.g., "Tor can be used to hide the source of a network request from the destination.") We developed the knowledge questions by consulting domain experts in computer security and tested their reliability with two independent samples (see Appendix C for details).

3.3 Technical knowledge level

All 28 participants filled out the same post-task survey. Besides differences in academic background, technical participants performed significantly better than lay participants in both the self-rated familiarity questions (mean: technical = 3.59, lay = 2.47, t [26] = 4.32, $p < .001$) and correctness on the true/false questions (mean number correct: technical = 4.27, lay = 1.53, t [26] = 5.83, $p < .001$).

3.4 Data Analysis

We qualitatively analyzed participants' think aloud responses to identify key differences across mental models. We conducted our analysis iteratively, carrying out three rounds of data collection and subsequent analysis, allowing the first analysis process to guide our second round of data collection, and then the third. Our initial analysis occurred after the first 12 sessions with participants (predominantly non-technical participants). We focused on the diagrams they generated during our sessions as well as the video and audio recorded during our sessions. By comparing and contrasting across user models, we generated a set of codes that indicated dimensions on which the models varied. To verify and extend codes and themes identified in our first round of data analysis, we conducted a second round of analysis, extending codes identified in our first round based on new features of the second set of models. In the last round of data collection, we added a few questions to the interview based on results from the previous two rounds. The third round of data collection expanded the age range of our sample and let us examine the influence of users' past experience and concerns on their perception and behavior. Six interview recordings were lost due to equipment problems but field notes on paper were available. The remaining 22 of the 28 interviews were recorded and transcribed (9 technical, 7 nontechnical, and 6 community participants). Aside from analyzing the drawings, we performed qualitative data analysis of the verbal transcripts and field notes using a grounded theory approach [10]. The data were coded in Dedoose (http://www.dedoose.com/). A second researcher independently coded 15% of all the interviews. Our analysis showed a good inter-coder agreement between the two researchers (kappa = 0.79).

4. RESULTS

Our analysis showed that participants with different technical education and personal experiences had very different mental models of how the Internet works. These models were related to participants' perceptions of privacy threat and what happens to their data on the Internet. However, technical education and mental models did not seem to be very predictive of how participants acted to protect their privacy or security. Those actions appeared to be more informed by participants' personal experience. In the following sections, we first discuss users' knowledge of how the Internet works as a system and their awareness of security and privacy features in the system. Next, we present people's different perceptions of their personal data on the

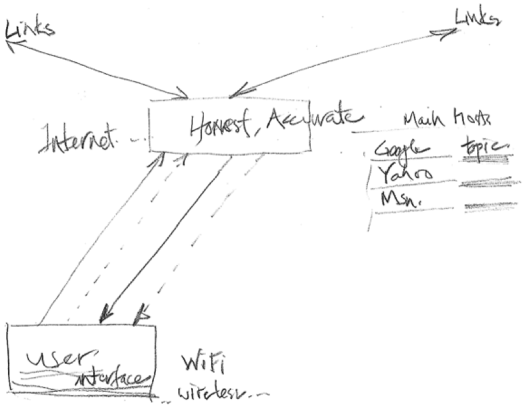


Figure 1. Internet as service (C01)

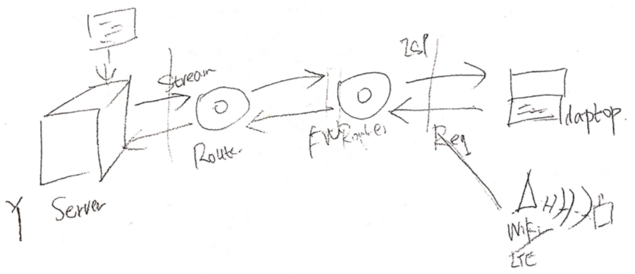


Figure 2. Articulated model with hardware components (T10)

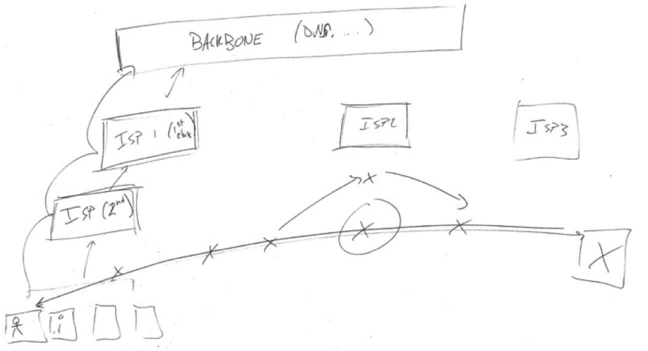


Figure 3. Articulated model with multiple layers of the network (T06)

Internet. Lastly, we show the methods participants take to prevent their data from being seen and discuss the connections between their knowledge, perception and the protective actions.

4.1 Users' knowledge of the Internet

Participant models varied in their representation of the Internet as a simple system or service (the "Internet" in Figure 1) or as an articulated, technically complex system (Figure 2 and 3).

4.1.1 Simple vs. articulated system mental models

A majority of the lay participants represented the Internet as a comparatively simple system or service consisting of the user connected to a "server," data bank, or storage facility. These participants used metaphors such as earth, cloud, main hub, or library that receives and sends out data. Thirteen lay participants and one technical participant belonged in this category. Their models showed that the Internet receives and sends out data, indexes webpages, and responds to their different requests. A few users considered Google or Yahoo the main provider that connected them to other webpages.

"So everything that I do on the Internet or that other people do on the Internet is basically asking the Internet for information, and the Internet is sending us to various places where the information is and then bringing it back." (C01, Figure 1)

Most lay participants only expressed surface-level awareness of organizations and services that they interacted with directly such as Google and Facebook, but did not mention any of the underlying infrastructure. When talking about making online payments, for example, they mentioned a number of different organizations involved in the process such as "the bank," "Amazon," and "PayPal." Some were aware of physical objects that helped them connect to the Internet (see N05's drawing of a router in Figure 4). Three lay participants also drew mobile towers when describing a cellular network. Three thought satellites played a role in connecting them to the Internet, but none of the technical participants mentioned this.

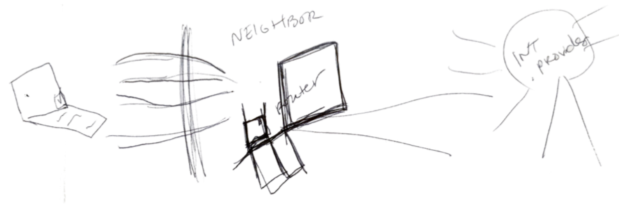


Figure 4. Drawing of how she uses neighbor's Wi-Fi (N05)

In most technical participants' drawings, we seldom saw a simple system or service representation of the Internet. Instead, users had more articulated models of the Internet as a complex system with varied hardware components and a more involved set of connections among components (Figure 2 and Figure 3). Ten technical and four lay participants belonged in this category. The number and presence of entities and organizations within participants' sketches mirrored to some extent their Internet literacy levels. The presence of other computers, servers, ISPs, DNS, routers, servers/clients, and infrastructure hardware spoke to a participant's knowledge and understanding of the Internet as a complex system.

Some technical participants articulated their view of multiple layers of the network (Figure 3), whereas most lay participants described one layer of the network. A few technical participants mentioned physical layers ("fiber cable", T05), or concepts potentially associated with a physical layer such as physical location (such as a "U.S. server," or a university as a physical entity). Most technical participants (9 out of 11) expressed broader awareness of entities and organizations involved in the Internet. For example, 6 technical participants noted there were many different ISPs. Furthermore, technically advanced users had specialized knowledge. Five technical participants mentioned network protocols such as "TCP/IP", "SMTP", or "IMAP", but none of the lay participants mentioned these concepts. Some

Table 2. Differences between simple and articulated models

	Description of the models
Simple and service-oriented models: 13 lay participants; 1 technical participant	Represent the Internet as a vague concept or a service; Only show awareness of organizations or services they directly interact with; Lack awareness of underlying layers, structures and connections; Use inconsistent or made-up terminologies.
Articulated technical models: 4 lay participants; 10 technical participants	Represent the Internet as a complex, multi-level system; Show broader awareness of components and organizations in the network; Express awareness of layers, structures and connections; Use accurate, detailed, consistent terms.

technical participants also mentioned logical elements such as "routing" or "peering." The differences between these two types of mental models are explained in Table 2.

There were aspects of the mental models both groups had in common. Regardless of their technical background, participants said that the Internet connects computers and supports communications. For instance, a 49 year-old local flower shop owner was quite excited about all the changes the Internet has brought to his life, and mentioned that the Internet enables him to *"talk to friends that I've lost contact over the years."* (C04) A technical participant focused more on the infrastructure: "*There's a level at which there're ISPs that communicate with each other*." (T06)

4.1.2 Awareness of security and privacy[1]

We analyzed the comments related to security and privacy that naturally emerged during the interview as a measure of people's general awareness and attention to security and privacy. We did not explicitly prompt people to talk about security mechanisms of the Internet. The concepts that emerged concerned private vs. public spaces, protection mechanisms, trust, and perception of security on mobile phones vs. computers.

4.1.2.1 Public vs. private communication

Six lay participants and two technical participants talked about distinctions between public vs. private information or connections. For instance, one nontechnical participant thought that home Wi-Fi is more secure than public Wi-Fi because it has firewall and security settings (N09). Several participants thought sending an email or doing an online payment is private while watching YouTube videos is public. A few participants mentioned privacy settings on YouTube or Facebook that they could use to control whether their information was public or private.

"I think there's a user profile [on YouTube]. I mean that to me is a much more public space." (C06)

4.1.2.2 Protection mechanisms

We coded users' expressed awareness of protection mechanisms such as encryption, passwords, certification of websites, and verification steps implemented by websites. One lay participant and seven technical participants said that their email, online payments, or connections could be encrypted. T04 said, *"If I'm going to use Gmail then I assume that, by default, the connection is going to be encrypted between my PC and the Gmail server."* Another technical participant drew a little lock sign in his model to indicate that the connections are encrypted (Figure 5).

One lay participant (N06) said, *"I don't put [my credit card info] in when there's not like that little lock up on top of the screen. I think it's pretty secure."* Also, when talking about sending an email or making an online payment, some participants mentioned the bank or email server would verify the requester's identity (T04, T08, and N11). In Figure 5, the technical participant included a certificate authority ("CA") in his model of online payments.

4.1.2.3 Trust

Eight lay participants and three technical participants expressed shared beliefs about the security provided by big companies or institutions, and considerable trust in those they knew. The cues participants used to decide whether or not they would trust a website included their knowledge that other people had used the same service, that it was a reputable brand, terms of service, certificates, warnings, and whether or not they had had a bad experience on the site.

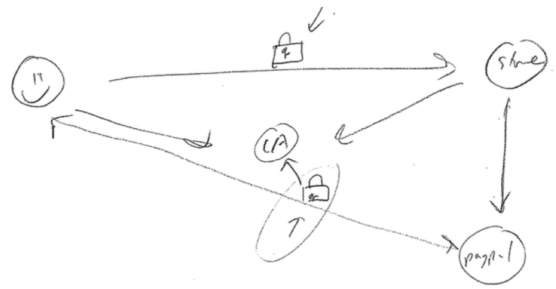


Figure 5. Model of making an online payment to a shoe store (T09)

"I think if this was Amazon, their site is probably protected." (C05)

One participant transferred his trust of the physical bank to the online world.

"I talk to the employees there in person a lot, and they just seem to have a level head on their shoulders. I don't think they would give out their information to anybody over the phone without verifying who they were with some kind of credential verification." (T11)

4.1.2.4 Mobile phones vs. computers

Participants offered mixed opinions about whether it is more secure to connect through the phone or through their computer. N10 said it is less secure to do banking or payment related activities on a mobile phone, because he felt it was like "*sharing wireless connections with other people in a public network*." He thought the difference between connecting from his computer vs. connecting from his smartphone was that the connection on mobile phone was wireless.

By contrast, T10 always used his smartphone to make payments because he was worried that his computer might have a virus or tracking software and thought his phone would be more secure. C01 thought a mobile hotspot was more secure than connecting to a public Wi-Fi at a coffee shop because he was the only one on it.

4.2 Users' perceptions of their data

A great deal of privacy-related policies and research efforts concerns organizational practices in the collection, retention, disclosure, and use of personal information. In our study, we asked users about their perceptions of how personal data is dealt with on the Internet.

4.2.1 Where does my data go?

Most participants were aware that their data is sent to the servers of the company who provides them services such as Google. Two lay participants had a very vague idea of where their data went (C03 and C04). When asked about where his data goes on the Internet, the flower shop owner said:

[1] This section and following sections are based on the 22 interview transcripts, including 9 technical and 13 lay participants.

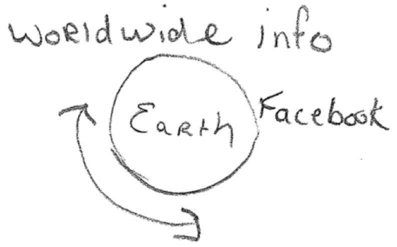


Figure 6. A depiction of where his information goes online (C04)

"I think it goes everywhere. Information just goes, we'll say like the earth. I think everybody has access." (C04, Figure 6)

Regarding where their data is stored, participants mentioned "Google's large storage banks," cloud storage, ISPs, and advertising companies. One participant said, *"Once something is online, it's there forever."* (T11) A few others were not sure if information would be stored permanently, using the evidence of having seen webpages removed.

Many participants were familiar with the partnerships among different organizations, an idea they mostly learned from news articles or personalized advertisements and services. N11 mentioned the *"paid relationship between Google and Amazon."* C02 said, "*Government can piggyback off the different servers and get all the information of what they are looking for.*" Eight lay participants and eight technical participants talked about personalized advertisement and personalized service such as tailored search results and video suggestions. Recommendations or ads tailored to their interests made people aware of a data partnership among different companies, but most of them could not spell out to whom their data was sold.

4.2.2 Who can see my data?

After each participant completed their drawing of the Internet, the interviewer asked, "Are there any other people, organizations, or companies that can see your connections and activities?" Privacy threats participants identified in frequency order include: companies that host the website (e.g., YouTube, Amazon) (mentioned by 18 out of 22 participants), third parties (e.g., advertisers or trackers) (mentioned by 14 participants), the government (mentioned by 12 participants), hackers or 'man in the middle' (mentioned by 12 participants), other people (e.g., other users online, other people using the same Wi-Fi) (mentioned by 11 participants), internet service providers (mentioned by 8 participants), employer (mentioned by 2 participants), and browser owners (mentioned by 1 participant). Figure 7 shows a fairly complete representation of all the people and organizations that the participant thought had access to his information, including the government, hackers, company, ISP, and third parties. This participant (T11) studied computer science in school, but stated that his current job was not related to technology.

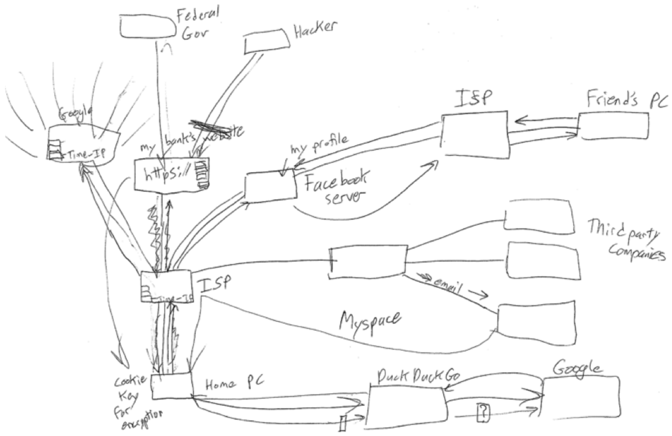


Figure 7. Model of the Internet including who can access his information (T11)

We compared how much lay and technical participants' mentioned the six most frequently mentioned threats. These two groups did not differ significantly in their general awareness of who has access their data. Lay participants mentioned on average 3.23 threats (out of 6), whereas technical participants mentioned on average 3.67 threats, a small non-significant difference overall. As shown in Figure 8, technical participants were significantly more likely, however, to mention hackers having access to their data than lay participants did. Across the categories of threat, they were more specific in identifying threat such as ISPs, whereas lay participants mentioned more vague threat such as third parties: "*whoever tries to make money off of you.*" (C02) This generality was probably due to the more simplistic mental models lay participants had about the Internet.

Although technical education did not seem to influence participants' overall perception of privacy threat, the mental models (simple vs. articulated) were somewhat predictive of the number of threats people perceived. We found that, on average, participants with articulated models mentioned more sources that might have access to their data than those with simple models (mean number of threats mentioned by people with articulated models = 4 and the number mentioned by those with simple models = 2.56, t [20] = 2.80, p = .01). Those with articulated mental models expressed higher awareness of privacy threats from government, hackers, and ISPs. This higher level of awareness may be caused by these people's better understanding of where risks could occur in the network. For example, with a mental model like Figure 1, there is no way the user would know what privacy risk his ISP could bring to his data on the Internet.

Besides these specific threats, some participants thought that "everyone" could access their information, either in the general

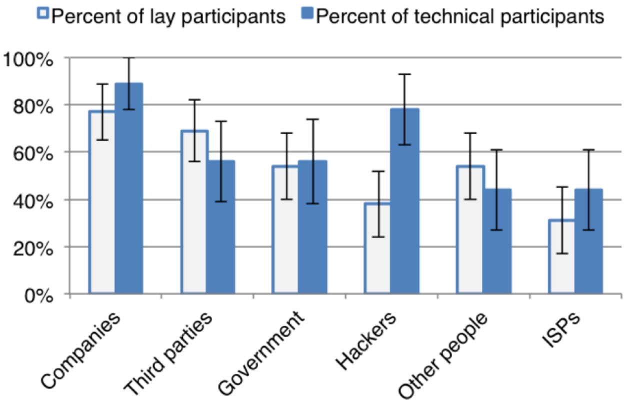


Figure 8. Percent of lay or technical participants who mentioned each group that might have access to their data.

sense, or in certain situations. T06 stated that, *"the Internet is not designed to be private"* and explained the technical details of why this is the case – *"at the end of the day you're relying on correct implementations of logically sound security protocols, and historically most implementations aren't correct and most protocols aren't logically sound. So, it's just a question of an arms race of who's paying more attention."* Two lay participants also held similar opinions about their information online – *"anybody that has the capability of getting through passwords or encryptions can get it [personal information]"* (C02 and C03). N07 thought that YouTube is open to "*a lot of other people*," so the data is available to everyone. Similarly, two community participants (C04 and C05) thought the Internet in general grants everybody access.

As described earlier in the paper, participants tended to deem sending an email and making online payment as more private than activities like posting on social media. Therefore, when asked whether others could see their transaction of an online payment, T10 said *"I don't think so."* Two other technical participants (T07 and T08) thought no one could intercept their email, because they had a password or encryption to protect their email content. N06 also thought that no one could see his email, but was not able to provide any further explanation except that *"email is more private."* A technical participant (T11) mentioned that he expected Netflix not to sell his data because it's a paid service, but he was uncertain of how exactly it works: *"I try to browse through the terms and conditions but there's so much there I really don't retain it."*

4.2.3 Different types of information

Previous research has shown that users consider some personal information more sensitive than others [3]. From our interviews, we saw different user privacy expectations for different types of information, including not only personal information, but also technical identifiers. For instance, three lay participants thought companies could access their purchase history but not credit card information (N06, C02, C06). N11 suspected companies would be more interested in what she watched on YouTube than her emails, so she expected more protection on emails. Some participants were aware of the differences between identifiable information (such as names) and non-identifiable information like an ID number or IP address (C05), but she also said *"they could find it [my name] from this ID."* T09 pointed out that even for encrypted messages, his ISP could see all the packets and they could still tell *"where the origin, which is me, and what it's going."*

4.3 How do people protect their information?

4.3.1 Protective actions

We asked people: "Did you do anything to prevent any others from seeing your connections or activities?" Participants mentioned a wide range of protective actions they had tried, such as not logging in to websites, watching for HTTPS, and using cookie blockers or tracker blockers. We categorized the actions participants used into four categories as shown in Table 3. *Proactive risk management* includes general precautionary steps people take in daily use of the Internet. *Event-based risk management* includes people's actions towards specific requests or intrusions. *Controlling digital traces* includes actions that mask or remove people's digital or physical footprints. *Securing connections* indicates methods people take to make sure their connection to a certain site or their general Internet connection is secure or anonymized.

Table 3. Protective actions used by lay participants and technical participants

Types of protective action	N	# of lay participants who have used this type of action (out of 13)	# of technical participants who have used this type of action (out of 9)	Actions
Proactive risk management	15	9 (69%)	6 (67%)	Use anti-virus program Back up personal data Be cautious when using public Wi-Fi Change password regularly Do not use or use less social media Take care of physical safety of credit card Use tape to cover computer camera Switch devices
Event-based risk management	8	5 (38%)	3 (33%)	Change email password when asked Do not accept many friend requests Do not give email address when asked Do not open pop ups Exit malicious website Not sign up or not log in
Controlling digital traces	15	10 (77%)	5 (56%)	Use anonymous search engine Use cookie blocker or other tracker blocker Cut off address from package Limit or change information shared online Delete cookies, caches, history Use private browsing mode Use fake accounts or multiple accounts
Securing connections	12	5 (38%)	7 (78%)	Encrypt data Watch for https in websites Use Tor Use password to secure Wi-Fi

Although our technical participants were more knowledgeable of how the Internet works in the backend, they did not in general take more steps to protect their information online, in comparison with lay participants (Mean types of actions used by technical participants = 2.33, lay participants = 2.23, *n.s.*). As shown in Table 3, the only difference was that technical participants were somewhat more likely to mention securing their connections than lay participants and the comparison shows a trend approaching significance (t [20] = 1.99, p = .07). This finding contrasts with some of the prior work that has shown a correlation between technical knowledge and privacy practices [28], but this ostensible contradiction may stem from how we and other authors explored the influence of technical knowledge. Our study was focused on how people understand the Internet and its infrastructure whereas other studies [18,28] have mainly focused on users' Internet literacy and their familiarity with privacy practices.

We counted the diversity of privacy threats that participants mentioned among six frequently-mentioned sources of threat in Figure 8: companies, third parties, government, hackers, other people, and ISPs. We then compared how perceptions of threat were related to protective action, by conducting a nonparametric correlation analysis on the number of threats they mentioned and the number of protective action types they took. The analysis yielded a moderate correlation (r_s = .40, p = .06). This result indicates that the awareness of privacy threats is probably a stronger indicator of people's protective actions than their general technical background. This comparison points to a difference in the impact of general technical knowledge, which does not seem to predict actions, and the awareness of Internet privacy risks.

Many participants had some knowledge of protective actions but had not used them. This may be one consequence of the privacy paradox [4] whereby people have general desire for privacy but do not act on this desire. We wanted to know what our participants would say about why they did not take steps to protect their information.

4.3.2 What prevents people from taking action?

Four categories emerged when participants talked about why they did not take actions to protect their information from being seen. The most common explanation was similar to the statement, "I've nothing to hide" [37].

Eleven participants (8 lay and 3 technical participants) said they were not worried about their information being accessed or monitored or did not have the need to use tools. Many participants were not concerned because they did not do anything very subversive, illegal, or had little to protect. T10 said, *"I don't care who sees and reads my email"* although he was aware that *"hackers can act as mail servers."* Two participants were not worried also because *"I don't put that much information out there."* (C03 and C04) Three participants said they had too little money to protect, *"I don't have much money to worry about."* (C03) C01 said he was not worried because his data is among *"an awful lot of data."* T11 said he knew a lot of methods that other people had used to mask their IP address, such as proxy servers, but he never pirated so much music that he felt the need to do so. T04 mentioned Tor as a protective method during the interview, but also said, "*Till now I haven't had the need to use Tor.*"

The second reason given for not taking protective measures was that doing so would sacrifice effectiveness or convenience. T11 started to use DuckDuckGo (https://duckduckgo.com/), an anonymous search engine, to conduct anonymous searching but switched back to Google after several months, because Google gave better search results, tailored to his interests. T06 quit Facebook but did not quit Google, because "*their services are a lot more useful*." C06 said she is willing to take risks because doing things online is much more convenient than the "*old-fashioned way*."

Another reason given for not taking protective measures was the poor usability of privacy protection tools or software. T07 said that it is hard to do incognito browsing on smartphones. N10 knew that he could get a blocker but suspected some of the blockers might include viruses and would add clutter to his browsing experience.

For a minority, a feeling of helplessness and lack of procedural knowledge prevented them from taking any action [36]. C05 said that hackers would probably hack into the website servers instead of individual users, and there was nothing he could do about it. Four lay participants said they lacked enough information to discuss actions they could do to prevent others' access to their information. C03 said he deleted cookies and then said, *"I don't know how to do anything else."*

The relationship of risk perception and action is also shown in participants' remarks. A technical background could influence awareness of threats and risks to some extent, but risk perception could also be shaped by personal experience. T11 started using DuckDuckGo after hearing about news related to Target's data breach and NSA monitoring. He became worried about how many people could see his information online. T11 had also been harassed by a Craigslist job poster because he gave out his phone number and email address. The Target data breach was also mentioned by C02, C07 and T11. After C07 was notified of the breach, she was not sure whether she was a victim or not, so she kept checking her statements carefully for a few months. Consistent with previous research [23], these instances suggest that past negative experience triggers more secure online behavior and a heightened level of privacy concern. In contrast, people who had not experienced a negative event seemed to be habituated to the convenience brought by the Internet and were less motivated to take protective actions online. A community participant (C04) had a friend who experienced identity theft, but hearing about this story did not make him worry about his information, and he stated, *"unless it happens to you it's hard to walk in somebody else's shoes."*

5. DISCUSSION AND IMPLICATIONS

Our study suggested technical education determined whether people viewed the Internet as a simple, service-like system or as an articulated technical system. Those with a more articulated model of the Internet expressed higher awareness of the different people or organizations that could access their data. However, technical participants did not take more steps to protect their online information than those with lower technical knowledge. After the second round of data collection, we invited five networking and computer security experts to review several lay and technical participants' models and discuss implications for security and privacy.

5.1 The role of knowledge in privacy decisions

Previous research is unclear as to whether or not Internet knowledge is associated with better management of one's privacy and security. We found little difference in the actions that people with more technical Internet knowledge took versus the actions lay participants used except that technical participants were slightly more likely to secure their connections (Table 3). Many technical participants expressed that they did not need to take action, and that the tools were inconvenient. These observations echo the finding in [15] that technical users complained about practical factors that prevented them from taking secure actions (e.g., "security is too expensive"). Also, expert reviewers pointed out that technical participants might be overconfident about their knowledge, which may cause a *"skewed view of security"*.

In comparison to general Internet knowledge, people's knowledge of privacy threats and risks might be more predictive of their privacy behaviors. Expert reviewers identified overlooking privacy and security risks as an important limitation of simpler mental models. They indicated that users who lacked awareness of Internet entities or organizations would have difficulty identifying the source of a problem or error when attacks, leaks, or other security issues occurred. One expert reviewer said that the lack of entity awareness in the simple mental model might engender too much trust in data privacy and security :

"*When it's just a magic black box, you tend to say well, I trust the magic black box, and so I would worry a little bit more that someone with this level of abstraction would not think as much about who could be sniffing on their communications or changing it or how they interpret security warnings and things like that.*"

Our data supported this argument, by showing that people with an articulated model on average expressed higher awareness of who could access their data. The number of threats people identified seemed to be correlated with protective actions they took.

Another dimension of knowledge is that of protection tools or systems. Expert reviewers were concerned that insufficient knowledge of encryption mechanisms could lead to data security risks. They speculated that users who were more aware of encryption would be better at controlling their data privacy and security. However, we did not find this association in our data. Participants who were more aware of protection mechanisms such as encryption or website certifications did not report taking more protective actions. There might be some skewness in our data because the majority of our participants were aware of protection mechanisms (17 out of the 22 we coded), so the relationship between knowledge of protection tools and people's actual action requires further investigation.

5.2 Uncertainty in knowledge and concerns

Across all three rounds of data collection, participants expressed a great deal of uncertainty or lack of knowledge about how the Internet works, how their data is collected, shared or stored, what protective actions they could use, and whether the protection is effective or not. This finding echoes Acquisti et al.'s [3] work demonstrating the privacy uncertainty. For example, N11 used a Google app to block trackers but she was not sure how effective it was and still concerned: *"I don't think it blocks everything."* Several nontechnical and community participants were confused about how attacks or problems happened. Finally, three technical participants expressed doubts about who had access to their data. These different uncertainties may prevent people from accurately estimating their privacy and security risks.

Another dimension of uncertainty in people's knowledge is whether or not their mental models can adapt to changes in technology. A few nontechnical participants' perception of the Internet seemed to be dominated by names of well-known content providers (e.g., "Yahoo", "Google", and "Facebook"). They also used name recognition as a safety heuristic—deciding that a website is secure because it is a well-known brand. However, advances in technology, security breaches reported in the press, and the rise of new companies could change these attitudes. As noted by one expert reviewer, participants did not seem to update their models as fast as the Internet changed. Only a few participants expressed awareness that their models might be outdated.

Much previous research about the privacy paradox discusses people who claim they are concerned but do not take steps to protect their information. Our study reveals another possibility: participants who showed less concern about privacy in some situations actually took protective actions in other situations. For instance, two participants (C02 and T11) mentioned reading websites' terms of service to figure out how the companies handled their data, which indicates they are pretty cautious about data privacy. But when the interviewer asked about their privacy concern levels, C02 said he was only moderately concerned. Although T11 said he was worried about privacy, the reason he switched back to Google from DuckDuckGo was *"I don't really have anything to hide."* A number of researchers have shown that general privacy segmentations like Westin's do not sufficiently capture people's complex privacy needs, and concerns do not align with their behavior [43]. Our finding suggests that we need more detailed measures of privacy concerns instead of a general privacy concern scale.

5.3 Implications for design and policy

People rely on their specific experiences and on observable cues to understand how their information is accessed, used, or protected online. Experiences include actions they've taken (e.g., password or monthly payment) and received (e.g., spam). Cues include interface cues (e.g., lock sign, dots replacing password), dynamic information (e.g., tailored advertisements), and social information (e.g., comments on a post). Most of our participants were aware of personalized services or advertisements, which spoke to their high awareness of companies and third parties having access to their data. A technical participant explicitly noted this transparency: *"They are totally telling you that they know what you're viewing, because they recommend videos for you"* (T06). Social cues on sites like YouTube and Facebook (e.g., user profiles, number of views, and uploader's profile) indicated the presence of other users, which rendered participants' activity on those sites more public. Regardless of their technical knowledge, participants seem to have made most of their privacy-related decisions based on these experiences and cues.

Most observable cues inform users about their privacy and security in the application layer and mainly deal with threats from governments and corporations. Other limited cues educate users about social threats from other people, such as supervisors, or security risks at other layers of the network. However, it can be easy to miss these limited cues. One design implication is to provide a "privacy indicator" for people's Internet activities, showing them who can see what information. Bernstein et al. [5] proposed that visualizing the size of one's audience on social media would help users understand the exposure of one's posts. Visualizing one's audience across applications and different network layers might help to increase users' awareness of privacy and security risks. At a minimum, applications could inform users about what control they have over their data, if any, once they put it online. Data access was the most important aspect of privacy emphasized by expert reviewers, but it was also the most difficult for participants to grasp. The challenge, as one expert reviewer noted, is in which data or security risk to surface or prioritize for user attention.

The dilemma of multiple sources of risk implies that even if we raise awareness about some sources of risk (e.g., tracking and third party advertisements), there are others, and if we try for more comprehensive warnings, we may cause overload, annoyance, and security tool abandonment or lack of adoption [41]. Moreover, warnings can raise user confidence, which can in turn increase their risk behaviors.

In half of the interviews we coded, participants said they trusted institutions or companies to take care of their security. Some expert reviewers and a few technical participants suggested that users are putting too much trust in the system or the software, and taking too little responsibility: *"If you want to* [*achieve*] *privacy, you have to take that into your own hands"* (T06). This attitude reflects a laissez-faire policy perspective that our data and prior work challenges. If users cannot understand or control their own privacy, or if they have a limited role, where should responsibility rest? Policy makers could enforce more strict laws and regulations to mandate organizational practices, but users may still feel uncertain and helpless if we fail to provide good education programs about the influence of policies on their personal data. In the last round of interviews, we asked the community participants whether or not they thought current laws provided enough protection for their privacy. All but one participant thought laws did not provide enough protection; however, when the interviewer asked participants about their knowledge of Internet-related policy, most participants could not articulate anything beyond what they heard in news reports, suggesting a strong need for investigative journalism.

5.4 Limitations and future work

Because we used a think-aloud style qualitative study, our observations were influenced by the questions we posed and the knowledge people recalled. Participants may have had more knowledge of the Internet or security mechanisms than they expressed. Another limitation of conducting a qualitative study is that we have a comparatively small sample size. The small sample size may prevent us from detecting small but real effects of declarative and procedural knowledge on motivations and behavior. We are conducting larger sample surveys that will provide more statistical power to detect correlations among users' knowledge, expressed awareness of threats, and use of protection tools.

Individual demographic differences such as age, profession, and area of the country may also influence people's Internet perceptions and behaviors but are more appropriately compared in a quantitative study. Our sample is especially sparse in some age ranges. In future work, the interplay between demographic factors and users' technical background should be examined.

Another limitation of this work is its scope. Our study specifically examined participants' knowledge of the underpinnings of the Internet, how they tried to control data access by others, and, in the post-test survey, their understanding of some tools and concepts for controlling privacy on the Internet. We did not measure participants' knowledge of how attacks occur or how they understood different privacy and security threats in detail. We used their awareness of who had access to their data as a proxy for their awareness of risks and threats. Our data showed gaps in people's understanding of how attacks occur, but we do not know how these gaps influenced their risk perceptions or behaviors. There is much more to learn about these and other dimensions of Internet knowledge. More extensive measures are needed to explore the relationships among technical knowledge, understanding of threats, and people's privacy behaviors. For instance, it will be important to understand if knowledge of the complexity, layers, entities, and operation of the Internet help people to understand how and where threats can occur, or whether they simply need to have in mind a mental list of threats and methods to lower risk. Future work should examine whether education or design interventions can improve specific aspects of users' mental models of the Internet such as entity awareness. We also need to understand better how people understand security versus privacy—or whether they even need such a distinction.

6. CONCLUSION

As the Internet becomes more technically complex and, at the same time, more intertwined with everyday life and the well being of organizations, we face the question of how to educate users to help them protect their privacy. We conducted a qualitative study to investigate users' mental models of the Internet and their knowledge of data flow on the Internet. We examined how they conceptualize the process of connecting to the Internet and how they think others can access their data online. Our analysis revealed strong differences among users with different educational backgrounds. The majority of those without computer science education had simple, service-oriented mental models whereas those with a background in computer science had an articulated many-layer model of the Internet that included key entities and organizations. People with a more articulated model expressed higher awareness of specifically who might have access to their personal data and communications. Yet technical background was not directly associated with more secure behavior online. Almost universally, participants' privacy protective actions or lack of action were informed by personal context and experiences, such as a feeling they had nothing to hide, and in some cases by immediate cues in the online environment such as a security emblem or famous company name. Our work suggests a need for more research into privacy protections that reduce the responsibility on users to understand how the Internet works and to make myriads of privacy protection decisions based on their technical knowledge.

7. ACKNOWLEDGEMENTS

We thank Katherine Mashek and Zachery Mullins for their help conducting interviews, and we thank Sauvik Das, members of the CoEX lab, David Naylor, Peter Steenkiste and other members of the XIA project for their feedback and suggestions. We also thank all the participants for sharing their experiences and thoughts, and the anonymous reviewers for their insightful comments. Support for this work was provided by NSF grants CNS1345305 and CNS1221006.

8. REFERENCES

[1] OpenSSL 'HeartBleed' vulnerability: https://www.us-cert.gov/ncas/alerts/TA14-098A Retrieved March 12, 2015.

[2] Target confirms massive credit-card data breach: http://www.usatoday.com/story/news/nation/2013/12/18/secret-service-target-data-breach/4119337/ Retrieved March 12, 2015.

[3] Ackerman, M. S., Cranor, L. F., & Reagle, J. Privacy in e-commerce: examining user scenarios and privacy preferences. In *Proceedings of the 1st ACM conference on Electronic commerce*. ACM (1999), 1-8.

[4] Acquisti, A., Brandimarte, L., & Loewenstein, G. (2015). Privacy and human behavior in the age of information*. Science,* 347(6221), 509-514.

[5] Bernstein, M. S., Bakshy, E., Burke, M., & Karrer, B. Quantifying the invisible audience in social networks. In Proc. of CHI 2013, ACM (2013), 21-30.

[6] boyd d and Hargittai E (2010) Facebook privacy settings: Who cares? First Monday 15(8). Available at: http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/3086/2589

[7] Bonné, B., Quax, P, & Lamotte, W. Your mobile phone is a traitor! – Raising awareness on ubiquitous privacy issues with SASQUATCH. International Journal on Information Technologies & Security, 3 (2014), 39-53.

[8] Bravo-Lillo, C. C., Downs, L., & J Komanduri, S. Bridging the gap in computer security warnings: A mental model approach. Security & Privacy, IEEE (2011), 18-26

[9] Brown, A., Mortier, R., & Rodden, T. MultiNet: reducing interaction overhead in domestic wireless networks. In Proc. of CHI '13. ACM (2013), New York, NY, USA, 1569-1578.

[10] Corbin, J.M. and Strauss, A.L. Basics of qualitative research: Techniques and procedures for developing grounded theory. Sage Publications, Inc, 2008.

[11] Das, S., Kim, T. H. J., Dabbish, L. A., & Hong, J. I. The effect of social influence on security sensitivity. In Proc. SOUPS (2014).

[12] DiMaggio, P., Hargittai, E., Neuman, W. R., & Robinson, J. P. (2001). Social implications of the Internet. Annual review of sociology, 307-336.

[13] Dommeyer, C. J., & Gross, B. L. What consumers know and what they do: An investigation of consumer knowledge, awareness, and use of privacy protection strategies. *Journal of Interactive Marketing* (2003), 17(2), 34-51.

[14] Ericsson, K. A., & Simon, H. A. Verbal reports as data. Psychological review, 87, 3 (1980), 215-251.

[15] Furnell, S. M., Bryant, P., & Phippen, A. D. Assessing the security perceptions of personal Internet users. *Computers & Security* (2007), *26*(5), 410-417.

[16] Friedman, B., Hurley, D., Howe, D. C., Nissenbaum, H., & Felten, E. Users' conceptions of risks and harms on the web: a comparative study. In *CHI'02 Extended Abstracts*, ACM (2002), 614-615.

[17] Hargittai, E. Survey measures of web-oriented digital literacy. Social Science Computer Review (2005), 23(3), 371-379.

[18] Hargittai, E., & Litt, E. New strategies for employment? internet skills and online privacy practices during people's job search. IEEE security & privacy (2013), 11(3), 38-45.

[19] Hmelo-Silver, C. E., & Pfeffer, M. G. Comparing expert and novice understanding of a complex system from the perspective of structures, behaviors, and functions. Cognitive Science (2004), 28(1), 127-138.

[20] Jacobson, M. J. Problem solving, cognition, and complex systems: Differences between experts and novices. Complexity (2001), 6(3), 41-49.

[21] Jensen, C. and Potts, C. Privacy Policies as Decision-Making Tools: An Evaluation of Online Privacy Notices. In *Proc. of CHI 04,* ACM Press (2004), 471–478.

[22] Jonassen, D. & Cho, Y. H. Understanding Models for Learning and Instruction, chapter Externalizing Mental Models with Mindtools, pages 145–159. Springer US, 2008.

[23] Kang, R., Brown, S., and Kiesler, S. Why do people seek anonymity on the internet?: informing policy and design. In *Proc. of CHI 13*. ACM (2013), New York, NY, USA, 2657-2666.

[24] Klasnja, P., Consolvo, S., Jung, J., Greenstein, B. M., LeGrand, L., Powledge, P., & Wetherall, D. When I am on Wi-Fi, I am fearless: privacy concerns & practices in everyday Wi-Fi use. In Proc. of CHI 2009, ACM (2009), 1993-2002.

[25] Litt, E. Measuring users' internet skills: A review of past assessments and a look toward the future. New Media & Society (2013), 15(4), 612-630.

[26] Nguyen, D. H., Kobsa, A., & Hayes, G. R. An empirical investigation of concerns of everyday tracking and recording technologies. In *Proc. of the 10th Ubicomp*, ACM (2008), 182-191.

[27] Page, K., & Uncles, M. Consumer knowledge of the World Wide Web: Conceptualization and measurement. Psychology & Marketing (2004), 21(8), 573-591.

[28] Park, Y. J. Digital literacy and privacy behavior online. Communication Research (2011), 40 (2), 215-236.

[29] Pew Research Center. What Internet Users Know about Technology and the Web. http://www.pewinternet.org/2014/11/25/web-iq/

[30] Poole, E. S., Chetty, M., Grinter, R. E., & Edwards, W. K. More than meets the eye: transforming the user experience of home network management. In Proc. of DIS 2008, ACM (2008), 455-464.

[31] Potosky, D. The Internet knowledge (iKnow) measure. Computers in Human behavior (2007), 23(6), 2760-2777.

[32] Raja, F., Hawkey, K., & Beznosov, K. Revealing hidden context: improving mental models of personal firewall users. In SOUPS 2009, ACM (2009).

[33] Resnick, M., & Wilensky, U. (1998). Diving into complexity: Developing probabilistic decentralized thinking through role-playing activities. The Journal of the Learning Sciences, 7(2), 153-172.

[34] Sen, S., Joe-Wong, C., Ha, S., & Chiang, M. A survey of smart data pricing: Past proposals, current plans, and future trends. *ACM Computing Surveys (CSUR),* 46, 2 (2013), 15.

[35] Smith, H. J., Dinev, T., & Xu, H. Information privacy research: an interdisciplinary review. MIS quarterly, 35, 4 (2011), 989-1016.

[36] Shklovski, I. A., Mainwaring, S. D., Skúladóttir, H. H., Borgthorsson, H. Leakiness and creepiness in app space:

Perceptions of privacy and mobile app use. In *Proc. of CHI 2014*, ACM (2014), 2437-2356.

[37] Solove, D. J. 'I've got nothing to hide'and other misunderstandings of privacy. *San Diego law review*, *44* (2007), 745-772.

[38] Tbahriti, S., Ghedira, C., Medjahed, B., & Mrissa, M. Privacy-Enhanced Web Service Composition, IEEE Transactions on Services Computing, 7, 2 (2013), 210-222.

[39] Turow, J. Americans & online privacy: The system is broken (2003). Annenberg Public Policy Center, University of Pennsylvania, 3-35.

[40] Ur, B., Leon, P. G., Cranor, L. F., Shay, R., & Wang, Y. Smart, useful, scary, creepy: perceptions of online behavioral advertising. In Proc. of SOUPS, ACM Press (2012).

[41] Vaniea, K. E., Rader, E., & Wash, R. Betrayed by updates: how negative experiences affect future security. In Proc. of CHI 2014, ACM (2014), 2671-2674.

[42] Wash, R. Folk models of home computer security. In Proc. of SOUPS 2010, ACM (2010).

[43] Woodruff, A., Pihur, V., Consolvo, S., Schmidt, L., Brandimarte, L., & Acquisti, A. (2014, July). Would a privacy fundamentalist sell their DNA for $1000... if nothing bad happened as a result? The Westin categories, behavioral intentions, and consequences. In Proc. of SOUPS 2014.

[44] Raja, F., Hawkey, K., & Beznosov, K. Revealing hidden context: improving mental models of personal firewall users. In SOUPS 2009, ACM (2009).

[45] Zhang-Kennedy, L., Chiasson, S., & Biddle, R. (2013, September). Password advice shouldn't be boring: Visualizing password guessing attacks. In eCrime Researchers Summit (eCRS), 2013 (pp. 1-11). IEEE.

[46] Zheng, J., Simplot-Ryl, D., Bisdikian, C., & Mouftah, H. (2011, November). The Internet of Things. IEEE Communications Magazine, 30-31.

APPENDIX A. Prescreen Survey

This survey was given to the technical participants in our study as a prescreen test of their technical knowledge about networking. It was also given to students in a graduate level computer networking class. We computed the scale reliability by combining these two datasets together (the participants in our interview study and the students in the networking class). The 8-item survey had a Cronbach's alpha of 0.61. Question 5 and Question 7 marked with an asterisk had item-total correlations lower than 0.50. After we removed those two items from the scale, Cronbach's alpha for the scale was 0.79 (N = 33). Note: The correct answers are marked in black boxes.

Technical Network Knowledge Scale

1. What is a three-way handshake in TCP/IP?

□ Three or more computers connected and communicating together
■ A method to establish a connection between two computers
□ Three computers on the same LAN or WLAN
□ A deal made between an ISP and a customer regarding Internet service
□ I'm not sure

2. Which of the following protocols work on the Data-Link layer of the OSI Model?

□ SMTP
□ HTTP
□ UDP
■ ARP
□ I'm not sure

3. Which of the following is the correct order for the OSI model layers?

□ Physical, Data Link, Transport, Network, Presentation, Session, Application
□ Physical, Data Link, Network, Transport, Presentation, Session, Application
■ Physical, Data Link, Network, Transport, Session, Presentation, Application
□ Physical, Data Link, Transport, Network, Session, Presentation, Application
□ I'm not sure

4. Which numbers below represent an IP address?

□ 2042.1.6.227
□ 125.120.255
□ 72.1380.12.86
■ 138.5.221.113
□ I'm not sure

*5. Which of the following capabilities does Tor software have?

□ Obscures your data even if someone is monitoring your network
■ Hides the source of a network request
□ Can only be used by domain experts
□ Acts as a VPN
□ I'm not sure

6. Which of these statements about SSL/CAs is NOT correct?

□ CAs can be compromised by attackers
□ A CA is a third party organization
□ A CA issues digital certificates
■ Using trusted certificates from a CA always guarantees the owner's identity
□ I'm not sure

*7. What does the wireless network encryption tool WEP stand for?

■ Wired Equivalent Privacy
□ Wireless Equivalent Privacy
□ Wireless Equivalent Protocol
□ None of the above
□ I'm not sure

8. Of the following choices, what is the best choice for a device to filter and cache content from web pages?

□ Web security gateway
□ VPN concentrator
■ Proxy server
□ MAC filtering
□ I'm not sure

APPENDIX B. Interview Script

Below is the text of our interviewer script along with our primary interview questions. Interviewers read this script to each participant prior to the drawing exercise and then went through the questions prompting the participant to illustrate their thoughts on paper while simultaneously explaining their diagram and thought process. Question 5, 6, and 7 marked with an asterisk were asked for each of the following activities: sending an email;

making a payment online; receiving an online advertisement; browsing a website.

Interviewer:

I'm going to ask you to explain your perceptions and ideas about how the Internet works—keeping in mind how things work "behind the scenes"—when you are doing certain activities online. This is a drawing exercise. I'm going to ask you to draw how you think the Internet works on these papers (hand over pen and papers). Please talk aloud and explain your thought processes while you are drawing.

Please keep in mind that there is no correct answer to these questions—just answer these questions based on your own knowledge and experiences.

1. First off, we'd like to get a picture of how you envision the Internet. Can you draw on this paper and explain for me how you think the Internet works, or how you connect to the Internet?

2. Where do you think your data on the Internet goes? How does your data flow on the Internet?

3. Are there any other people, organizations or companies that can see your connections and activities?

4. Do you do anything to prevent others from seeing your connections and activities?

**5. Please recall an instance when you [watch a YouTube video] on your laptop (or computer). Can you draw and explain for me how you think that works.*

**6. Do you do this same activity on a smartphone? How do you think it works when you are connecting through your smart phone? Is there any difference?*

**7. Is there any example of this system didn't work? Why? Did there anything surprising or unexpected happened? What do you think happened?*

APPENDIX C. Post-test Knowledge Survey

This survey was given to participants in our interview study to assess their technical knowledge of the Internet, privacy and security. It was also given to students in a graduate level computer networking class and MTurk participants in another research study. We computed the scale reliability by combining these three datasets together (the participants in this study, the students in the networking class, and the MTurk participants in another research study). Total N = 432. Cronbach's alpha for *Internet Know-How Self Report Scale* is 0.88. Cronbach's alpha for the *Technical Knowledge of Privacy Tools Scale* is 0.66. Note: The correct answers are marked in black boxes.

Internet Know-how Self Report Scale
How would you rate your familiarity with the following concepts or tools?

	I've never heard of this.	I've heard of this but I don't know what it is.	I know what this is but I don't know how it works.	I know generally how this works.	I know very well how this works.
IP address	□	□	□	□	□
Cookie	□	□	□	□	□
Incognito mode / private browsing mode in browsers	□	□	□	□	□
Encryption	□	□	□	□	□
Proxy server	□	□	□	□	□
Secure Sockets Layer (SSL)	□	□	□	□	□
Tor	□	□	□	□	□
Virtual Private Network (VPN)	□	□	□	□	□
Privacy settings	□	□	□	□	□

Technical Knowledge of Privacy Tools Scale
Please indicate whether you think each statement is true or false. Please select "I'm not sure" if you don't know the answer.

	True	False	I'm not sure
Incognito mode / private browsing mode in browsers prevents websites from collecting information about you.	□	■	□
Tor can be used to hide the source of a network request from the destination	■	□	□
A VPN is the same as a Proxy server.	□	■	□
IP addresses can always uniquely identify your computer.	□	■	□
HTTPS is standard HTTP with SSL to preserve the confidentiality of network traffic.	■	□	□
A request coming from a proxy server cannot be tracked to the original source.	□	■	□

User Perceptions of Sharing, Advertising, and Tracking

Farah Chanchary
School of Computer Science
Carleton University
Ottawa, Canada
farah.chanchary@carleton.ca

Sonia Chiasson
School of Computer Science
Carleton University
Ottawa, Canada
chiasson@scs.carleton.ca

ABSTRACT

Extending earlier work, we conducted an online user study to investigate users' understanding of online behavioral advertising (OBA) and tracking prevention tools (TPT), and whether users' willingness to share data with advertising companies varied depending on the type of first party website. We presented results of 368 participant responses across four types of websites - an online banking site, an online shopping site, a search engine and a social networking site.

In general, we identified that participants had positive responses for OBA and that they demonstrated clear preferences for which classes of information they would like to disclose online. Our results generalize over a variety of website categories containing data with different levels of sensitivity, as opposed to only the medical context as was shown in previous work by Leon et al. In our study, participants' privacy attitudes significantly dominated their sharing willingness. Interestingly, participants appreciated the idea of user-customized targeted ads and some would be more willing to share data if given prior control mechanisms for tracking protection tools.

1. INTRODUCTION

Internet advertising has become increasingly user-sensitive. Advertising networks track users online and create user profiles based on their online activities and preferences without consent from users. These profiles help advertising networks decide which ads are more likely to be of interest to a particular user. The main mechanism for online tracking is third party HTTP cookies by advertising domains [9]. Since users directly interact with first party websites and may be unaware of hidden third parties, the data collection process may presumably violate their online privacy. Previous studies showed that familiarity with advertising companies influenced participants' data sharing willingness [23] and participants' choice to disclose certain classes of information mostly depended on the third parties collecting the data [8]. Leon et al. [8] compared two similar online health/medical websites as the first party to explore how privacy practices of their assigned site might influence participants' data sharing willingness, but the study results did not reveal any significant impact of the first party site. In general, medical information is sensitive and contains many unique characteristics that might make it different from other domains. We re-investigate users' perceptions of OBA based on their interactions with first party websites of varying sensitivity. We further extend our investigation by incorporating the impact of participants' privacy attitude, privacy practices, and technical background.

According to Consumer Action's 2013 survey [5], 69% of consumers were unwilling to allow companies to track them in exchange for a free service or product, and 87% believed they should have the right to control what is collected about them online. A variety of privacy tools are available to control OBA [9]. Some tools use opt-out cookies to store a user's preference not to receive OBA, while other tools transmit *Do Not Track* headers to websites to signal a user's request. These tools are challenging for users to understand [9] and sometimes users cannot properly distinguish between tracking prevention tools and ad blocking tools [1]. We examined users' understanding of TPT and what control features might make them more willing to share information for OBA.

In this online study, we aimed to understand how users experience behavioral advertising online and how their preferences across website categories and privacy control features of TPT influenced their willingness to share data. Using an online survey, we collected responses from 368 participants. Confirming and extending Leon et al.'s work, our participants showed a relatively consistent level of willingness to share personal information with different web sites they visited. Participants with the highest general concern for privacy (Privacy Fundamentalists) were least willing to share any type of information online. Participants with technical (computer or IT related) background showed increased willingness to share information for OBA. User friendly tracking-prevention tool (TPT) features also made participants more inclined to share data. However, having access to view and edit user profiles had only moderate impact on their data sharing willingness. Overall, participants were not interested in paying money to block online tracking or targeted ads. On the contrary, their responses showed that they preferred to see relevant website ads and would

Symposium on Usable Privacy and Security (SOUPS) 2015, July 22–24, 2015, Ottawa, Canada.

share their personal information with online advertisers to receive these ads if they could control what information to share and with whom.

We summarized related work that motivated us to conduct this study in Section 2. In Section 3, we described our study methodology and analysis techniques. In Section 4, we presented study results covering participants' demographic information, understanding of OBA and TPT, willingness to share data online, and other factors that influenced their willingness. Discussion and limitations of our study are in Section 5 and our conclusions are in Section 6.

2. RELATED WORK

Recently, a number of studies have been conducted on the practices of OBA and the usability of privacy tools that allow users to control online advertising. In 2012, Ur et al. [23] presented results of 48 semi-structured interviews where participants found OBA to be simultaneously useful and privacy invasive. They also reported that participants had strong concerns about advertising companies collecting personally identifiable information but their attitudes were context dependent. Participants' willingness to share information varied depending on their familiarity with and the level of advertising activities of these ad companies. In a similar interview-based study by Agarwal et al. [1], the authors reported that the issue of online tracking made users concerned and sensitive about the content in online ads and how the context surrounding their browsing behavior could lead to varying levels of embarrassment. Moreover, the authors mentioned third-party-indifference as a major finding since they did not observe any difference between participants' sensitivities towards the trust levels across third parties. A study by Costante et al. [6] investigated Internet users' perception of the trustworthiness of websites using four types of websites (e-commerce, e-health, e-bank and e-portfolio) and showed that users' perception of trust varied with application domains and users' IT related knowledge. A 2012 survey [18] on the use of search engines showed that despite majority of users viewing these websites as useful and trustworthy, they neither agreed to share search data for receiving personalized results nor were aware of ways to restrict the data collection process. In 2014, Rader [19] examined participants' level of awareness of behavioral tracking and privacy concern based on first party data collection using a social network site (Facebook) and a search engine (Google). Her study results showed that despite having profound knowledge about first party data tracking, participants were much less aware of automatic collection, collaboration and data aggregation across various websites.

Users' lack of knowledge of tracking prevention tools also affect their intentions to adopt suitable privacy practices. McDonald and Cranor [14] found that the majority of American Internet users (86%) were aware of targeted ads but lacked the knowledge to make informed decisions to protect their privacy. The authors also reported users' misconceptions about the purpose of cookies and the effects of clearing them. They highlighted discrepency between people's willingness to pay to protect their privacy and their willingness to accept discounts in exchange for private information. A survey by McDonald and Peha [15] in 2011 also suggested a large gap between the actual implementation of *Do Not Track* in web browsers and what users expected from it, e.g., stopping complete data collection and data aggregation across websites. Leon et al. [9] conducted a laboratory study investigating the usability of nine privacy tools to restrict OBA. Participants misunderstood how these tools worked and mistakenly believed that they were protected against tracking, while in reality they might no longer see targeted ads but continue to be tracked.

Lack of transparency of data collection practices also raised privacy concerns. In a 2006 study, Awad and Krishnan [2] investigated relationships between information transparency features (e.g., data removal and time expirations of data) and consumers' willingness to share information for online personalization. The authors reported that participants who were concerned about these features were less willing to have an online profile. A 2009 survey [22] showed that 92% of users were in favor of a law that requires online advertising companies to delete all stored information about an individual on request. An interesting finding of Leon et al.'s [8] paper was that 52% of participants would be equally or less likely to share data if they were given access to view and edit data collected about them. In a recent paper, Rao et al. [20] explored transparency of data collection practices and accuracy of data in user profiles. They found large number of user profiles with as much as 80% inaccuracy.

Leon et al. [8] presented how users' willingness to share personal information with advertising companies changed depending on these companies' privacy practices. In contrast to Agarwal et al.'s results, participants were mostly concerned with the third party that collected the data, rather than the first-party site. Since the authors explored only a few choices for the first party (i.e., two versions of an online health web site), we investigated in our study whether users truly had no concerns regarding first party tracking. We also explored whether users understood the advantages of using privacy tools or whether they preferred simple ad blocking tools.

3. METHODOLOGY

We conducted a between-subjects online study to investigate users' understanding and preferences for sharing information online. We partially followed the research methodology adopted by Leon et al. [8]. In this section, we describe our recruitment process using the CrowdFlower platform, the research objectives, the structure of our survey questionnaire, and the techniques used to analyze data.

3.1 Recruitment using CrowdFlower

We recruited participants from around the world using an online crowdsourcing service, namely CrowdFlower[1], in two phases. Initially, we recruited 45 participants to ensure the usability of our questionnaire and correctness of the data collection process. In the second phase, we collected responses from 355 participants using a similar procedure. Our recruitment materials indicated that the study would be about how individuals experience the Internet and OBA.

[1] `http://www.crowdflower.com/`

There was no indication that privacy would be one of the research components of this study. Our survey did not collect any sensitive information and all participants remained anonymous. Participants received \$0.50 for completing the survey. This study was approved by our Institutional Research Ethics Board.

3.2 Research Questions

We sought answers to these questions regarding users' understanding and preferences for sharing information online:

Q1. What are participants' current practices, understanding, and perception of OBA and targeted ads?

Q2. Do participants' preferences vary based on categories of first party websites? Is first party more important than the third party?

Q3. Do users' privacy attitudes affect their sharing willingness?

Q4. What features of TPT influence participants' willingness to share?

3.3 Structure of the Questionnaire

Our survey questionnaire was divided into six parts.

1. *Demographic Information*: we collected participants' age, gender, highest level of education, occupation and the amount of time they spent online.

2. *Basic Understanding of Online Advertising*: we asked them to define website advertising, targeted ads, tracking prevention tool, and give their opinions about website advertising and online tracking.

3. *Informational Video*: we provided a link to a short informational video on OBA produced by the *Wall Street Journal* [2] to help them learn how OBA actually works, and we asked them two basic test questions on the concepts of third party cookies and behavioral targeting.

4. *Willingness to Share Information*: we explored participants' willingness to share information online using 5 point Likert-scales (from "Strongly Disagree" to "Strongly Agree"). We used 24 types of information that constituted a subset of 30 types used by Leon et al. [8]. These 24 types were selected because they contained both Personally Identifiable Information (PII) and Non-PII for users, as defined by the Network Advertising Initiative (NAI) [17]. Moreover, these were not related to properties of any specific type of website. For this part only, participants were evenly distributed into four groups and assigned to one type of website services, i.e., an Online Banking site (OB), Online Shopping site (OS), Search Engine (SE) or Social Network site (SN). Participants disclosed their willingness to share information with their assigned first party site. Next, we asked how concerned they were for both first and third party tracking using 5 point Likert-scales (from "Strongly Concerned" to "Strongly Unconcerned"). We also asked whether they would change their preferences if given a fee payment option to control the online data collection process or an option to access their online data for review, edit or deletion.

5. *Understanding of TPT*: we asked for participants' views on TPT, ad blocking tools, and privacy control features that might make them more comfortable with data sharing.

6. *Users' Privacy Attitudes and Practices*: we explored participants' general privacy views, using the Westin Index [21] and their previous privacy practices (e.g., deleting cookies, reading websites' privacy policies, refusing disclosure of sensitive personal information). Finally, we asked for their comments on OBA.

See Appendix B for the full questionnaire.

3.4 Test Questions

Using the two questions from Part 3, we performed a screening test to identify and discard information from participants who were not paying attention. We found 32 participants with incorrect answers (see Q23 and Q24 in Appendix B). All further data analysis used responses from 386 participants who passed both test questions.

3.5 Analysis

We performed statistical tests to identify significant patterns among several data elements collected through our survey questionnaire. All statistical tests were done with R version 3.1.2 and assumed a significance level of $p < 0.05$. We conducted a factor analysis to identify patterns in participants' sharing willingness and group closely related information together. This facilitated our investigation of how participants perceived concerns for online tracking of similar data types. We also employed the Westin Index to categorize participants according to their privacy outlook. We subsequently examined how participants with different privacy attitudes weighted online information disclosure. We first present our factor analysis and Westin Index analysis in this section. Results of these analyses will be used to help answer our research questions in Section 4.

3.5.1 Factor Analysis

To investigate how the categories of websites influenced participants' willingness to share 24 types of information, we performed factor analysis to reduce these 24 types to a smaller number of output variables. Factor analysis is a process that evaluates underlying associations of closely related variables and combines them into a single latent factor. If such underlying factors exist, then further analysis is performed based on these factors instead of the individual variables. A similar process was followed by Leon et al. [8].

Our exploratory factor analyses found that 17 variables could be grouped into 4 factors and the remaining 7 data

[2] http://www.tamingdata.com/2010/10/18/how-advertisers-use-internet-cookies-to-track-your-online-habits/

Table 1: Factor Analysis of willingness to disclose different types of information ($N = 386$). We present Cronbach's α for each resultant factor and the factor loading value for each variable. Percentage of agreement represents those who agreed or strongly agreed to disclose this information.

Factor **(Variables included)**	**Factor Loading**	**Agreement (%)**
Demographic Information **(α=0.897)**	–	**39**
Age	0.68	39
Gender	0.73	54
Weight and Height	0.65	35
Highest level of Education	0.70	44
Religion	0.69	32
Sexual Orientation	0.62	34
Marital Status	0.68	35
Personal Identification & Financial Information **(α=0.906)**	–	**13**
Address	0.76	14
Phone number	0.80	15
SIN/SSN	0.93	10
Credit Card No.	0.87	10
Credit Score Bracket	0.66	17
Location Information **(α=0.905)**	–	**46**
Country	0.67	55
State	0.82	44
Town	0.80	39
Computer Information **(α=0.826)**	–	**39**
Computer's OS	0.74	41
Computer's Browser	0.79	38
Variables that did not conform to any factor	–	–
Hobbies	NA	46
Name	NA	34
Zip code	NA	30
Email address	NA	28
Political preferences	NA	22
Income Bracket	NA	17
Computer's IP Address	NA	16

types did not conform to any particular factor. As in Leon et al. [8], we considered a variable part of a factor if it had a factor loading of at least 0.6 for the particular group, as well as factor loadings under 0.4 for all other groups. We named the resultant factors: *(1) Demographic Information*, *(2) Personal Identification & Financial Information*, *(3) Location Information*, and *(4) Computer Information*. The results of this factor analysis is given in Table 1. We used Cronbach's alpha (α) value for each factor to estimate the internal reliability of the factor analysis test. All four resultant factors had alpha values higher than 0.8, which is the standard to support high correlations between group members. All further analyses considered the four resultant factors. We created an index variable for each factor by averaging participants' responses to all the questions included in the factor.

3.5.2 Westin Index Analysis

The Westin Index [21] is a set of three questions (see Appendix B, Questions 74-76) designed to segment users into three groups: (1) Privacy Fundamentalists, who view privacy as having an especially high value which they feel very strongly about; (2) Privacy Pragmatists, who have strong feelings about privacy but can also see the benefits from surrendering some privacy in situations where they believe care is taken to prevent the misuse of this information; and (3) Privacy Unconcerned, who have no real concerns about privacy or about how other people and organizations use information about them.

The Westin Index has been widely used in the literature to measure users' attitudes towards privacy [4, 10–12]. In 2014, Woodruff et al. [24] argued based on their online survey results that generic privacy attitudes prescribed by the Westin Index did not correlate with individuals' attitudes and behavioral intentions for the protection or disclosure of personal information online. However, we followed the original segmentation index to remain consistent with earlier work since this was not central to our exploration.

Based on the Westin Index, we divided participants into three groups according to their privacy attitudes. We found that 30.4% of our participants were Privacy Fundamentalists, 45.9% were Privacy Pragmatics and 23.6% were Privacy Unconcerned. This conforms to typically observed demographics [21]. We used these groupings to explore how participants' privacy attitudes influenced their sharing willingness. Where appropriate, we further analyzed these correlations according to categories of websites.

4. RESULTS

In this section, we present an analysis of participants' survey responses addressing each of our research questions identified in Section 3.2. We analyzed responses from 386 participants between the ages 18 and 73 (mean=31.7 and σ=9.5). Participant demographics are summarized in Table 2.

4.1 Practices, Understanding & Perception (Q1)

We analyzed participants' responses by measuring their basic level of understanding of online advertising, tracking, and TPT. We asked open-ended questions requesting an explanation of these terms in their own words. We checked each answer and considered it as correct if it contained at least some basic keywords indicating that they understood the concepts. Our study results showed that 55% of participants could define website advertising. Only 6% of participants mentioned that website advertising was beneficial, while others thought it was spam (2%), annoying (6%) and false information (2%). Even though almost half of participants had degrees or work experience in computer related fields, we found that overall awareness about how targeted ads and privacy protection tools work was very low. We asked them to explain how targeted ads worked and 46% had at least partially correct answers. Only 38% of participants could correctly explain how TPT worked.

Table 2: Participants' Demographic Information.

Demographic	Number	Percent
Gender		
Female	99	27
Male	265	72
Decline to answer	4	1
Occupation		
Administrative support	29	8
Art, writing, or journalism	16	4
Business, management, or finance	45	12
Computer engineering	75	20
Education (e.g., Teacher)	26	7
Engineering	18	5
Homemaker	13	4
Service (e.g., retail clerks)	20	5
Skilled labor	21	6
Student	57	16
Unemployed	29	8
Other	12	4
Decline to answer	7	2
Educational Background		
No/Some high school	13	4
High school graduate	75	20
Some college	59	16
Associate's degree	35	10
Bachelor's degree	121	33
Graduate degree	61	17
Decline to answer	4	1
IT Background		
Yes	180	49
No	188	51
Internet Usage (hrs./day)		
1-5	85	23
5-9	149	40
9-13	74	20
13-17	45	12
>17	15	4

4.1.1 Website Ads and Online Tracking

Using 5-point Likert scales (1 = "most negative", 5 = "most positive"), participants expressed their views about different aspects of website advertising. Results are available in Figure 1. Half of participants agreed that website advertising is necessary to enjoy free services on the Internet, 42% found website advertising useful, and 42% believed that website advertising relevant to their interests can save time. However, half also said that they did not normally notice the ads that appeared on the websites that they visited.

Using another 5-point Likert scales (1 = "impossible", 5 = "very common"), participants expressed their understanding of online tracking. Approximately half of participants were aware of the various tracking capabilities. Figure 2 summarizes participants' opinions. Nearly one-fifth of participants believed it was impossible for online tracking systems to track all websites visited, and some wrongly believed that companies did not track individuals' online activities without users' permission (27%).

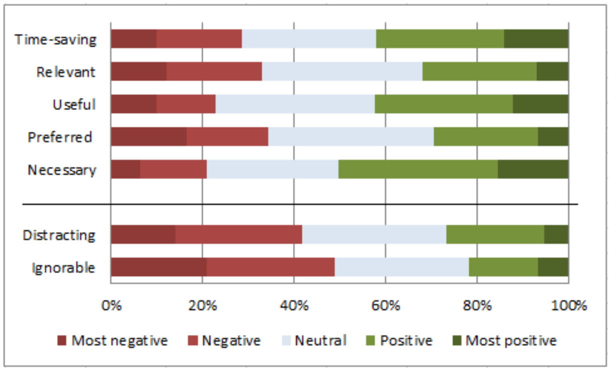


Figure 1: Views on website advertising. Statements included *"In general, I find website advertising..."*.

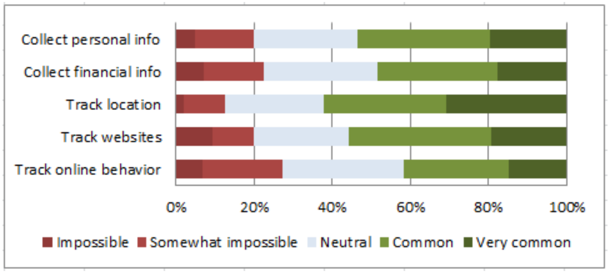


Figure 2: Views on online tracking. "Track websites" and "Track online behavior" are inverted for data presentation.

4.1.2 Targeted Ads

We inquired about users' perceptions of receiving targeted ads based on their online activities. Only 23% of participants liked receiving targeted ads reflecting their online activities, while 37% expressed clear dislike, and the remainder were neutral. In response to our open-ended question, *"Explain what, if anything, would make you feel more comfortable with receiving targeted ads?"*, participants displayed a variety of reactions, including criticisms for currently generated targeted ads. Participants did not perceive relevance or value from targeted ads based on their browsing histories. They saw much more value in seeing ads based on their actual expressed interests. This was clearly articulated by participants in our study: *"Most of the time I get ads that have nothing to do with me, being a girl doesn't mean I'm looking for makeup or trying to get skinny or whatever other stereotyped information that make ads show up"* or *"I'm tired of keep getting ads that I searched over 1 month ago"*.

4.1.3 Current Privacy Practices

We explored participants' previous online behavior to measure how concerned they were about their online privacy in practice. Figure 3 shows that the majority of users (>80%) demonstrated conscious responses to preserve their online privacy either by refusing to provide unnecessary personal information to websites, deleting cookies from web browsers, or terminating online transactions when they were uncertain about the data retention and usage policies. The least popular practice was activating the *Do Not Track* option in web browsers or installing TPT on their computers (58%). We

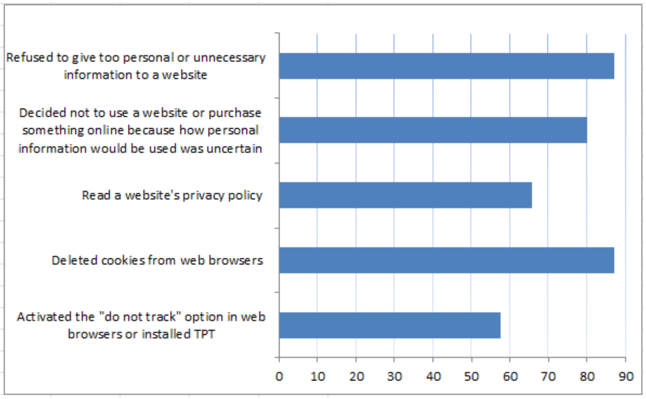


Figure 3: Percentage of participants who have previously employed 5 privacy practices.

found that while users are taking steps to prevent online data leakage, they use only a subset of available safeguards.

4.1.4 *Answer to Q1*

Q1 asks *What are participants' current practices, understanding, and perception of OBA and targeted ads?* Half of our participants were aware of OBA and were actively protecting their online privacy, but most of them were oblivious to the functionalities of TPT. In general, participants were not satisfied with receiving targeted ads based on their online activities. While half of participants appreciated the idea of user-customized targeted ads, half (not mutually exclusive) reported generally ignoring current targeted ads.

4.2 Impact of First and Third Parties (Q2)

This section summarizes participants' willingness to share their information online for the purpose of showing targeted ads on websites.

4.2.1 *Effects of First Party and Data Types*

We were interested in whether Internet users' preferences vary for different types of first party websites. We compared financial websites, shopping sites, search engines, and social networks. We did not find any major differences between website categories, confirmed by statistical analysis.

However, participants do distinguish between different types of information. Figure 4 shows participants' responses for sharing willingness based on a 5 point Likert scale across all websites. We found that participants expressed relatively consistent preferences for 24 types of information across the four website categories. Responses from individual website categories are available in the appendix (Figures 8, 9, 10 and 11). Overall, participants were more willing to share their country (55%), gender (54%), hobby (46%) or state (44%). Few wanted to disclose their credit card number (10%), social identification or security number SIN/SSN (10%), phone number (15%) or exact address (14%). Factor analysis results also uniformly confirms that more participants were willing to share their location information (46%), demographic and computer information (39%) than their personal identification and financial information (13%) (see Table 1). These results are consistent with that of the previous study published by Leon et al. [8] where they used health websites. However, our results confirmed that preferences also holds for variety of first party websites (i.e., financial websites, shopping sites, search engines and social networks).

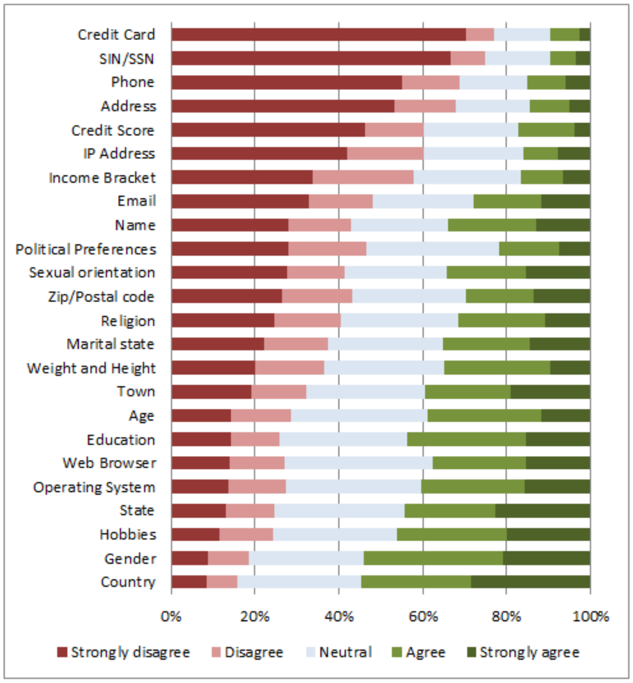


Figure 4: Willingness to disclose to a first party website.

Table 3: Level of Concern for First and Third parties.

Group	Concerned (%)		Unconcerned (%)	
	1^{st} party	3^{rd} party	1^{st} party	3^{rd} party
OB	37	55	20	14
OS	51	53	21	15
SE	42	41	9	20
SN	43	46	20	19

4.2.2 *Concern for First and Third Party Tracking*

We asked participants to express their concern for first party tracking (based on their designated website) and third party tracking without mentioning the name of any particular third party (see Appendix B, Questions 54 - 55). As shown in Table 3, only participants of the online banking (OB) group expressed increased concern (55%) for third party tracking compared to their online banking sites (37%). We suggest two possible reasons for this result. First, online banking sites generally do not show a large number of online ads compared to other sites so any ads may be viewed suspiciously. Secondly, users manage highly sensitive financial data through OB sites, and wish to avoid third party tracking of such data. In general, participants from other groups expressed approximately equal levels of concern for both first and third parties.

4.2.3 Answer to Q2

Q2 asks *Do participants' preferences vary based on categories of first party websites? Is first party more important than the third party?* There was no significant difference between first parties on participants' data sharing willingness. And except for online banking (OB), participants were equally concerned between first and third party tracking.

4.3 Impact of Privacy Attitudes (Q3)

We found that participants' privacy attitudes had significant impact on their willingness to share data. We used the three categories of participants derived from the Westin Index (see Section 3.5.2) for analyzing data in this section.

Table 4 shows all significant differences between Privacy Fundamentalists and other participants (i.e., Privacy Pragmatics and Privacy Unconcerned). Overall Privacy Fundamentalists were less willing to share their demographic data, and their personal identification information and financial information. Each row of this table represents one set of differences. For example, the first row of the table represents a significant difference in overall sharing willingness for demographic information among all participants (Kruskal-Wallis, $N = 386, \chi^2(2) = 18.125, p = 0.001$). Pairwise comparisons using Wilcoxon rank sum test (PW) show that more Privacy Fundamentalists were unwilling (16%) to disclose demographic information than Privacy Pragmatics ($8\%, p = 0.001$) and Privacy Unconcerned ($7\%, p = 0.001$).

4.3.1 Answer to Q3

Q3 asks *Do users' privacy attitudes affect their sharing willingness?* Participants' privacy attitudes significantly affected their data sharing willingness for two out of four overall factors: personal identification, financial and demographic data. In all cases, Privacy Fundamentalists showed the least interest to share any type of information.

4.4 Impact of TPT Features (Q4)

In this section, we explored what privacy control features of TPT might influence participants' willingness to share their data.

4.4.1 Usefulness of TPT

Agarwal et al. [1] mentioned that users were unsatisfied with mechanisms that only control tracking or OBA. Rather, users demanded selective filtering of ad contents. We presented hypothetical tools with specific features (see Table 5), generally matching to TPT and ad blocking tools (ABT) (without specifically mentioning their names), and asked participants to rate these tools with a 5-point Likert Scale (1 = "least useful", 5 = "most useful"). The description for TPT explained that it would control third party tracking on selected topics and hide related targeted ads, but not generic ads. The description for ABT explained that it would block embarrassing or irrelevant ads selected by participants, but would not stop third-party tracking of online activities.

As shown in Figure 5, 55% of participants thought TPT was useful; in comparison, only 37% found the ad blocking tool useful. We asked them which tool they preferred and 72% of participants chose TPT.

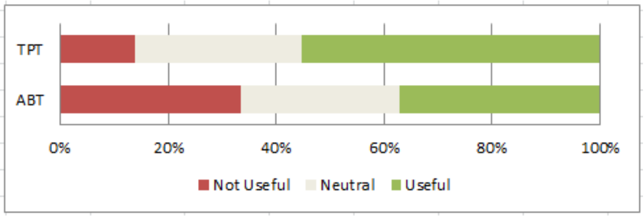


Figure 5: Participants' opinion of TPT and ABT.

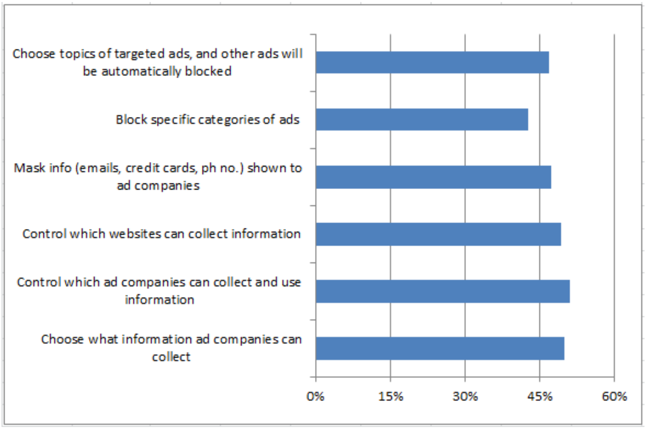


Figure 6: Percentage of participants more willing to share if six control features were provided by privacy tools.

4.4.2 Control Features of TPT

Current privacy protection tools for controlling OBA differ significantly from one another [9]. For example, opt-out tools only block particular advertising networks from showing targeted ads based on users' browsing behavior. The *Do Not Track* browser plugin attempts to block both first or third party cookies by sending a DNT header to visited websites. General blocking tools provide a range of options, including selectively blocking/unblocking groups of ad companies, setting opt-out cookies for ad networks, and installing filter subscriptions maintained by third parties to block websites. In this section, we investigate what features would increase participants' willingness to share information.

We presented six hypothetical control features to participants. Figure 6 shows the control features and the percentage of participants who would be more willing to share if each feature was available. Half of users personally want to control which sites can collect information (regardless of whether they are first or third parties). They also want to control which types of information to share (50%) and want the ability to customize targeted ads (47%). We further analyzed whether Privacy Fundamentalists were more inclined to adopt these tools (TPT/ABT) and found that participants' privacy attitudes had no significant impact on their preference for these tools.

Table 4: Statistical results comparing Fundamentalists (F), Pragmatics (PR) and Unconcerned (U) participants' willingness to share. The % *Unwilling* column represents the percentage of participants from each group who were unwilling to share the specified *Factorized data*. PW=p-value in Wilcoxon rank sum test pairwise comparisons with Bonferroni correction, n.s.=no significant difference. Only factors with significant results are included.

Factorized	**Group**	**% Unwilling**			**Kruskal-Wallis**			**PW**	
data		F	PR	U	N	χ^2	p	F - PR	F - U
Demographic Information	Overall	16	8	7	368	18.125	0.001	0.001	0.001
	SE	12	12	0	92	8.847	0.01	n.s.	0.01
Personal ID & Financial Information	Overall	55	36	21	368	47.189	0.001	0.001	0.001
	OB	73	32	25	91	8.403	0.01	n.s.	0.01
	OS	52	28	29	94	9.465	0.01	0.01	0.05
	SE	39	42	17	92	17.211	0.001	0.005	0.001
	SN	57	43	13	91	15.745	0.001	n.s.	0.001
Location Information	OB	35	10	21	91	8.999	0.01	0.01	0.05
PC Information	SN	37	21	4	91	6.342	0.05	n.s.	0.05

Table 5: Features offered by TPT and ABT.

Feature	**TPT**	**ABT**
Control third party tracking	Yes	No
Hide targeted ads	Yes	No
Hide generic ads	No	No
Block embarrassing ads	No	Yes
Block irrelevant ads	No	Yes
Selection option available	Yes	Yes

4.4.3 Answer to Q4

Q4 asks *What features of TPT influence participants' willingness to share?* Participants clearly distinguished between TPT and ABT, and the majority considered TPT more useful than ABT. Nearly half of participants, across all websites and irrespective of their privacy attitudes, were more willing to share data if they could restrict both first and third parties from collecting data, select types of information to share, and customize topics of targeted ads.

4.5 Other Factors Affecting Willingness to Share

Sharing willingness might depend on many other factors apart from website categories and privacy attitudes. So we conducted post-hoc exploration of a few other options and report the results.

4.5.1 Frequency of Website Visit

Frequency of visiting a particular type of website significantly influenced overall willingness to share for location data (Kruskal-Wallis test: $N = 386, \chi^2(5) = 11.936, p < 0.05$) and personal identification & financial data (Kruskal-Wallis test: $N = 92, \chi^2(5) = 19.559, p = 0.001$). Pairwise comparisons using Wilcoxon tests showed that frequent visitors (daily visits) were more willing (34%) to share location information than infrequent visitors (12%). We also found that frequent visitors of search engines (SE) were less likely to share personal identification information and financial data than who visited SE websites only a few times in the last year.

Many websites, like shopping sites and search engines, provide location-based selection or search facilities for their client services. Frequent Internet users might perceive this as a useful feature and hence be more willing to share these data. However, financial (e.g., credit card number) or personal identification data (e.g., SIN/SSN) are too sensitive and frequent users appear aware of the risk of online exposure, thus oppose disclosure of this information.

4.5.2 Computer Related Background

Participants' computer or IT related background had significant impacts on sharing willingness (IT = technical background, non-IT = with no technical background). Wilcoxon tests revealed that IT participants were significantly more willing to share their personal identification data & financial information ($W = 12367.5, p = 0.001$) and computer related information ($W = 14704, p < 0.05$) than the non-IT users. Study results showed that 42% of non-IT participants refused to share personal identification & financial information compared to 27% of IT participants. Similarly, 13% of non-IT participants refused to share computer related data compared to 5% who had computer related background. We may assume that people with degrees or work experience in computer related fields are more confident in their abilities to handle the risk of information leaking and thus are more willing to share these data.

4.5.3 Intentions to Explore Online Ads

Some of our participants expressed interests in exploring online ads by clicking links on websites. We identified significant impact of this *intent to explore* on participants' concerns for receiving targeted ads (Wilcoxon rank test: $N = 386, W = 8341, p = 0.001$). More users who clicked links to explore online ads (25%) would like to receive targeted ads based on their online activities than users who did not explore ads (17%).

We further found significant impact of this intention on participants' concern for third party tracking (Wilcoxon rank sum test: $N = 386, W = 9462, p < 0.05$). Users who clicked links to explore online ads (52%) knew that they might be at risk and showed increased concern for third party tracking compared to users who did not explore ads (38%). However, participants' intentions to explore ads did not influence their concern for first party tracking.

4.5.4 *Access to Collected Data*

We next found that data retention policies had moderate impact on the sharing willingness. We proposed three hypothetical scenarios to participants. One scenario was based on users' access to collected data and two others were based on fee payment by Internet users to control online information tracking.

We asked participants whether they would change their willingness to share if given an option for having access to collected data for reviewing, editing or even permanently deleting from the online platforms. We found that 25% of participants were more willing to share information if they were given this access. They specifically mentioned that they would be in favor of targeted ads based on their online activities if they were in control of selecting what data could be used for generating these ads.

To increase data collection transparency, some companies recently allow users to access and edit their online profiles [8]. Some companies provide users access to their profiles based on browser cookies (e.g., Google [7], Yahoo! Pulse [25] and BlueKai [3]), while others like Microsoft [16] provide users access to information through a privacy dashboard that requires users to create an account with them [20]. 25% of participants felt this option was acceptable and became more willing to share information.

Interestingly, the majority of participants did not change their sharing willingness. Some participants with negative views expressed privacy related concerns such as "*I don't like my private info to be on the Internet, it's just for me*", concerns relating to time costs, "*Who has time for that. I don't want information collected about me, period. I'm supposed to do that for every website I visit? Craziness..*". Some participants did not attribute much value to targeted ads, "*I don't really care if my information are correct or not, if it is for ad purpose*". Many participants would not trust this mechanism, "*They should not collect that information on first place without our consent. Even if I wanted to remove I wouldn't trust them to actually discard my data*".

4.5.5 *Fee Payment*

We presented two fee payment options (Scenario 1 & 2) to our participants to measure the extent of their interest in controlling online information tracking.

Scenario 1: Their favorite websites would charge a monthly fee in exchange for not showing any ads, but companies might still collect information from users for other purposes.

Scenario 2: This payment method would stop advertising companies from collecting any information about users' online activities on the website but display general ads.

The majority of our participants were unwilling to pay to stop targeted ads (61%) or online tracking (51%). Overall responses for each scenario are shown in Figure 7. This result supports our findings from Section 4.4.1 and also matches with results published in Leon et al. [8].

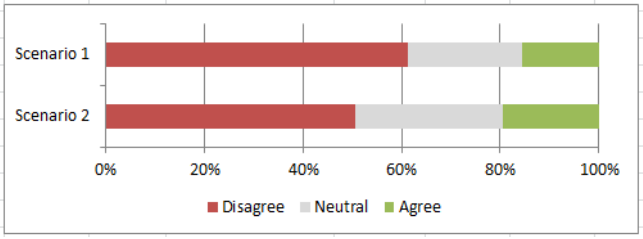


Figure 7: Percentage of participants who are willing to payment for controlling targeted ads and online tracking.

4.6 Summary of Results

To summarize overall results, we list all the factors examined in this study and their relative impact on participants' sharing willingness in Table 6. We identified four factors that greatly influenced participants willingness to share various types of PII and non-PII data: (1) participants' privacy attitudes, (2) frequency of visiting a specific type of website, (3) having technical background, and (4) intention to explore online ads. The choice of first party websites had no impact on participants' data sharing willingness, suggesting that Leon et al.'s findings [8] may be generalizable. Our participants also showed preferences for the types of data they were willing to share online.

Some factors influenced a subset of our participants, such as options that allowed access to participants' user profiles for performing necessary modification, and TPT features to restrict data collection or to select topics for targeted ads.

Table 6: Factors affecting participants' sharing willingness.

Factors	**Impact Level**	**Section**
First party websites	None	4.3
Control features of TPT	Moderate	4.5.2
Access to collected data	Moderate	4.6.4
Privacy attitude	High	4.4
Frequency of website visit	High	4.6.1
Computer/IT background	High	4.6.2
Exploring Online ads	High	4.6.3

5. DISCUSSION

Leon et al. [8] investigated the impact of privacy practices using two health-themed first party websites (i.e., a familiar online medical site and a fictitious online medical site) on participants' willingness to share data. They suggested that hidden third party tracking was more important than site familiarity for users' willingness to disclose information online. Other studies demonstrated that participants considered companies' non-OBA related activities when deciding whether to allow data collection [23] and that participants' showed indifference towards third parties when sharing information online [1].

Confirming and extending these prior studies, we investigated users' sharing preferences across different first party website categories (banking sites, shopping sites, search engines and social networking sites) for the purpose of receiving targeted ads. We carefully selected a range of first party sites that would be familiar to users and that would cover scenarios with data of varying sensitivity. We found that the type of first party website had no major impact on participants' willingness to sharing. Furthermore, participants expressed equal concern for both first and third party tracking. However, we confirm that participants' privacy attitudes significantly influenced their sharing willingness. In general, participants with strong concern for privacy were unwilling to disclose personal, financial and demographic data for any type of website. These types of data are considered sensitive by NAI [17] and therefore, should only be collected with users' consent. Consent mechanisms should offer some assurance that opt out preferences are being observed. Other types of data were also of concern to smaller segments of the population; providing opportunity to voice a preference would also be beneficial in these cases.

In line with the results of Leon et al.'s study [8], participant responses clearly showed that some data items can be openly shared for OBA (e.g., 46% of the participants agreed to share hobbies, 55% for country, 54% for gender, and 43% for education), but these are user-specific. Advertising companies can maintain user profiles combining these details with the categories of ads preselected by users. A significant number of our participants were open to targeted ads, as long as they had some control over what information is being collected for their profile.

While a number of studies individually investigated users' concerns towards online behavioral advertising [14,23], users' understanding of tracking prevention tools [9] and preferences for ad blocking tools to control embarrassing ads rather than third party tracking [1], we combined exploration of participants' level of understanding of OBA and TPT, and preferences over the types of tools (TPT and ABT) and their control mechanisms. We also found that participants' having computer related background or a strong preference for online ads were more willing to share information online. Furthermore, sharing willingness of frequent website visitors varied significantly based on website categories. It would be interesting to further investigate the group-wise usage patterns to find what makes them more inclined to share.

As users' agitation about seeing embarrassing online ads had been emphasized by Agarwal et al. [1], we investigated users' preferences for tools specifically to block embarrassing or irrelavant ads. Our study results indicated that most users were more concerned over online tracking than blocking unwanted ad networks or topics. We assume that the definition of embarrassment is sensitive to both geographical location and culture, and users' concerns on this topic needs further investigation.

Current control mechanisms of privacy protection tools are controversial and have poor usability [9, 15]. Our hypothetical TPT features showed increase in participants' sharing willingness (>43%) across all categories of websites. As expected, half of participants prefered features that would provide them control over the types of collectable data as well as over the data collecting entities. The main downside of today's OBA mechanism is that it creates users' profiles based on their browsing histories and not on their actual interests in seeing ads [1]. For example, 47% of our participants would share more data online if they could choose topics for targeted ads. Therefore we suggest a more open privacy-choice mechanism for OBA, which would communicate with users regarding data collection by asking their preferences instead of showing ads that might surprise or annoy them based on users' general profiles. Rather than alienating users through "creepy" OBA practices, companies may be better served in starting a dialog with users and collecting information that users are comfortable revealing.

5.1 Limitations

The main limitation of our study is the data we collected are self-reported values based on participants' views towards OBA and perceived willingness to share personal information in hypothetical scenarios. From our data, we are unable to confirm how well this maps to users' actual behavior. In our recruitment notice, we intentionally avoided explicit mention of privacy, but the study design might have influenced responses nonetheless. These limitations are common with several other related studies available in the literature.

6. CONCLUSIONS

We conducted an online survey using CrowdFlower to investigate whether participants' willingness to share 24 types of personal information with online advertising companies varied depending on the type of first party websites they visited. We also explored users' understanding of online behavioral advertising and tracking prevention tools. Furthermore, we investigated how other aspects, such as participants' privacy attitudes, practices and features of privacy protection tools influenced their sharing willingness. Our work confirms and extends previous work, such as Leon et al.'s study exploring only one type of first party website.

We found that half of participants were well informed about OBA and the majority demonstrated at least some activities to protect their online privacy. However, their overall awareness about tracking prevention was low. Participants expressed clear preferences for which classes of data they were willing to share and these were mostly consistent regardless of which first party site was visited. In fact, the type of first party website had no significant impact on users' decisions. Our results generalize over several types of first-party sites where users would typically disclose data of varying sensitivity. Moreover, participants were similarly worried about first and third party tracking. We confirm significant differences in sharing willingness based on privacy attitudes (Westin Index), with Privacy Fundamentalists being most concerned. Our participants appreciated the idea of user-customized targeted ads and some would be more willing to share if given prior control mechanisms to specify which information can be collected by whom, and what types of targeted ads they wish to see. We recommend active involvement of users in decision-making about OBA and targeted ads.

7. ACKNOWLEDGEMENTS

Sonia Chiasson holds a Canada Research Chair in Human Oriented Computer Security. She acknowledges the Natural Sciences and Engineering Research Council of Canada (NSERC) for funding the Chair and her Discovery Grant.

8. REFERENCES

[1] L. Agarwal, N. Shrivastava, S. Jaiswal, S. Panjwani, Do Not Embarrass: Re-Examining User Concerns for Online Tracking and Advertising, In *Proc. Symposium on Usable Privacy and Security (SOUPS)*, July 24-26, 2013.

[2] N. F. Awad and M. Krishnan. The personalization privacy paradox: An empirical evaluation of information transparency and the willingeness to be profiled online for personalization. In *Management Information Systems Quarterly*, 30(1), 2006.

[3] The BlueKai Registry, `http://bluekai.com/registry/`, Accessed: March 2015.

[4] S. Consolvo, I. E. Smith, T. Matthews, A. LaMarca, J. Tabert, and P. Powledge. Location disclosure to social relations: why, when, & what people want to share. In *Proc. of the SIGCHI Conference on Human Factors in Computing Systems*, pages 81-90. ACM, 2005.

[5] Consumer Action "Do Not Track" Survey Results, `http://www.consumer-action.org/downloads/english/Summary_DNT_survey.pdf`, Accessed: February 2015.

[6] E. Costante, J. den Hartog, and M. Petkovic, On-line trust perception: What really matters. In *Proc. STAST*, 2011.

[7] Google Ads Settings, `https://www.google.com/settings/u/0/ads`, Accessed: March 2015.

[8] P. G. Leon, B. Ur, Y. Wang, M. Sleeper, R. Balebako, R. Shay, L. Bauer, M. Christodorescu, L. F. Cranor, What Matters to Users? Factors that Affect Users' Willingness to Share Information with Online Advertisers, In *Proc. Symposium on Usable Privacy and Security (SOUPS)*, July 24-26, 2013.

[9] P. G. Leon, B. Ur, R. Balebako, L. F. Cranor, R. Shay, Y. Wang, Why Johnny Can't Opt Out: A Usability Evaluation of Tools to Limit Online Behavioral Advertising, In *Proc. CHI 2012*, ACM Press, 589-598, 2012.

[10] Kumaraguru and L. F. Cranor. Privacy Indexes: A Survey of Westin's Studies. *Technical report, Carnegie Mellon University CMU-ISRI-5-138*, 2005.

[11] M. Kwasny, K. Caine, W. A. Rogers, and A. D. Fisk. Privacy and technology: Folk definitions and perspectives. In *CHI'08 Extended Abstracts on Human Factors in Computing Systems*, pages 3291-3296. ACM, 2008.

[12] M. Malheiros, S. Preibusch, and M. Sasse. 'fairly truthful': The impact of perceived effort, fairness, relevance, and sensitivity on personal data disclosure. In *Proc. of the 6th International Conference on Trust & Trustworthy Computing (TRUST 2013)*, pages 250-266, 2013.

[13] J. R. Mayer and J. C. Mitchell, Third-party web tracking: Policy and technology. In *IEEE Symposium on Security and Privacy*, 2012.

[14] A. M. McDonald and L. F. Cranor, Americans' Attitudes About Internet Behavioral Advertising Practices. In *Workshop on Privacy in the Electronic Society*, October 4, 2010.

[15] A. McDonald and J. Peha. Track gap: Policy implications of user expectations for the "Do Not Track" internet privacy feature. In *Information Privacy Law eJournal*, 5, 2012.

[16] Microsoft. Microsoft personalized ad preferences. `https://choice.microsoft.com/en-US/opt-out`, Accessed: March 2015.

[17] NAI Code of Conduct 2013, `http://www.networkadvertising.org/2013_Principles.pdf`, Accessed: March 2015.

[18] K. Purcell, J. Brenner, and L. Rainie. Search engine use 2012. In *PewResearchCenter Technical Report*, March 2012.

[19] E. Rader, Awareness of Behavioral Tracking and Information Privacy Concern in Facebook and Google, In *Proc. Symposium on Usable Privacy and Security (SOUPS)*, July 9-11, 2014.

[20] A. Rao, F. Schaub, N. Sadeh, What do they know about me? Contents and Concerns of Online Behavioral Profiles, In *Proc. of ASE International Conference on Privacy, Security, Risk and Trust (PASSAT, 2014)*, December 14-16, 2014.

[21] H. Taylor, Most People are "Privacy Pragmatists" Who, While Concerned about Privacy, Will Sometimes Trade It Off for Other Benefits. Harris Interactive, 2003.

[22] J. Turow, J. King, C. J. Hoofnagle, A. Bleakley, and M. Hennessy. Americans reject tailored advertising and three activities that enable it. `http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1478214`, 2009.

[23] B. Ur, P. G. Leon, L. F. Cranor, R. Shay, Y. Wang, Smart, Useful, Scary, Creepy: Perceptions of Online Behavioral Advertising, In *Proc. Symposium On Usable Privacy and Security (SOUPS)*, July 11-13, 2012.

[24] A. Woodruff, V. Pihur, S. Consolvo, L. Schmidt, L. Brandimarte and A. Acquisti, Would a privacy fundamentalist sell their DNA for $1000... if nothing bad happened as a result? The Westin categories, behavioral intentions, and consequences, In *Proc. Symposium on Usable Privacy and Security (SOUPS)*, July 9-11, 2014.

[25] Yahoo Ad Interest Manager. `https://info.yahoo.com/privacy/us/yahoo/opt_out/targeting/details.html`, Accessed: March 2015.

APPENDIX

A. PARTICIPANTS' SHARING WILLINGNESS

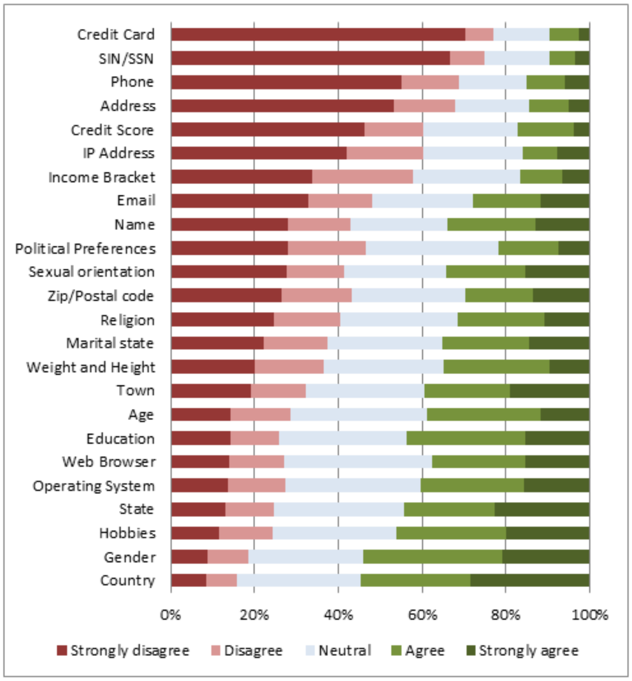


Figure 8: Willingness to disclose information with an online banking website.

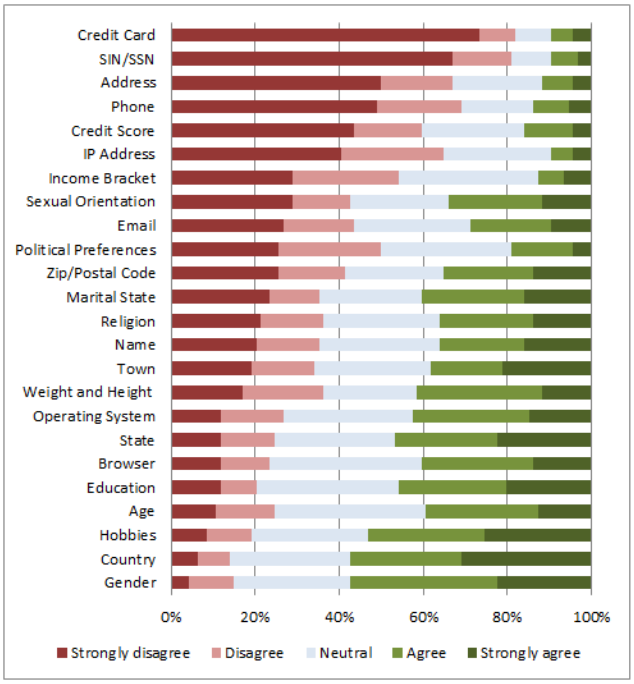


Figure 9: Willingness to disclose information with an online shopping site.

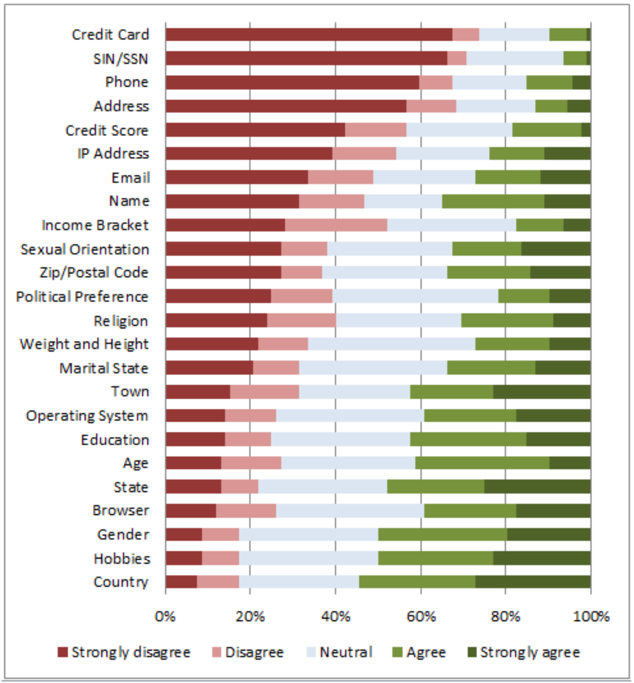


Figure 10: Willingness to disclose information with a search engine.

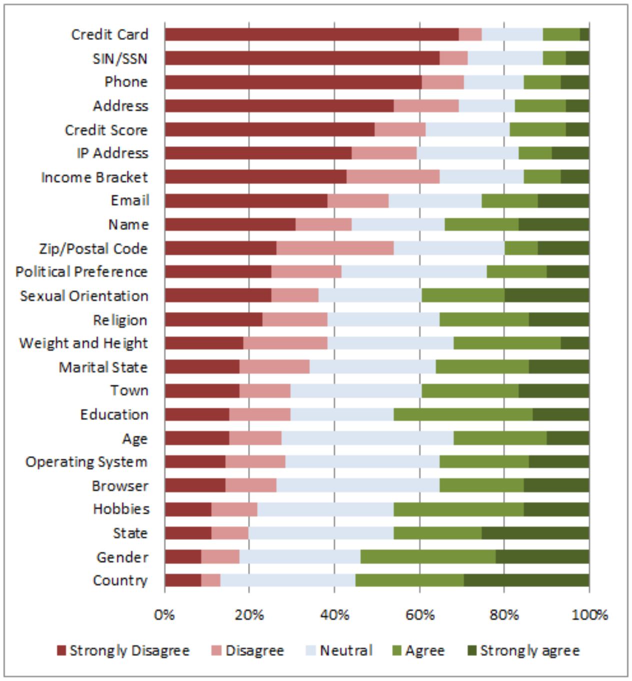


Figure 11: Willingness to disclose information with a social networking site.

B. SURVEY QUESTIONNAIRE

Please think thoroughly before answering each question. Your precise responses are very important for us. You may give an incomplete answer or say you do not know.

Part 1 - Demographic Information

In this part of the questionnaire we collect some demographic information. You can always decline to answer should you feel uncomfortable with a question.

Q1. What is your gender?
( ) Male ( ) Female ( ) Decline to answer

Q2. What is your age (in years)?

Q3. Which of the following best describes your primary occupation?
( ) Administrative support (e.g., secretary, assistant)
( ) Art, writing, or journalism (e.g., author, reporter)
( ) Business, management, or financial (e.g., manager, accountant, banker)
( ) Computer engineer or IT professional (e.g., systems administrator, programmer, IT consultant)
( ) Education (e.g., teacher)
( ) Engineer in other fields (e.g., civil engineer, bio-engineer)
( ) Homemaker
( ) Legal (e.g., lawyer, law clerk)
( ) Medical (e.g., doctor, nurse, dentist)
( ) Scientist (e.g., researcher, professor)
( ) Service (e.g., retail clerks, server)
( ) Skilled labor (e.g., electrician, plumber, carpenter)
( ) Student
( ) Unemployed
( ) Decline to answer

Q4. Which of the following best describes your highest achieved education level?
( ) No high school
( ) Some high school
( ) High school graduate
( ) Some college
( ) Associates/2 year degree
( ) Bachelors/4 year degree
( ) Graduate degree - Masters, PhD, professional, etc.
( ) Decline to answer

Q5. Do you have a college degree or work experience in computer science, software development, web development or similar computer-related fields?
( ) Yes ( ) No

Q6. Approximately how many hours do you spend on the Internet each day?
( ) None ( ) Fewer than 1 ( ) Between 1 and 5 ( ) Between 5 and 9 ( ) Between 9 and 13 ( ) Between 13 and 17 ( ) More than 17

Part 2 - Basic Understanding

We are interested in understanding how you experience things online. We will start with some questions that seek your views about website advertising. Here, "website advertising" refers to ads that are displayed on the web pages that you visit but it excludes pop-up windows or advertising sent over email.

Q7. In your own words, define website advertising.

Q8. In your own words, describe how targeted ads work.

Q9. In your own words, describe how Tracking Prevention tools work.

How much do you agree or disagree with the following statements?
(5 Point Likert-Scale from "Strongly disagree" to "Strongly agree")
Q10. Website advertising is necessary to enjoy free services on the Internet.
Q11. In general, I like website advertising.
Q12. In general, I find that website advertising is useful.
Q13. In general, I find that website advertising is distracting.
Q14. In general, I find website advertising to be relevant to my interests.
Q15. In general, I find that website advertising relevant to my interests can save my time.
Q16. I usually don't look at the ads that appear on the websites that I visit.

Q17. Have you ever clicked on an ad that appeared on a website to get more information about the advertised product? (Yes/No)

How common are the following scenarios?
(5 Point Likert-Scale from "Impossible" to "Very common")

Q18. Companies collect detailed personal information about individuals, such as health conditions, without telling them.
Q19. Online companies collect detailed financial information about individuals, even when they are not purchasing something online.
Q20. Companies track individuals' locations when they are using a mobile phone.
Q21. Online tracking systems cannot follow an individual to all websites he has visited.
Q22. Companies do not track where individuals go and what they do online without their permission.

Part 3 - Willingness to Share

Please read this information carefully, then answer the questions below.

Many websites contract with online advertising companies. The advertising companies pay websites for every ad they show, allowing the websites to provide free services to its visitors. Clicking on the link below will open a new tab or window in your browser displaying a short video explaining how Online Behavioral Advertising (OBA) works. Please watch the video at your own pace and then based only on the information that you have learned from the video choose the correct answer for the following questions.
`http://www.tamingdata.com/2010/10/18/how-advertisers-use-internet-cookies-to-track-your-online-habits/`

Q23. A cookie is ...
(a) a software for browsing Internet
(b) a textfile that contains a ID number to recognize users
(c) your username for a website

Q24. Behavioral targeting ...
(a) tracks visitors' activities using third party cookies from

different websites
(b) does not create profiles for specific visitors
(c) uses information only from the original website visited by a user

Please answer the questions below indicating what information you would allow Advertising Companies to collect for the purpose of showing you targeted ads.

Q25. How often have you visited your favourite [Online Banking Website/Online Shopping site/Search engine/Social network] in the last 12 months?
( ) None ( ) Only once ( ) A few times ( ) A few times per month ( ) A few times per week ( ) A few times per day

Based on the information you know about OBA now, please indicate what information you would allow your [Online Banking Website/Online Shopping site/Search engine/Social network] to collect for the purpose of showing you targeted ads on any website.

(5 Point Likert-Scale from "Strongly disagree" to "Strongly agree")

Q26. My age
Q27. My gender
Q28. My weight and height
Q29. My highest level of education
Q30. My income bracket
Q31. My religion
Q32. My political preferences
Q33. My sexual orientation
Q34. My marital status
Q35. My hobbies
Q36. My credit score bracket
Q37. My country
Q38. My state / province
Q39. My town or city
Q40. My zip code / postal code
Q41. My exact address
Q42. My name
Q43. My email address
Q44. My phone number
Q45. My social security number / social insurance number
Q46. My credit card number
Q47. My computer's operating system
Q48. My computer's IP address
Q49. My web browser

Q50. Will you change your willingness to share if a website allows you to review, edit and delete the information collected about you? For example, you now have the option to confirm that your information and preferences are accurate and remove information that you no longer feel comfortable sharing.

If your favourite [Online Banking Website/Online Shopping site/Search engine/Social network] allows editing of your info, you will be...
( ) less willing to share
( ) equally willing to share
( ) more willing to share

Q51. Please explain the reason(s) for your answer in Q50.

Q52. Suppose your favourite [Online Banking Website/Online Shopping site/Search engine/Social network] offers you the opportunity to pay a monthly fee in exchange for not showing you any ads, but advertising companies may still collect information from you for other purposes. To what extend would you agree to pay? You will pay a fee for your favourite [Online Banking Website/Online Shopping site/Search engine/Social network] to hide ads but still collect your info.

(5 Point Likert-Scale from "Strongly disagree" to "Strongly agree")

Q53. Suppose your favourite [Online Banking Website/Online Shopping site/Search engine/Social network] offers you the opportunity to pay a monthly fee in exchange for stopping advertising companies from collecting any information about you or your online activities on this website, to what extend would you agree to pay? You will pay a fee for your favourite [Online Banking Website/Online Shopping site/Search engine/Social network] to stop collecting your info but display general ads.

(5 Point Likert-Scale from "Strongly disagree" to "Strongly agree")

Q54. How concerned are you for first party tracking (first party tracking means the website you are currently visiting is collecting information about you)?

(5 Point Likert-Scale from "Least concerned" to "Most concerned")

Q55. How concerned are you for third party tracking (third party tracking means a website having some sort of contracts with the first party is collecting information about you)?

(5 Point Likert-Scale from "Least concerned" to "Most concerned")

Q56. What do you consider the main benefit, if any, of receiving ads that are targeted based on your online activities?

Q57. What do you consider the main downside, if any, of receiving ads that are targeted based on your online activities?

Q58. Overall, how do you feel about receiving ads that are targeted based on your online activities?

(5 Point Likert-Scale from "Strongly dislike" to "Strongly like")

Q59. Explain what, if anything, would make you feel more comfortable with receiving targeted ads?

Part 4 - Understanding about Tracking Prevention Tools
Imagine you have two tools that you could install on your browser.

Q60. Tool-A protects you from being tracked online on particular topics. It will control online tracking on the topics you select and hide related targeted ads. However, it will show generic ads. How useful is this tool?

(5 Point Likert-Scale from “least useful” to “very useful”)

Q61. Tool-B blocks ads which you find embarrassing or irrelevant, but it does not stop tracking of your activities. How useful is this tool?

(5 Point Likert-Scale from “least useful” to “very useful”)

Q62. If you have to pick only one, which tool would you choose?
a) Tool A
b) Tool B

Please state how much you agree or disagree with the following statements.

(5 Point Likert-Scale from “Strongly disagree” to “Strongly agree”)

I would be more willing to share my personal information for the purpose of receiving targeted ads if tracking prevention tools...
Q63. allowed me to choose ahead of time what information advertising companies can learn about me
Q64. allowed me to control which advertising companies can collect and use my information
Q65. allowed me to control on which websites my information can be collected
Q66. allowed me to mask my information (accounts, emails, credit cards, phone numbers) to show to these advertising companies at different points in time
Q67. allowed me to block some specific categories of ad.
Q68. allowed me to choose some topics to see targeted ads, and other ads will be automatically blocked.

Part 5
Please indicate whether you have ever done any of the following. (Yes / No / Decline to answer)

Q69. Refused to give information to a website because you felt it was too personal or unnecessary
Q70. Decided not to use a website or not to purchase something online because you were not sure how your personal information would be used
Q71. Read a website’s privacy policy
Q72. Deleted cookies from your web browser
Q73. Activated the ”do not track” option in your web browser or installed tracking prevention tools

Do you agree or disagree with the following statements:
Q74. I feel that consumers have lost all control over how personal information is collected and used by companies.
Q75. I feel that most institutions handle consumers’ personal information in a proper and confidential way.
Q76. I feel that existing laws and organizational practices provide a reasonable level of protection for consumer privacy.

Q77. Do you have any further comments?

Leading Johnny to Water: Designing for Usability and Trust

Erinn Atwater, Cecylia Bocovich, Urs Hengartner, Ed Lank, Ian Goldberg
Cheriton School of Computer Science
University of Waterloo
{erinn.atwater, cbocovic, urs.hengartner, lank, iang}@uwaterloo.ca

ABSTRACT

Although the means and the motivation for securing private messages and emails with strong end-to-end encryption exist, we have yet to see the widespread adoption of existing implementations. Previous studies have suggested that this is due to the lack of usability and understanding of existing systems such as PGP. A recent study by Ruoti et al. suggested that transparent, standalone encryption software that shows ciphertext and allows users to manually participate in the encryption process is more trustworthy than integrated, opaque software and just as usable.

In this work, we critically examine this suggestion by revisiting their study, deliberately investigating the effect of integration and transparency on users' trust. We also implement systems that adhere to the OpenPGP standard and use end-to-end encryption without reliance on third-party key escrow servers.

We find that while approximately a third of users do in fact trust standalone encryption applications more than browser extensions that integrate into their webmail client, it is not due to being able to see and interact with ciphertext. Rather, we find that users hold a belief that desktop applications are less likely to transmit their personal messages back to the developer of the software. We also find that despite this trust difference, users still overwhelmingly prefer integrated encryption software, due to the enhanced user experience it provides. Finally, we provide a set of design principles to guide the development of future consumer-friendly end-to-end encryption tools.

1. INTRODUCTION

Despite recent revelations of mass government surveillance of the Internet [11] and more than 20 years of availability of end-to-end (E2E) encryption for email, we have yet to see the widespread adoption of encrypted email tools by the general population. Usability is undoubtedly one issue blocking adoption; since the 1999 paper "Why Johnny Can't Encrypt" by Whitten and Tygar [21], E2E encryption tools have been repeatedly criticized for their significant usability issues [4, 10, 18]. Alongside usability, Renaud et al. [16] identified a number of additional issues including awareness, concern, understanding, and knowledge of privacy that must be overcome before privacy-enhancing software is widely adopted.

While Renaud et al. argue that these problems must be solved with user education, we believe they instead show that E2E encryption must be integrated into email clients by *default* for typical users. Indeed, Gaw et al. [7] found that even activists with significant vested interest in securing their communications frequently did not realize the need for E2E encryption. Thus we believe service providers have a duty to help protect their users with the implementation of sane and secure default settings. Email service providers will not be willing to make E2E the default, however, until the usability issues have been solved in a satisfactory way so as to not drive users away from their offering. This paper aims to tackle exactly this issue.

Recent work by Ruoti et al. [17] examined the integration of E2E email encryption tools. Specifically, they proposed an E2E solution for the Gmail webmail client that used a key escrow system, requiring dependence on a trusted third party, to address the usability issues that stem from key management. They also compared an integrated solution called Pwm (pronounced "poem") against a standalone encryption program called Message Protector (MP). Their work postulated that users trust the system more when they must manually interact with ciphertext. This raises questions that tie into the work by Renaud et al. [16]: does interaction with ciphertext foster awareness, understanding, or knowledge, and thus increase trust?

We argue that the difference between Ruoti et al.'s two email implementations, Pwm and MP, exists around two orthogonal factors: integration of E2E encryption into the email client, and visibility of ciphertext. We describe the design and results of a study that evaluates user trust around these two dimensions of integration and transparency. To do this, we create a new integrated solution that makes use of the OpenPGP standard for interoperability. We use focus group testing and pilot studies to guide the design and enhance the usability of three E2E email encryption tools: a browser extension that encrypts email transparently (i.e., displays the ciphertext to the user), a browser extension that encrypts email while hiding the details of encryption, and a standalone encryption tool that requires direct interaction with ciphertext. We then evaluate these approaches with users by replicating and extending the original study by Ruoti et al. using our three different implementations of

Symposium on Usable Privacy and Security (SOUPS) 2015, July 22–24, 2015, Ottawa, Canada.

E2E email encryption.

Our work makes the following contributions:

- We show that it is possible to make an encrypted email client that is well-liked by users, while also adhering to the OpenPGP standard, by incorporating recent proposals for key management systems [13].
- We show that users have a clear preference for integrated secure email clients, when compared against manual encryption tools with similar effort put into their implementation and user experience.
- We present evidence that the transparency of encryption tools (the degree to which the details and results of encryption are shown to the user) does not have an effect on user trust. Our results show that trust hinges primarily on the software developer's online reputation. We also find that users think browser extensions are more likely to access their personal information than desktop applications, despite the reverse typically being true (as browser extensions operate within a heavily sandboxed environment).

In what follows, Section 2 will discuss work related to usable encrypted email. Section 3 will talk about the focus groups and pilot studies we used to guide our designs, and Section 4 will outline our evaluation methodology. We present the results of the user study in Section 5. Section 6 discusses design implications learned through the process, and Section 7 concludes.

2. RELATED WORK

The lack of adoption of encrypted email systems is certainly not due to the lack of implementations available [1]. There are a variety of existing tools for secure email, including browser extensions, websites, plugins, and standalone programs, available for both users of webmail and desktop mail clients. Still, the results of usability studies and the obvious failure of any implementation at widespread adoption indicate major problems with the adoption and continued use of these systems.

A great deal of recent literature on the usability of email encryption has been devoted to pinning down the precise reason as to why the widespread adoption of long-standing systems such as PGP has failed. Some authors maintain that usability is the deciding factor. In fact, since the bleak diagnosis provided by Whitten and Tygar [21] 16 years ago, subsequent PGP studies have shown little improvement to its usability. In the original study, only four out of twelve participants were able to successfully encrypt a message using PGP. Only one participant successfully completed all secure email tasks. After changes to PGP, including more automation and automatic decryption, the original usability study was repeated in 2006 [18] to show similar results: none of the participants was able to successfully encrypt a message and key exchange was still difficult to perform and not well understood.

Alternatives to PGP such as Key Continuity Management (KCM) [12] have been proposed to increase the usability of secure email by automating key generation and management. In this system, users create their own S/MIME certificates, much in the same way they generate their own PGP public keys. Garfinkel and Miller [6] conducted a usability study of KCM in 2005 to assess the advantages of existing S/MIME tools over the PGP tools available at the time. The key improvements in S/MIME addressed adoption difficulties by automating key generation upon sending email from a new address, and attaching a certificate (public key) to every outgoing email. The KCM system utilizes certificate pinning and alerts users of inconsistencies between the sender and their saved certificate. The results of this study showed that, although these improvements allowed users to correctly identify attacks in which an adversary spoofed a sender's identity, they still experienced difficulty with encrypting sensitive messages for the correct recipients. While users were technically able to send encrypted messages, this success did not extend to the wider problem of providing users with the means to make simple trust decisions (such as whether or not to trust a key initially) and correctly identify who could read their secured emails.

In 2013, Moecke and Volkamer [15] compiled a set of desirable criteria for usable E2E encryption tools. They argued that a usable secure email system must allow users to easily join, give them the tools and information necessary to make clear and informed trust decisions, not require them to understand the underlying details of public-key cryptography, and allow for secure communication without out-of-bounds verification. After an analysis of existing systems, including systems that used PGP and S/MIME, they conclude that no existing systems for secure webmail satisfy all of these criteria. Many systems that ranked slightly higher (such as Hushmail[1]) require trust in a third-party service provider.

There is another camp of authors who claim usability is not the key reason why many users still choose not to encrypt their communications. Recently, Renaud et al. [16] presented a hierarchy of mental states a user needs to progress through before she is able to adopt and use E2E encryption. The authors place usability of E2E far at the end of the list; they argue users must first be aware of privacy violations, be concerned about them, and see a need and capability to act before successfully adopting these systems for daily use. After conducting an exploratory study on 21 participants, they found most participants lacked a fundamental understanding of how email worked, could be exploited, or how to solve these issues once they are established.

Whitten and Tygar addressed the problem of education in their seminal paper, calling for *metaphor tailoring* to help users unpack the meaning behind public and private keys, encryption vs. decryption, and signing and verification. A recent study titled *Why King George III can encrypt* [20] revisits this idea by walking through basic secure email tasks, replacing technical terms with their analogous paper-mail counterparts. They motivated encryption with scenarios users can easily imagine being relevant to an 18$^{\text{th}}$-century monarch. A study on participant comprehension of these concepts found that while these metaphors sped up the explanation of secure email, they were not necessarily more effective in the end than the original technical language. After implementing these metaphors in a user interface, the authors still found that participants needed extra help to correctly perform encryption and decryption tasks.

There have been recent efforts outside of academia to address the concerns of usability and adoption by deploying and evaluating different approaches to email encryption.

[1] `https://www.hushmail.com/`

Open Tech Fund has compiled a partial list [1] of past and ongoing projects. Existing solutions include browser extensions similar to those discussed in this paper, such as Mailvelope[2] and Google End-to-End[3]. Other solutions, such as Mailpile[4] and Thunderbird's Enigmail[5] are email clients with integrated support for PGP. OpenITP recently released the results of a usability study on new Mailpile features [8]. Their study found that many of Mailpile's features enhanced the usability of key management and distribution, and also identified factors (such as awareness of encryption) that require further investigation. While the development of these tools shows great strides towards the adoption of encrypted email solutions, we have yet to see concrete evidence of the effect of individual features and design principles on usability and user trust. Our work is motivated by this large influx of tools and aims to lay some theoretical groundwork upon which more secure tools can be built in the future.

Our study is based on work done in 2013 by Ruoti et al. [17]. Their proposed system, Pwm, generates symmetric keys on a key escrow server to automate key generation and enable users to encrypt messages to contacts without obtaining public keys. As a result, the system hides almost all of the cryptographic details from the user. They found their system to be significantly more usable than select existing systems and users generally liked the effortless experience. However, they noticed that half of their users preferred their implementation of a standalone encryption application named Message Protector (MP). They inferred that this preference was due to a higher degree of trust that stems from being more involved and aware of the encryption process. However, this conclusion was derived from qualitative analysis as the experiment was not designed to probe such factors. In our work, we modify the experimental design significantly in order to explicitly probe whether this awareness is the real reason for the difference in trust, or if it can be attributed to other factors (such as differences in the design quality of the different applications).

3. DESIGN AND IMPLEMENTATION

We conducted a focus group and two pilot studies to guide the design of our encrypted email tools. We designed three separate tools, each with different levels of integration and transparency. Figure 1 shows the extent to which each of our three tools are integrated in the webmail client (also automated in the sense that the user is not required to perform additional encryption steps) and the extent to which they are transparent (i.e., show the details and results of encryption to both the email sender and email recipient).

We created two integrated encryption tools in the form of a Chrome browser extension with different levels of transparency. The transparent version automatically encrypts and decrypts email messages with the click of a button and deliberately shows the encrypted ciphertext to the user. The opaque version provides exactly the same functionality, but instead hides the ciphertext with a user-friendly overlay.

We created a transparent standalone encryption program that requires users to manually copy and paste messages from and into a separate software program. This tool shows the resulting ciphertext to the user as they step through the encryption process. We chose not to implement an opaque, standalone encryption tool as this would be more difficult to implement, likely not interoperable with other PGP software, and would not aid in investigating the hypotheses we describe in Section 4.

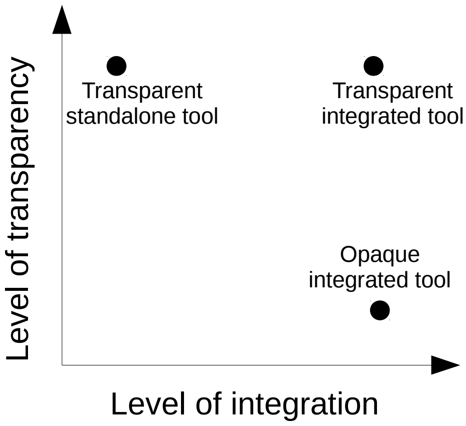


Figure 1: Placement of our tools on the transparency-integration scale

In the following sections, we discuss: how the focus group was conducted and how the results influenced our designs; how the initial pilot study was conducted and briefly, its results; how the final pilot study was conducted; and lastly, how the information and lessons learned from these studies influenced the final design of our software applications.

3.1 Focus Group

Before we began to design our tools, we conducted a focus group on existing encrypted email tools to get a sense of which features were the most desirable and which tasks posed the greatest problem to users. We gathered three participants (F0, F1, and F2) and had them perform a variety of tasks involving key management, encrypting emails before sending them, and decrypting emails upon receiving them. We had them perform the tasks using 1) Pwm, the integrated tool developed by Ruoti et al. [17], 2) Mailvelope, an encrypted email browser extension, and 3) the Enigmail plugin for Thunderbird. We did not include Message Protector, one of the programs developed by Ruoti et al., as it was not available to us at the time. The task order was varied using a 3x3 Latin square. All participants performed the tasks in the same environment using an email account we provided them solely for the focus group. All data analysis from this focus group was qualitative.

F0 and F1 were both computer science graduate students with experience using PGP software at some point in the past. They uncovered a number of issues with all three pieces of software, in both usability and generic software bugs. Their experience proved helpful in spotting a number of potential issues for ordinary users: for example, F1 noted that "it is as easy to send someone your private key as your public key [in Mailvelope]". F2 was a graduate student in Pure Mathematics with little-to-no prior exposure to encrypted email or cryptography.

We will now summarize our findings by software program.

[2] https://www.mailvelope.com/

[3] https://github.com/google/end-to-end

[4] https://www.mailpile.is/

[5] https://www.enigmail.net/

Mailvelope: All three participants had trouble finding the functions they needed (e.g., for generating private keys and importing public keys) in the Options window of Mailvelope. Each participant imported public keys in a slightly different way, including F0, who copy/pasted its contents despite the existence of a direct import-from-email feature in Mailvelope. All three expressed confusion as to whether the key had been imported successfully or not. Participants also expressed an annoyance with the lack of labels on the webmail overlay (F1: "What are these buttons?"). All three participants had difficulty understanding the dialog for selecting which users to encrypt the email for (which were not determined automatically from the list of email recipients). Participants also frequently chose to interact with ciphertext that was not intended to be human-editable: F1 managed to successfully import an incorrect public key by modifying the public key block.

Pwm: Participants were generally impressed with Pwm's simple, integrated interface (F2: "Oh really? No way!"). However, F1 missed the "encrypt" button in the overlay and unknowingly sent an unencrypted email. F2 *almost* sent an unencrypted email, but caught themselves at the last moment and decided to try clicking the lock button. F0 and F2 noted that the icon (an unlocked lock) is ambiguous as to whether a user should click to unlock (thus the message is locked by default) or if they had to click to lock the message. This finding is similar to the issues participants had with Mailvelope's use of unlabeled icons for buttons.

Enigmail: The Enigmail extension for Thunderbird is intended to open the setup wizard on first use, but there was a bug which prevented this from occurring on the current version of Ubuntu we set up for the focus group. We thus had to instruct all three participants to manually start the wizard from a buried menu option. This wizard contains a significant amount of exposition and technical detail, and all three participants expressed exhaustion reading it and eventually gave up in favour of spamming the "next" button. F1 and F2 found the "generating randomness" explanation confusing and thought they had failed to follow the correct instructions. All three found the key management interface confusing, with F2 opening public keys in a text editor frequently instead of importing them. All participants also found the Sign/Encrypt checkboxes difficult to locate, and clicked the menu that reveals them numerous times to convince themselves that they were still checked. F2 forgot the passphrase they set to locally encrypt their private key during the setup phase.

In summary, participants encountered usability issues with all three software applications, but had a far easier time with Pwm. We believe this is due to the design decision of Pwm to prevent the user from interacting with any options panes whatsoever in order to accomplish the software's basic tasks. While the extra tutorial and options provided by Enigmail appear helpful, they in fact turned out to be detailed and verbose to the point that users found them overwhelming and started ignoring them. The results of this focus group provided us with three main design goals to guide the design of our encrypted email tools.

1. Easy setup: A wizard should prompt users for the minimum information needed to start using the system and should not overwhelm them with exposition.

2. Simple to use: Users should not need any interaction with the options screen whatsoever in order to accomplish basic tasks (i.e., encryption, decryption, and importing keys).

3. Clear and explicit: It should be obvious to users what the result of clicking a button will be, and whether or not their email will be sent securely *before* they click the send button.

3.2 Initial Pilot Study

With these goals in mind, we proceeded to design and implement our own integrated and standalone encryption tools. We will present the final versions in Section 3.4. We evaluated them by running an initial study with six participants. We recruited participants with no prior experience with email encryption tools from the Faculty of Mathematics graduate program to participate in our pilot study. Before running the study, we received clearance from our Office of Research Ethics. We used the results of this study to increase the usability of our tools and eliminate confounding factors that would affect their perceived trustworthiness.

We noticed a trend, as in the focus group, of users editing or interacting with text in ways they were not supposed to. Three out of six participants failed to copy the `--- BEGIN MESSAGE ...` header we placed at the top of messages for our standalone system. The purpose of this header was to indicate the beginning of a ciphertext message. In the integrated system, some participants added a plaintext copy of the message into the ciphertext block, below the header. This is a significant problem in the design of both systems, as it poses a threat to the confidentiality and integrity of messages. We alleviated this problem in the standalone system by removing the headers altogether and creating separate buttons for the encryption and decryption tasks.

We observed that four out of six participants encrypted to the wrong recipients using our standalone tool. (This problem did not exist in our integrated tool, which automatically detected email recipients from the Gmail compose window.) Some chose to encrypt to everyone in their address book, or to no one. We partially countered this problem by requiring them to select exactly one recipient to encrypt for, but solutions to this problem are limited by the nature of a standalone approach to encryption tools: there is no way for an unintegrated solution to compare who the message is being *sent to* with who it is being *encrypted to*.

Finally, we observed several isolated instances of confusion over various advanced aspects of the software. One participant got lost in the "Advanced Options" menu, and one had trouble with the manual key management process of the integrated tool. We addressed these issues by reorganizing and simplifying much of the UI.

3.3 Final Pilot Study

Before commencing the final version of our study, we recruited five participants as pilots (using the recruitment process described in Section 4.1). We used these pilot participants to find any remaining bugs in our implementation, problems with the wording of our instructions, and any final issues with our UI choices. We fixed issues between pilot participants as time permitted, and then fixed all outstanding issues once the five pilot runs were completed; no changes were made in the duration of the user study.

The recurring issue of users interacting with structured ciphertext appeared again, this time in the integrated tool.

To fix this, we locked the contents of the message body once it had been encrypted; users had to decrypt it first in order to edit it. We also found users entering the wrong text in some fields of the standalone system, so we added strict validation and feedback to as many input fields as we could.

3.4 Implementation Details

In this section, we describe the final versions of our three encrypted email tools. Although some example screenshots are given here to highlight key concepts, we have included a more thorough set of screenshots documenting our tools in Appendix C. Note that we used the branding of "Mailvelope" for our integrated tool and "Message Protector" for the standalone tool in order to give the impression of a completed product to the participants. We attempted to make the tools as similar as possible, limiting the differences to those required to achieve different levels of transparency and integration. We used the same UI widgets, consistent terminology, and similar messages for each tool.

3.4.1 Key Management

One of the most challenging aspects of designing an encrypted email client for everyday computer users is the concept of key management. In order to use PGP encryption, users need to first (securely) obtain a key from each intended recipient. Current PGP implementations [19] use a Web of Trust model in which users signed each others' keys to establish paths of trust, but this idea has failed to catch on with non-technical users [14]. Later specifications (such as S/MIME) proposed using trusted authorities to verify keys' authenticity, but this defeats the decentralization ideal offered by end-to-end encryption. Newer proposals such as verified keyservers (e.g., Keybase.io[6]) and Google End-to-End [9] (similar to certificate transparency [13]) suggest a middle ground, in which lesser security than direct out-of-band verification is obtained but less trust is placed in central authorities (such as key escrow servers). We believe these proposals represent a good *default* level of security for the everyday user: all users are protected from passive adversaries, but more complex options (such as out-of-band verification or trust paths) can still be applied within the same system, allowing advanced users to gain some protection against active attackers.

To this end, we implement key management using a simulated version of Keybase.io. The core principle of Keybase.io is that (unlike PGP's HTTP Key Protocol servers to which anyone can post a public key) the person uploading a public key for an email address must be able to *receive* email at that address, preventing dishonest users from uploading keys for arbitrary other users. This means the keyserver can be polled for a recipient key with high probability of the actual recipient being the owner of said key, as long as there are no active adversaries present in the system. With this capability, encryption software is able to transparently perform key management on behalf of the user for any correspondents that have similar software installed. The user interface in the failure case of this scenario simply needs to explain to the user that the recipient does not have compatible software installed. We believe this concept is easily grasped by most everyday users, as they experience it already with segregated IM and social networking applications.

3.4.2 Integrated Tool

We chose to implement our encrypted email overlay by modifying Mailvelope, the browser extension we used in our focus group. We chose Mailvelope because it is open source, implements the OpenPGP standard, works with the current versions of Chrome, Firefox and Gmail, and also has a feature to support arbitrary webmail providers. Due to time constraints we chose to limit our development efforts to supporting Gmail accessed via Chrome on Ubuntu, although there would be minimal development effort required to expand our changes to include the full set of platforms and browsers. A Chrome extension package (CRX) for our modified browser extension is available [2].

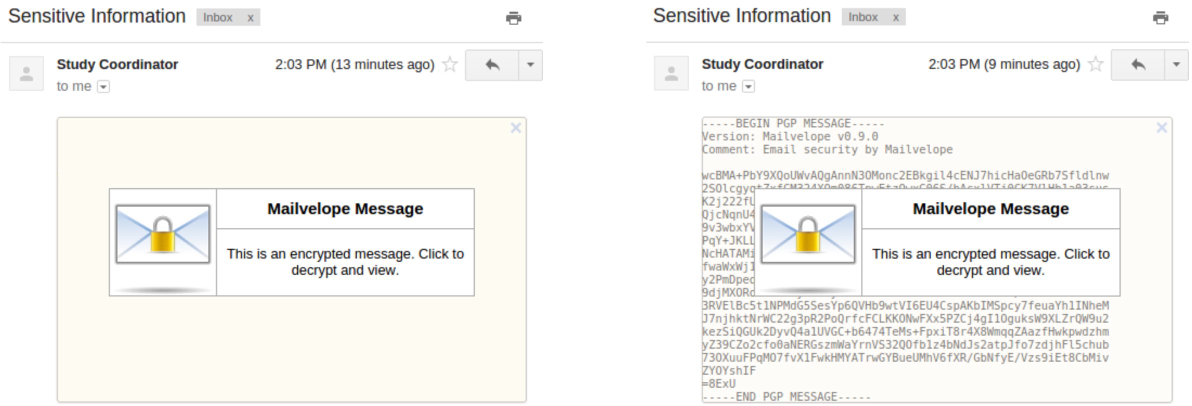
In support of our *easy setup* goal above, we implemented a wizard that opens the moment the extension is installed. The wizard asks the user for only their name and email address, which is then used to generate a PGP keypair for them using sane default settings. Similar to Ruoti et al., we removed the requirement to set a local password protecting the private key and set it to be empty by default.

We successfully met our second goal of preventing the user from interacting with the options pane by adding mechanisms to the Gmail overlay that accomplished all required tasks. This included automatically setting sane default options. We implemented the key management mechanism described previously by emulating tie-in with Keybase.io. When the user checks "Encrypt this email", our emulated Keybase.io server is polled for keys corresponding to the recipients in the To: field. If a key for all recipients is found, the message is simply encrypted on the spot. If there are recipients for which no key is found, a prompt appears informing the user why the message cannot be encrypted ("the recipient does not have [software] installed"), and gives them the option to automatically send an invitation to the recipient to install it themselves. When the recipient accepts the invitation, a message is sent back to the user informing them they can now encrypt messages to that recipient.

Our third goal of being *clear and explicit* was met by replacing all of the UI widgets in the webmail overlay with labeled components, and providing explicit feedback to the user in strategic places. For example, we replaced the text of the "Send" button with "Send Unencrypted" with a checkbox labeled "Encrypt this email" placed right next to it. Once checked, the extension checks to see if there are public keys available for the typed recipients in the Compose window. If there are, the email is encrypted with PGP and the Send button is reverted to its normal text. We also replaced the iconography used to represent special messages (encrypted emails, key requests, etc.) with a box stating that it was an encrypted email object and a brief description of what would happen if the user clicks on it.

To facilitate investigation of our research questions in Section 4, we created two versions of our integrated tool ("transparent" and "opaque") with slight differences in the UI. In the opaque version, no ciphertext is ever displayed to the user. In the transparent version, the overlay on received encrypted messages is partially transparent, so the user can see the ciphertext behind it. Figure 2 shows these differences side by side. We also hide the ciphertext in the Compose Message window in the opaque version, replacing it with a message stating "This message is now encrypted".

[6] `https://keybase.io/`



Mailvelope Message

This is an encrypted message. Click to decrypt and view.

(a) Receiving an encrypted message in the opaque version

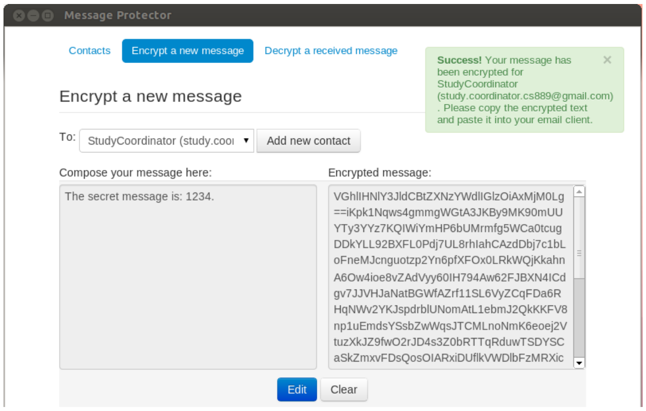
```
-----BEGIN PGP MESSAGE-----
Version: Mailvelope v0.9.0
Comment: Email security by Mailvelope

wcBMA+PbY9XQoUWvAQgAnnN3OMonc2EBkgil4cENJ7hicHaOeGRb7Sfldlnw
2SOlcgyq
K2j222fU
QjcNqnU4
9v3wbxYV
PqY+JKLL
NcHATAMi
fwaWxWjl
y2PmDpec
9djMXORd
3RVElBc5t1NPMdG5SesYp6QVHb9wtVI6EU4CspAKbIMSpcy7feuaYh1INheM
J7njhktNrWC22g3pR2PoQrfcFCLKKONwFXx5PZCj4gI1OguksW9XLZrQW9u2
kezSiQGUk2DyvQ4a1UVGC+b6474TeMs+FpxiT8r4X8WmqqZAazfHwkpwdzhm
yZ39CZo2cfo0aNERGszmWaYrnVS32Q0fb1z4bNdJs2atpJfo7zdjhFl5chub
73OXuuFPqMO7fvX1FwkHMYATrwGYBueUMhV6fXR/GbNfyE/Vzs9iEt8CbMiv
ZYOYshIF
=8ExU
-----END PGP MESSAGE-----
```

Mailvelope Message

This is an encrypted message. Click to decrypt and view.

(b) Receiving an encrypted message in the transparent version

Figure 2: Comparison of opaque and transparent versions of the integrated tool



Success! Your message has been encrypted for StudyCoordinator (study.coordinator.cs889@gmail.com) . Please copy the encrypted text and paste it into your email client.

Encrypt a new message

To: StudyCoordinator (study.cooı

Compose your message here:

The secret message is: 1234.

Encrypted message:

```
VGhlIHNlY3JldCBtZXNzYWdlIGlzOiAxMjM0Lg
==iKpk1Nqws4gmmgWGtA3JKBy9MK90mUU
YTy3YYz7KQIWiYmHP6bUMrmfg5WCa0tcug
DDkYLL92BXFL0Pdj7UL8rhIahCAzdDbj7c1bL
oFneMJcnguotzp2Yn6pfXFOx0LRkWQjKkahn
A6Ow4ioe8vZAdVyy60IH794Aw62FJBXN4ICd
gv7JJVHJaNatBGWfAZrf11SL6VyZCqFDa6R
HqNWv2YKJspdrblUNomAtL1ebmJ2QkKKFV8
np1uEmdsYSsbZwWqsJTCMLnoNmK6eoej2V
tuzXkJZ9fwO2rJD4s3Z0bRTTqRduwTSDYSC
aSkZmxvFDsQosOIARxiDUflkVWDlbFzMRXic
```

Figure 3: Encrypting a message using the standalone tool

3.4.3 *Standalone Tool*

We wrote our standalone program to resemble the Message Protector (MP) program from the original study as closely as possible. As Ruoti et al. made a point of MP being a standalone software application and not a web app, we created a desktop app using the Chromium Embedded Framework, which allowed us to implement it as a web app using Javascript and HTML5 and then install it locally to a user's desktop. Unfortunately, we had only a single screenshot and a description of MP's functionality to go from, and so were forced to infer some details of its UI. We also made a number of changes motivated by our focus and pilot studies, in order to make a best-effort approach to implementing MP's user experience. The code for the final application is available on our website [2].

Figure 3 shows Alice in the process of encrypting a message with sensitive information to Bob. After encryption, users must copy and paste the ciphertext from the output field into a webmail client of their choice. When receiving an encrypted message, they copy and paste the received ciphertext into the Decrypt pane of the standalone software. We implemented key management to more closely resemble the functionality of our integrated tool. When the user attempts to add a new entry to their contact list, the Keybase.io server is polled. If a key is found, the entry is simply added to the contact list and encryption to that contact can be done immediately. If no key exists, a message is displayed informing the user and asking them to copy/paste an invitation to the recipient, instructing them to install the tool. The contact entry is added to a pending contacts list, and the Keybase server is then polled periodically in order to see if the recipient has joined the system yet. Once they have, a notification is displayed in the software, and encrypted messages can then be sent to the contact.

4. METHODOLOGY

The goals of our experiment were two-fold; we wanted to demonstrate the usability of our integrated email encryption tool, and to investigate which software features are most related to user trust. We designed our experiment with the following hypotheses in mind:

H01: The integrated client is a usable encrypted email tool, as determined by the System Usability Scale (SUS).

H02: Standalone encryption tools provide a less desirable user experience than integrated encryption tools.

H03: The extent to which encryption software is automatic does not have a significant effect on user trust.

H04: Users trust transparent encryption software more than opaque software.

We conducted a mixed measures user study to evaluate these hypotheses. We used a between-subjects variable to evaluate the effect of transparency on user trust, and a within-subjects variable to evaluate the preference of an integrated, automatic encryption tool over a standalone, manual encryption tool. We will accept or reject the last three hypotheses based on participant responses to probing questions on tool preference and trust.

4.1 Subjects

We asked 36 participants (15 male, 21 female) to perform a sequence of email encryption tasks on two different systems (integrated and standalone). 18 participants were given the transparent version of the integrated client, wherein the ciphertext was visible after encryption and before decryption. The other half were given the opaque version of the client, with hidden ciphertext. We also alternated between making participants perform the integrated tasks first and second. As with our focus and pilot studies, this study received clearance from our Office of Research Ethics.

We required all participants to be active webmail users (Gmail or Hotmail) to reduce the effect of differing experiences with email clients on the results. To help avoid bias from technical expertise, we excluded all current or former computer science students, and those with any background or knowledge of cryptography. We advertised our study on Craigslist and Kijiji, two popular local online classifieds websites. We also advertised our study to an introductory computing course for non-majors, a university-wide graduate student mailing list, and via posters put up around campus. We ended up with 33 students and 3 non-students. About 17% of participants were engineering students, 42% were science students (chemistry, biology, environment, health, and physics), 25% were studying mathematics (statistics, pure mathematics, and actuarial sciences), two were business students, and one was studying political science. Our non-student participants were a mechanic, a civil engineer, and a cashier. Participants knew from the recruitment materials and information letter that they were participating in a study on the usability of email privacy tools. They had no knowledge of the tools beforehand, and were not told that we were measuring trust.

The self-rated level of computer expertise in our participant pool ranged from minimal to expert. Of the 36 participants, 67% reported having previously sent sensitive information over webmail or Facebook. No participants had ever used email encryption, though a few reported taking precautionary measures when sending sensitive information. Some of these measures were effective: P06 and P09 reported previously sending sensitive information in password protected zip files, and P22 used the university's secure file transfer service. Others practiced mildly effective techniques: P02 made an attempt at concealing information by writing it in Chinese, and P03 sent protected information over mobile voice calls to increase the difficulty of interception. Others still exhibited a misunderstanding of features. P27 reported sending sensitive information over Gmail with Chrome's incognito mode, and P01 sent sensitive information over emails addressed with BCC. The majority of participants who sent sensitive information reported taking no precautionary measures (67%). Despite this trend, all but one participant considered maintaining the privacy of messages containing sensitive information to be "important" or "very important".

4.2 Tasks

We asked each participant to perform a set of tasks with both the integrated and standalone clients. These tasks included setting up the system, sending secure messages, and receiving secure messages. After each set of tasks, the participants then filled out a questionnaire. They answered questions about usability, suggestions for improvement, and trust. We used an interactive online survey for participant instructions and questionnaires.

We conducted the study in an empty room, with one computer for the participant and one for the study coordinator. They had access to a Jane Doe Gmail account, created for the purpose of the study. To protect the privacy of our participants, we asked them not to use their own personal email account. The participants worked on a virtual machine running Ubuntu 12.04LTS. We carefully explained the location and purpose of each icon at the beginning of the study. The launcher only contained icons for the programs they were required to use (Chromium web browser version 34, and the standalone client). We freely answered questions about the UI that were unrelated to the encryption software. We ran a desktop recording program in the background to review their actions and movements after they had completed the study. We also collected audio recordings during their completion of the study tasks and the post-study interview.

Participants had no knowledge or introduction to the tools before being asked to complete the tasks. We instructed them to complete the tasks to the best of their ability, and asked them to explore the interface if they were unsure about what to do next. We intervened or notified them of the correct actions only when they became stuck for an extended period of time. The purpose was to get a clear idea of how inexperienced users would interact with the systems, and what natural tendencies led to encryption mistakes.

In what follows, we describe the tasks performed using the integrated and standalone clients, as well as the questionnaire and interview process given to the participants.

4.2.1 Integrated Client Tasks

Setup: At the beginning of the integrated portion of the study, participants were asked to log into Jane Doe's email account to see a single message with a request to install a secure webmail extension. The text of this email contained a link from which participants could install and set up the Chrome extension. Normally users could be directed to the Chrome Extension Store to obtain the extension, but we used a separate download website for our study; some of the consequences of this are discussed in Section 5.

Email Decryption: After setup, participants received an encrypted email message from the study coordinator. To decrypt, they followed the instructions on the overlay and copy and pasted the decrypted text into the interactive survey.

Email Encryption (Part 1): Participants were asked to perform two different encryption tasks: in the first task, we asked them to reply to the message they received from the study coordinator in the previous task.

Email Encryption (Part 2): In the second encryption task, we asked participants to send a secure message to a new contact, for whom they did not yet possess an encryption key. The participant first had to email the contact with plaintext instructions on how to install and setup the integrated browser extension. After they received confirmation that their contact had signed up (in the form of an encrypted email), they were able to send an encrypted message.

4.2.2 Standalone Client Tasks

Setup: Our standalone tool was set up as a pre-installed standalone desktop application. We assume users are familiar enough with installing software to leave this part out of the usability study (and we did not want to influence them with the unfamiliar process of installing software in Ubuntu). When opening the tool for the first time, participants were asked to enter their email information and to add the contacts randomFriend@hotmail.com, mom@familyWebsite.com, and study.coordinator.cs889@gmail.com.

Email Encryption (Part 1): We then asked participants to use the standalone client to send a secure message to study.coordinator.cs889@gmail.com.

Email Decryption: Next, they received a reply from the study coordinator with an encrypted message. They were then asked to decrypt this message using the standalone client and paste the contents back into the interactive survey.

Email Encryption (Part 2): As in the integrated client tasks, we concluded by asking participants to use the standalone client to send a secure message to a new contact for whom they did not possess a public key. Participants were first instructed to send plaintext instructions to this contact. Again, the client simulated polling Keybase.io. When it found an added contact's key, it displayed a notification that the participant could now communicate with the contact securely.

4.3 Questionnaires

After completing the tasks for each system, participants answered a few survey questions about their experience with and trust level of the system (full text given in Appendix A). They first provided feedback on the usability of the system in the form of ten Likert scale statements. We used the results of this feedback to calculate the System Usability Scale rating of each system [5].

The remaining questions asked users what they liked and disliked about each system, how often they envisioned using the system, and quizzed their understanding of who could read the messages sent between them and the study coordinator. These were used to collect qualitative feedback on how usable and trustworthy users found the systems.

We concluded the questionnaire by asking participants to choose which system they preferred and their thought process behind that decision. The purpose of these questions were to gather more qualitative feedback on the usability of our systems, and to find out which aspects of the system were most important to users in terms of usability or trust.

4.4 Interview

To gain a deeper understanding of the answers to the last part of the questionnaire, we finished the study by asking each participant a series of increasingly probing questions about which system they preferred. We began by asking them to restate their answers at the end of the questionnaire, and then asked them explicitly about trust if they did not factor that into the reasoning behind their decision. Finally, we asked them to imagine a scenario in which they managed a business that required employees to send sensitive information to clients. We asked them which tool they would prefer their employees to use. The purpose of this question was to gain insight into what problems they considered important and how they assumed the general public would interact with encryption tools.

5. RESULTS

5.1 System Usability Scale

The System Usability Scale (SUS) was originally proposed in 1996 by John Brook as a means of evaluating the usability of products in an industrial setting [5]. SUS is based on a Likert scale measurement in which the evaluator indicates their agreement on a 5 point scale with 5 positive and 5 negative statements about the product. We use this scale to evaluate the usability of our tools in order to compare our work with the results of the similar experiments conducted by Ruoti et al. [17].

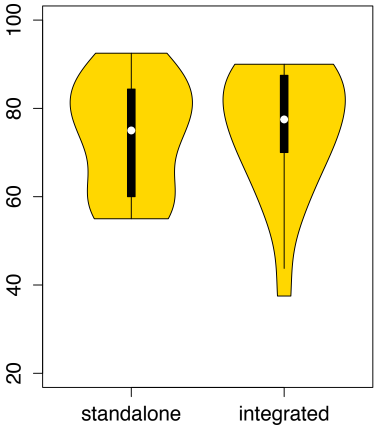


Figure 4: Absolute SUS scores for our encryption tools

We first calculated the SUS scores for the integrated and standalone clients, using only the half of participants who used each tool first. This was to obtain an absolute usability rating of each tool and avoid bias from participants comparing them to the tool they used before. The results were a SUS score of 75 ± 14 for the integrated client and 74 ± 13 for the standalone. These tools both receive an adjective rating of "good" according to Bangor's adjective ratings [3]. A summary of the SUS scores for each group of 18 participants are shown as violin plots in Figure 4. These enhanced box plots show the distribution of the calculated SUS scores from each participant. We found no significant difference ($p > 0.5$) using the Mann-Whitney U test for non-parametric data in the absolute usability of these two tools. The similarity in user experience ratings between these two tools gives a strong indication that the differences in qualitative feedback are due to our experimental factors (i.e. transparency and integration), and not confounding design factors. Our SUS scores for both tools are comparable to those reported by Ruoti et al. (76 for the integrated tool and 74 for the standalone).

We then calculated the SUS scores for each tool from participants that had performed tasks with the other tool previously. The usability rating for the integrated client, among participants who had already seen the standalone client was 78 ± 16, and the usability rating for the standalone client among participants who had previously seen the integrated client was 52 ± 21. This latter result showed a statistically significant difference ($p < 0.01$) in usability ratings for the standalone client between participants who were seeing it for the first time and those who had already tried out the integrated client. Figure 5 shows the interaction of ordering with the SUS scores for the standalone and integrated tools. Each plot represents 18 people; the absolute plot contains SUS scores from the participants that used the tool first, and the comparative plot shows the SUS scores from the participants that used it second.

When comparing the integrated and standalone usability scores of all participants, we found that the usability score of our integrated tool (73±15) was significantly higher than the standalone counterpart (63±20, $p < 0.01$). We did not see a statistically significant difference between the SUS scores of the transparent and opaque versions of the integrated client ($p > 0.05$). The respective scores were 73 ± 15 and 80 ± 14.

	Behaviour ID	Occurrences	Description
Integrated tasks:	i-1-3	5	Asked if they were supposed to install extension after seeing warning
	i-3-5	7	Confused as to how to send a secure message
	i-3-4	3	Tried to use the standalone client to send a secure email
	i-4-6	14	Sent encrypted email before knowing the new contact had joined
	i-x-2	3	Sent a plaintext message
Standalone tasks:	s-1-1	6	Forgot to add contacts
	s-4-1	18	Sent encrypted email before knowing the new contact had joined
	s-4-4	4	Confused by invitation process
	s-4-7	3	Added new contact twice

Table 1: Common behaviours encountered in the performance of study tasks

5.2 Observational Results

We noticed several common behaviours our participants exhibited while performing the study tasks. The most common of these are summarized in Table 1 and discussed in more detail below. See Appendix B for a full table of issues and behaviours. Our numbering scheme identifies issues by tool and task number in the following manner: {i/s}-{task #}-{id} where the id is arbitrary.

i-1-3: The integrated setup task asked participants to install a browser extension. We did not have our extension on the Chrome Application store, and this triggered several browser warnings (see Figure 6). Five of our 36 participants showed hesitation in completing the installation process and asked us whether it was okay for them to proceed. It is possible that these warnings affected their trust of the system.

i-3-5: The third study task required participants to send an encrypted email to the study coordinator for the first time. As we withheld tutorial information about how to do so (for reasons explained in the previous section), we noticed some confusion as to how the encryption process worked. Several participants asked us how to send a message with the installed software instead of Gmail. We responded that they should experiment with the user interface and proceed with sending an email as they normally would. After this prompting, most participants noticed the existence of the encryption checkbox upon opening a new compose window.

i-3-4: Other participants exhibited a behaviour similar to i-3-5, in which they were confused about how to send a secure message with the integrated tool, but instead of asking for direction, they opened the standalone application visible in the application launcher of their desktop. We stepped in at this point and told them to proceed with sending an email through Gmail as they normally would.

i-4-6: The last study task for the integrated client asked participants to send an encrypted email to a contact for whom they did not possess a public key. The purpose of this was to observe how participants interact with, understand, and trust our simplified key management scheme. After participants sent an invitation email to the new contact, the study coordinator accepted the invitation right away and uploaded a public key to our Keybase.io server. The majority of participants waited for a confirmation email from the contact, explaining that they had installed the system and could now communicate securely. However, 39% of the participants sent the encrypted email without waiting for the confirmation. This was possible because key management was done by automatically connecting to a key server in the background. This behaviour may suggest that participants are not aware of the steps that are normally involved in key management. It also suggests that they are willing to take advantage of transparent key management schemes.

i-x-2: Only three of the 36 participants (8%) sent "sensitive information" in plaintext during the encryption tasks. Two of these participants had not noticed the encryption checkbox or "Send Unencrypted" button and expressed surprise when we pointed out their existence at the bottom of the compose window. The other participant did manage to encrypt the messages, but appended a plaintext copy before sending it to the study coordinator.

s-1-1: Six participants skipped the majority of the first study task, not bothering to add contacts to the system. They realized this fact upon reaching the encryption phase of the study, since they were unable to encrypt a message without first selecting a contact to send the message to.

s-4-4: As in the integrated client tasks, we asked participants to send an encrypted email to someone for whom they did not possess a public key. Some participants asked us how to "send something through" the standalone client. We explained that they had to add the contact and then copy and paste the instructions into their webmail client.

s-4-1: After participants invited the new contact to install the standalone application, many then proceeded to encrypt a message to the contact without seeing or noticing confirmation that the contact had installed the system. Our behind-the-scenes key management process made it possible to use the software without an understanding or awareness of the exchange of public keys. This willingness to proceed without confirmation or feedback is the same behaviour we witnessed during the integrated tasks (behaviour i-4-6).

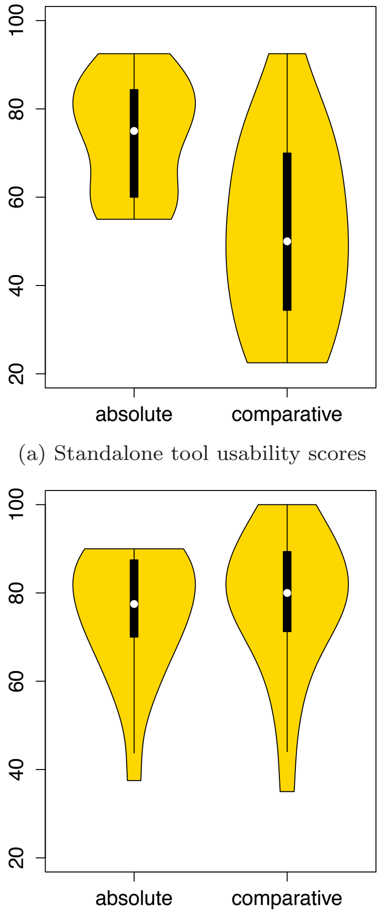
s-4-7: After receiving confirmation that the new contact had installed the standalone application and could now receive encrypted emails, three participants tried to add the contact to the system again. This could be a further indication of confusion surrounding key distribution, or it could be a failing of the user interface.

5.3 Qualitative Results

Five general themes surrounding usability and trust emerged from the questionnaire and post-study interview. We now discuss these themes in detail.

5.3.1 Preference for Integration

81% of participants preferred the integrated client over the standalone. The majority of them cited the integration into Gmail as the reason for their preference:



(a) Standalone tool usability scores

(b) Integrated tool usability scores

Figure 5: Interaction of ordering and SUS scores. "Absolute" refers to participants who used the respective tool first, whereas "comparative" refers to participants who had already used the other tool beforehand.

> *I find it more convenient... I don't have to open up another program to send the encrypted message, I can just choose whether or not to encrypt it when sending an email... (P07)*

> *[The integrated tool] is much more convenient to use, to the point where I wouldn't mind encrypting even everyday, non-sensitive emails (P24)*

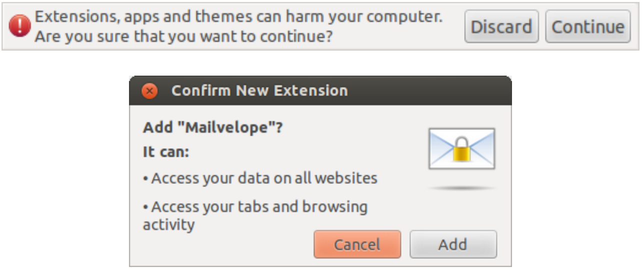
> *I liked being able to use something that complemented the email system I currently use instead of having to learn something new then apply that to what I already use. (P30)*

Others pointed out specific aspects of the standalone tool that were cumbersome or tedious:

> *I would have to go into a whole other program, open it up, encrypt it and if I'm . . . sending 50-60 messages a day it'd be difficult to do so and always have to go back and forth. (P22)*

> *With [the standalone client] I had to go through this whole process... when you send many emails eventually that gets tedious... it's definitely something I would want to avoid. (P34)*

This sentiment was especially apparent in the interview when we asked participants which tool they would prefer if they were in a managerial position.



Extensions, apps and themes can harm your computer.
Are you sure that you want to continue?

Confirm New Extension

Add "Mailvelope"?
It can:
- Access your data on all websites
- Access your tabs and browsing activity

Figure 6: Two warnings that appear when installing extensions in the Chrome browser. The first appears only when installing from sources other than the Chrome Web Store.

> *I would rather have them use [the integrated tool] because if I give them a more complex system like [the standalone tool], there might have to be some kind of training going into it. If it's much more complicated, I might not even use it because it's so, well, tedious and I feel my employees might not even be using it. There's always that chance so I feel that [the integrated tool] is a much better option. (P01)*

Some participants were also aware that the introduction of extra steps also introduced opportunities for human error, with P16 saying they could "see people accidentally not copying the entire message by missing a character and the message may become incomplete".

> *[The standalone tool] requires an additional program to be open, and copy and pasting to occur which I could mess up on. (P22)*

Many participants also added that this tool would be easy to set up to encrypt messages by default. They expressed during the interview that in a business environment, they would prefer to have encryption always on. One participant expressed concerns that employees could forget to encrypt emails before sending even with the integrated client.

> *... it would be easy to train [employees] to check the box before they send it, but I suppose that would also make it easy for them to forget... you asked about the possibility of having it always on, so that would make it a good thing I think to use it that way. (P09)*

5.3.2 Lack of Trust Preference

The overwhelming majority of participants (69%) did not trust one system over the other, even when prompted to think about trust. Most participants addressed the question of whether they trusted one system more with a simple "no", but others explicitly stated that they felt the two tools "did the same job" (P14).

Some participants with no trust preference did not consider email an appropriate method of secure communication.

> *Currently, I do not feel there is a need to encrypt my sensitive information. I typically do not share sensitive information via email... email encryption is too troublesome, especially when no one around me has used it before. (P21)*

> *...for some extremely important information, I would like to talk face to face or through cellphone. (P04)*

5.3.3 Distrust of Integration

Of the 31% of participants who did have a trust preference, most (10 out of 11) trusted the integrated client less than the standalone one because of its integration with the browser:

> *[The standalone tool] acts as a different entity that is on your desktop and not integrated online. So it feels more secure to encrypt and decrypt messages separately (P08).*

Another participant pointed out that the seamless integration "...might give the sense that [this tool] is 'less safe' and more intrusive" (P26).

While the idea of using physical security measures to protect sensitive information (e.g., keeping a password in a locked drawer as opposed to a text file in a home directory) is correct, the notion that our standalone client and programs like it are offline and prevented from leaking information to third parties is incorrect. In fact, browser extensions can be considered "more secure" than separate applications due to the sandboxing they are subjected to. Only one participant mentioned that the integrated client "...seemed safer since it is an extension on the browser compared to [the standalone program] which is a software on my computer" (P32).

5.3.4 Reputation-Based Trust

Of the 69% of participants who had no trust preference, many (9 out of 25) mentioned that they would prefer to do research on the companies producing the software, or to hear about the tools' word-of-mouth reputation. One participant, without prompting, added that "actually I don't know whether I should trust these two systems" (P17) while stating their preference of the integrated client over the standalone. Another stated that their trust depended on "basically how well I know the company and if I don't know I'll research it and see if it's professional-looking." (P13)

5.3.5 Misunderstanding of Key Management

Several participants demonstrated a misunderstanding of who could read messages encrypted with the integrated tool. One participant thought that the standalone client was safer because it encrypted only for the recipient they selected, as opposed to the integrated client which had the same "encryption" (key) for every contact.

> *[The standalone client] had a different encryption for each contact, like it allowed different encryption routes for my privacy. (P27)*

> *...with [the standalone tool], only certain people can receive the email and decrypt it. You have to add the contact. (P02)*

Another participant did not like going through the invitation process and wanted to be able to send encrypted messages immediately. They wished for the invitation to include the ciphertext for the first message, so the recipient could read it as soon as they installed either piece of software.

5.4 Discussion

The quantitative data we collected for the evaluation of the usability of our encrypted email tools on the System Usability Scale, together with the qualitative feedback we received in the questionnaire and interview portion of our study, provide strong evidence in support of our hypothesis that our integrated client is a usable encrypted email tool (H01). It had a SUS score of 75, receiving an adjective rating of "good", and was comparable to the encryption tools proposed by Ruoti et al. (which received SUS scores of 76 and 74). The qualitative feedback on the integrated tool, summarized in the previous section, was positive. Most users stressed its positive aspects and integration when describing their preference for our system over the standalone system.

Our second hypothesis (H02), that standalone, manual encryption tools provide a less desirable user experience than integrated, automatic tools is also supported by the results from our quantitative and qualitative evaluation of the usability of both systems. The preference for integration was the strongest theme we encountered in participant feedback, with 81% of participants citing this as the primary reason they preferred the integrated client over the standalone one.

In our third hypothesis (H03), we expected to see no preferences in user trust between systems that were integrated (automatic) and those that were standalone applications (manual). Although the majority of our participants did not trust standalone software more than the integrated browser extension, 28% expressed a belief that the standalone system was more secure. Most reasoned that this was because it was not integrated into the browser. This is an interesting trend that perhaps mimics other systems in which diversification prevents losses due to a single point of failure (e.g. financial investments, nutrition, etc.). It is also possible that these few participants expressed this belief because of the browser warnings that were displayed when installing the integrated system. Five participants stopped the study to ask whether they were allowed to install the extension after seeing such a warning. Several cited this as the reason they trusted the standalone software more.

> *I would think that [the standalone tool] would be the safer one. That's just the feeling I got from it... the browser extension asks you for permissions when you go in, it asks to see what tabs you have and all that stuff. (P31)*

We expected the differences in user trust to instead stem from the degree to which they saw the internal workings of the software, and the result of encrypting or decrypting a message (H04). Only one out of 36 participants expressed concern about using an opaque version of the extension.

> *I guess the [standalone] one ... it looks very encoded. The [integrated] one, it's black, but it doesn't actually... (P15)*

More participants judged user error to be an important factor in trust than software transparency.

We did not create a hypothesis for the effect of transparency on usability, but we did find a noticeable, yet not statistically significant difference in the usability scores ($p > 0.05$) of the transparent and opaque versions of our integrated tool. One user specifically mentioned their preference for opacity from a usability perspective, which we recommend exploring in future work:

> *After you encrypted it, it shows a bunch of letters which is pretty long (P35)*

We have strong evidence for our first two hypotheses (H01 and H02): that it is possible to make usable integrated encrypted email tools, and that they are preferred over similar standalone versions. Our replication of Ruoti et al.'s study contradicted the lack of preference they witnessed between integrated and standalone solutions. While roughly equal numbers of participants preferred each of their tools, we saw an overwhelming preference for an integrated solution.

Our evidence contradicts hypotheses H03 and H04 on the basis of trust preferences in different types of privacy software. Our methodology separated the two aspects of Ruoti et al.'s original tools—the level of transparency and integration—to find the design features that contribute to user feelings of trust. We found that it was the level of integration, and not the involvement and awareness of the encryption process that led some users to believe their information was more or less secure. This finding is contradictory to the hypothesis of Ruoti et al. We also found that the majority of participants did not feel a different level of trust towards either of our tools. Instead, they either did not think about trust at all or based trust on company reputation and tool popularity rather than on software features.

5.5 Limitations

We conducted extensive focus group and pilot study sessions in an attempt to make all three versions of our encrypted email tools as usable as possible. This was to eliminate confounding factors due to tool design, and allow us to focus on the factors of transparency and integration in determining trust and usability. We aimed to make the integrated and standalone clients as similar as possible, within the constraints imposed by the level of integration. The System Usability Scale ratings discussed in Section 5.1 provide evidence for the accomplishment of this goal. However, it is still possible that minor differences affected the level of trust and quality of user experience that participants expressed.

It is also likely, given that we received the majority of our participants through our on-campus recruiting efforts (92% of our participants were students), that our participant pool was not representative of the general public. The age and tech-savvy bias of our participants may give a skewed vision of the overall usability of our tool, and could have an additional effect in the trust themes we discuss in Section 5.3.

There was a minor difference in the task order between the integrated and standalone portions of the study. As discussed in Section 4, the integrated tasks had participants decrypt a received message before encrypting a reply. In contrast, the standalone tasks had participants first send an encrypted message and then decrypt a reply. Our reasoning behind these differences was to emulate a normal workflow for each tool, given the differences in the setup procedures. We think it is unlikely that these differences would invalidate our results. As mentioned in Section 5.4, it is possible that the installation of the browser extension fed into the lack of trust a few users expressed in the integrated tool.

6. DESIGN IMPLICATIONS

The strongest result we observed was an overwhelming preference for the integrated, automatic encryption tool. This preference was apparent from both the comparative usability scores and the feedback in the questionnaires and post-study interview. Users liked the seamless integration of the encryption functions with a system they already use and are familiar with. They also recognized the tediousness of frequently copy-pasting ciphertext from a standalone client. At the same time, participants reported that they considered trustworthiness as a factor when choosing software. Some based this trust on intrinsic properties of the tools (such as their degree of integration), while others based trust on the reputation and popularity of the software developers.

This gives us more insight into designing usable encryption tools that also engender trust from users. With widespread adoption and ongoing use being the eventual goals, our results suggest we should focus on making integrated solutions more straightforward and trustworthy, rather than making standalone systems more usable. Furthermore, the usability of standalone systems is limited by their nature: key management will continue to be a manual and tedious task if standalone tools are unable to interact with users' contact lists. While integrated systems can automatically select encryption keys to match an email's recipients, a standalone tool requires users take care to select the correct recipient twice—once in the standalone encryption tool, and once in the email client itself. The other main usability complaint we received regarded the continual need to copy and paste ciphertext; this, again, is solved only by integrating the encryption software with the webmail client.

There are two main avenues we can explore to effect user trust of integrated systems. The first is to re-enforce the notion of sandboxing—several participants cited sandboxing as their reason for placing higher trust on the standalone system. This trust, however, overlooks the fact that browser extensions and plugins are generally placed under significantly more restrictions than desktop applications. Both systems polled servers for keys, and therefore had the ability to send and receive metadata without the explicit knowledge of the user (although it could be inferred indirectly from the invitation task). Through design, we may be able to indicate to users that their keys are stored locally with the integrated system as well, and re-enforce the notion that private information they enter will not be sent to their email provider, browser, or software developers. The second method of inspiring trust is to publish the integrated tool under a reputable developer's name, and provide the download from a trustworthy source (such as the major app stores). It is also important to permit, and even encourage, reviews from users of the software in a place where they can be viewed by other users. Users frequently put their trust in word-of-mouth opinions, and some would also search the Internet for impressions of the software from reputable authorities (such as reviews in online news outlets).

7. CONCLUSION

In this work, we showed that encrypted email clients based on the OpenPGP standard can be both usable and well-liked by ordinary webmail users. We showed that users have a strong preference for encryption tools that integrate tightly with their existing email client, as opposed to standalone encryption software that must be used separately. Our results demonstrate that, contrary to previous research findings, such standalone software does not inspire trust by forcing users to interact with encrypted objects. Rather, we found that a fraction of users believe desktop applications are more likely to operate in a purely offline manner, refraining from sending user data back to the developers, as compared to browser extensions. The majority of users, however, felt unable to make a distinction in trustworthiness between the two types of software alone, and would rather defer to popular opinion by way of online reviews or company reputation.

This work shows methods that could be applied to improve the usability of existing PGP tools. When working with unfamiliar concepts such as encryption, explicit UI labels help users feel more confident versus unlabelled icons. Common mistakes (such as sending plaintext email unknowingly) can be partially defended against with prominent indicators. Semi-automated key distribution allows users to send secure email without an understanding of how PGP works. We hope that successful integration with the popular webmail service Gmail will encourage these services to move towards making E2E encryption the default for all users.

8. ACKNOWLEDGMENTS

This work was made possible with funding from the Natural Sciences and Engineering Research Council of Canada and the Ontario Research Fund.

9. REFERENCES

[1] Overview of projects working on next-generation secure email. `https://github.com/OpenTechFund/secure-email`. Accessed Feb 2015.

[2] E. Atwater, C. Bocovich, U. Hengartner, E. Lank, and I. Goldberg. Paper companion website. `https://crysp.uwaterloo.ca/software/leadingjohnny/`, 2015.

[3] A. Bangor, P. Kortum, and J. Miller. Determining what individual SUS scores mean: Adding an adjective rating scale. *Journal of usability studies*, 4(3):114–123, 2009.

[4] P. Bright and D. Goodin. Encrypted e-mail: How much annoyance will you tolerate to keep the NSA away? *Ars Technica*, June 2013.

[5] J. Brooke. SUS-a quick and dirty usability scale. *Usability evaluation in industry*, 189:194, 1996.

[6] S. L. Garfinkel and R. C. Miller. Johnny 2: A user test of key continuity management with S/MIME and Outlook Express. In *Proceedings of the 2005 Symposium on Usable Privacy and Security*, SOUPS '05, pages 13–24, New York, NY, USA, 2005. ACM.

[7] S. Gaw, E. W. Felten, and P. Fernandez-Kelly. Secrecy, flagging, and paranoia: Adoption criteria in encrypted email. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '06, pages 591–600, New York, NY, USA, 2006. ACM.

[8] Gillian Andrews. Usability Report: Proposed Mailpile Features. `https://openitp.org/sup/field-notes/`, December 2014.

[9] Google End-to-End Wiki. Key Distribution. `https://github.com/google/end-to-end/wiki/Key-Distribution`. Accessed Feb 2015.

[10] M. Green. The Daunting Challenge of Secure E-mail. *The New Yorker*, November 2013.

[11] G. Greenwald, E. MacAskill, and L. Poitras. Edward Snowden: the whistleblower behind the NSA surveillance revelations. *The Guardian*, 2013.

[12] P. Gutmann. Why isn't the Internet secure yet, dammit. In *AusCERT Asia Pacific Information Technology Security Conference 2004; Computer Security: Are we there yet?*, May 2004.

[13] B. Laurie, A. Langley, and E. Kasper. Certificate transparency. RFC 6962, RFC Editor, June 2013.

[14] M. Lee. Ed Snowden Taught Me To Smuggle Secrets Past Incredible Danger. Now I Teach You. *The Intercept*, October 2014.

[15] C. T. Moecke and M. Volkamer. Usable secure email communications: criteria and evaluation of existing approaches. *Inf. Manag. Comput. Security*, 21(1):41–52, 2013.

[16] K. Renaud, M. Volkamer, and A. Renkema-Padmos. Why doesn't Jane protect her privacy? In E. De Cristofaro and S. Murdoch, editors, *Privacy Enhancing Technologies*, volume 8555 of *Lecture Notes in Computer Science*, pages 244–262. Springer International Publishing, 2014.

[17] S. Ruoti, N. Kim, B. Burgon, T. van der Horst, and K. Seamons. Confused Johnny: When automatic encryption leads to confusion and mistakes. In *Proceedings of the Ninth Symposium on Usable Privacy and Security*, SOUPS '13, pages 5:1–5:12, New York, NY, USA, 2013. ACM.

[18] S. Sheng, L. Broderick, C. Koranda, and J. Hyland. Why Johnny still can't encrypt: evaluating the usability of email encryption software. In *2006 Symposium On Usable Privacy and Security - Poster Session*, 2006.

[19] The Free Software Foundation. The GNU Privacy Handbook. `https://www.gnupg.org/gph/en/manual.html`, 1999.

[20] W. Tong, S. Gold, S. Gichohi, M. Roman, and J. Frankle. Why King George III can encrypt. `http://randomwalker.info/teaching/spring-2014-privacy-technologies/king-george-iii-encrypt.pdf`, 2014.

[21] A. Whitten and J. D. Tygar. Why Johnny can't encrypt: A usability evaluation of PGP 5.0. In *Proceedings of the 8th Conference on USENIX Security Symposium - Volume 8*, SSYM'99, pages 14–14, Berkeley, CA, USA, 1999. USENIX Association.

APPENDIX

A. USER STUDY INSTRUCTIONS

These are the questionnaires and instructions given to participants for the pilot study and main study. These questions are available in HTML format from our website [2], in addition to the script used to record answers and the instructions for the focus group described in Section 3.1.

A.1 Pre-study questionnaire

Welcome to our study!

Thank you for your participation. During this study, you will be asked to perform certain tasks using Gmail and then provide feedback to help us improve our software. During the course of this study, all acts taking place on the screen will be recorded along with audio of anything we discuss. This will help us learn whether or not our software is easy to use.

You will have access to a temporary Gmail account for use in completing tasks during this study. You will not be asked to use your own Gmail login name or password at any time. Do not enter or access any of your own personal data during the study since everything on the screen will be recorded.

Are you a student? (Yes; No)

What is your occupation or area of study?

What is your gender? (Male; Female; Other)

What is your approximate age? (18-23; 24-30; 31-40; 41-50; 51-65; 66+)

How long have you been a Gmail user? (I don't use Gmail; <1 Year; 1-2 Years; 3+ Years)

Approximately how often do you use webmail (Gmail)?

How would you rate your level of computer expertise? (Minimal; Minimal to Average; Average; Advanced; Expert)

Have you ever sent private or sensitive information via Web email or Facebook? (Yes; No)

If so, how did you send that information? (briefly explain):

How important is maintaining the privacy of your messages containing sensitive information? (Very Important; Important; Neither important nor unimportant; Unimportant; Very Unimportant)

Have you ever encrypted an email or Facebook message? (Yes; No)

If you answered yes to the previous question, please briefly explain how you did so.

A.2 Standalone client tasks

Message Protector Tasks

Message Protector (MP) is a computer program that allows users to protect Internet messages (e.g., email, Facebook private messages) via encryption. In this portion of the study, you will execute various tasks that comprise the primary functionality of MP and answer a few related questions.

Task 1: Set up MP

Open Message Protector from the toolbar on the left-hand side of your screen.

[image] This is what the MP icon looks like.

MP requires an email address and the email account password to allow the user's contacts to be able to read their protected messages. For this study, we have created the following test account for you to use:

Email Address: `jane.doe.cs889@gmail.com`

Password: `xxxxx`

Allow the following contacts to read your protected messages: `randomFriend@hotmail.com`, `mom@familyWebsite.com`, and `study.coordinator.cs889@gmail.com`.

In this scenario, you will encrypt and decrypt email messages with MP. The encrypted emails are sent and received, however, using Gmail. Click here to go to Gmail and log in with the credentials above.

Task 2: MP Email Encryption

Use MP to encrypt an email and send it to `study.coordinator.cs889@gmail.com` (note that you will need to copy and paste the message contents from the MP window into a new Gmail message). Include the phrase "The last four digits of my SSN are 6789" in the message.

Task 3: MP Email Decryption

After completing the previous task, you will receive a protected reply email from `study.coordinator.cs889@gmail.com`. Use MP to decrypt the message.

[image] It may take a few minutes for the email to arrive.

Type the decrypted message below:

Task 4: Adding contacts

Try sending a **secure** message to

`study.coordinator2.cs889@gmail.com` *(notice the "2" in "coordinator2")*

using MP. You will notice they do not have MP installed yet, so follow the instructions that appear to send them an invitation.

[image] It will take a few minutes for the study coordinator to receive and accept your invitation.

Once they have accepted the invitation, send the secure message. Include the phrase "The vault combination is 1234" in the message.

A.3 Standalone post-study questionnaire

SUS Questions:

Please answer the following question about MP. Try to give your immediate reaction to each statement without pausing to think for a long time. Mark the N/A column if you don't have a response to a particular statement. *Choose from 1 (strongly disagree) to 5 (strongly agree)*

1. I think that I would like to use this system frequently
2. I found the system unnecessarily complex
3. I thought the system was easy to use
4. I think that I would need the support of a technical person to be able to use this system
5. I found the various functions in this system were well integrated
6. I thought there was too much inconsistency in this system
7. I found the system very cumbersome to use
8. I would imagine that most people would learn to use this system very quickly
9. I felt very confident using the system
10. I needed to learn a lot of things before I could get going with this system
11. My level of understanding of MP directly affects whether I would use it to protect my email messages.

Remaining Questions:

Who can read messages that you protect with MP? (Anyone that has MP installed, receives the message, and that I have selected to communicate with securely; Anyone who receives the message and who I have selected to communicate with securely; Anyone who receives the message; Anyone who has MP installed; I don't know)

After MP is installed, what actions must recipients take to read MP protected messages? (Access the MP software; Copy the message and paste to MP; Copy the message, paste to MP, click the Encrypt button; Copy the message, paste to MP, click the Decrypt button; I don't know)

How often would you use MP to protect your email messages? (Always; Very Often; Occasionally; Rarely; Very Rarely; Never)

What did you like about MP?

How could MP be improved?

A.4 Integrated client tasks

Mailvelope Tasks

Mailvelope is a computer program that allows users to protect email messages via encryption. In this portion of the study, you will execute various tasks that comprise the primary functionality of Mailvelope and answer a few related questions.

Task 1: Set up Mailvelope

Please login to our test Gmail account with the login name and password shown below. Read the first message and follow the instructions given in the message.

Click here to open Gmail

Username: `jane.doe.cs889@gmail.com`

Password: `xxxxx`

At some point, you will be prompted for your name and email. Please use the information below:

Name: `Jane Doe`

Email: `jane.doe.cs889@gmail.com`

Task 2: Mailvelope Email Decryption

Wait for a new message from the study coordinator.

[image] It may take a few minutes for the email to arrive.

Read the message and enter the secret phrase below:

Proceed to the next task.

Task 3: Mailvelope Email Encryption

Send a **secure** message to

`study.coordinator.cs889@gmail.com`

using Mailvelope. Include the secret phrase you were given in your message.

Task 4: Adding new Mailvelope Contacts

Try sending a **secure** message to

`study.coordinator2.cs889@gmail.com` *(notice the "2" in "coordinator2")*

using Mailvelope. You will notice they do not have Mailvelope installed yet, so follow the instructions that appear to send them an invitation.

[image] It will take a few minutes for the study coordinator to receive and accept your invitation.

Once they have accepted the invitation, send the secure message. Include the phrase "The vault combination is 1234" in the message.

A.5 Integrated post-study questionnaire

(Same SUS questions from A.3.)

Remaining Questions:

What did you like about Mailvelope?

What did you dislike about Mailvelope and how would you like it to be changed?

If you started using Mailvelope on your own, would you prefer protection for new messages to be? (Always on; Only on for the messages I decide are private; Usually off, unless I click a separate button on the Gmail page)

A.6 Post-study questionnaire

Post Study Survey

Please answer the following questions. Try to give your immediate reaction to each statement without pausing to think for a long time. Mark the N/A column if you don't have a response to a particular statement.*Choose from 1 (strongly disagree) to 5 (strongly agree)*

1. I trust Gmail employees to not disclose, misuse, or abuse my email messages
2. I am concerned about Gmail scanning my messages
3. I worry that some messages aren't really from who they say they are from
4. I feel safe sending important information through email
5. I feel safe creating accounts with usernames and passwords on new sites
6. I feel safe installing browser extensions or plugins
7. Creating accounts for new websites is easy
8. Installing browser extensions is easy
9. I feel safe clicking on links in email messages
10. I feel safe clicking on links in email messages from people I know
11. I never click on links in email messages
12. I would trust a company other than Facebook or Gmail (e.g., MP, Mailvelope) to protect my email messages
13. I feel that it is important to encrypt my emails and messages that contain sensitive or private information
14. I would use a different Internet Encryption tool for every website that I store or share sensitive information

Have you installed browser extensions, add-ons or plugins before today? (Yes; No)

What has prevented you from installing browser extensions, add-ons or plugins in the past?

When deciding whether you will trust a browser extension, add-on or plugin, what influences your decision?

Have you ever been asked to send sensitive information you were not comfortable sending through email? (Yes; No)

What type of sensitive information were you asked to send?

Did you send the requested information? (Yes; No)

Have you ever received information you were not comfortable receiving through email? (Yes; No)

What type of sensitive information did you receive?

Which system would you prefer to use? (MP; Mailvelope; None of the above)

Please explain your answer to the previous question:

Thank you for completing our study! Before you go, please let us know if there is any additional information you would like us to have.

Additional Info:

B. FULL STUDY DATA

Tables 2 and 3 list the frequency with which various issues were encountered by participants using the integrated and standalone tools, respectively. Table 4 shows which tool each participant preferred, and their stated reason for their choice.

This data, plus the answers provided to our other survey questions described above, are available in spreadsheet form from the paper's companion website [2].

Behaviour ID	Occurrences	Description
i-1-1	2	Asked how to install the browser extension
i-1-2	1	Incorrectly enterred data during setup
i-1-3	5	Asked if they were supposed to install extension after seeing warning
i-1-4	6	Didn't click "Finish" button at the end of the setup phase
i-1-5	2	Opened advanced setup options and then closed them
i-1-7	1	Didn't know they had successfully installed extension
i-2-1	1	Immediately clicked overlay without reading it
i-2-2	1	Was unsure about clicking the "Encrypt message checkbox
i-3-1	1	Almost sent an unencrypted message, but encrypted at the last minute
i-3-2	1	Didn't see encrypte checkbox at first
i-3-3	1	Tried clicking encrypt checkbox first, and then composing the message
i-3-4	3	Tried to use the standalone client to send a secure email
i-3-5	7	Confused as to how to send a secure message
i-3-6	1	Closed compose window to modify an encrypted message instead of unchecking the checkbox
i-3-7	5	"This message has been encrypted" didn't show up (bug)
i-3-8	2	Were unsure they had sent the right message
i-4-1	2	Encrypted message only for Jane Doe (bug)
i-4-2	10	The "Encrypt message checkbox failed to appear (bug)
i-4-3	1	Sent two emails instead of one
i-4-4	1	Confused about how to get someone else to install the system
i-4-5	1	Didn't know how to modify the encrypted message
i-4-6	14	Sent encrypted email before knowing the new contact had joined
i-x-1	1	Skipped a task
i-x-2	3	Sent a plaintext message

Table 2: Behaviours encountered in the performance of integrated tool tasks.

Behaviour ID	Occurrences	Description
s-1-1	6	Forgot to add contacts
s-1-2	1	Added contact didn't show (user error)
s-1-3	2	Didn't enter contact names
s-1-4	2	Unsure how to add a contact
s-1-5	3	Expressed uncertainty as to how contacts were loaded/stored in MP
s-1-6	1	Asked whether tool would read read protected messages
s-2-1	1	Thought they had made a mistake while encrypting (but had not)
s-2-2	1	Did not select the correct contact to encrypt to
s-2-3	1	Appended plaintext to ciphertext after encrypting
s-2-4	1	Put plaintext in Gmail draft before encrypting
s-2-5	2	Did not know to copy encrypted message into Gmail
s-3-1	1	Incorrectly copied ciphertext
s-3-2	1	Tried to encrypt before adding contacts
s-4-1	18	Sent encrypted email before knowing the new contact had joined
s-4-2	1	Did not see or understand added contact notification
s-4-3	1	Encrypted invitation instead of sending it in plaintext
s-4-4	4	Confused by invitation process
s-4-5	2	Sent invitation but didn't add the contact to the system
s-4-6	1	Sent the invitation to the wrong person
s-4-7	3	Added new contact twice
s-x-1	1	Sent only plaintext

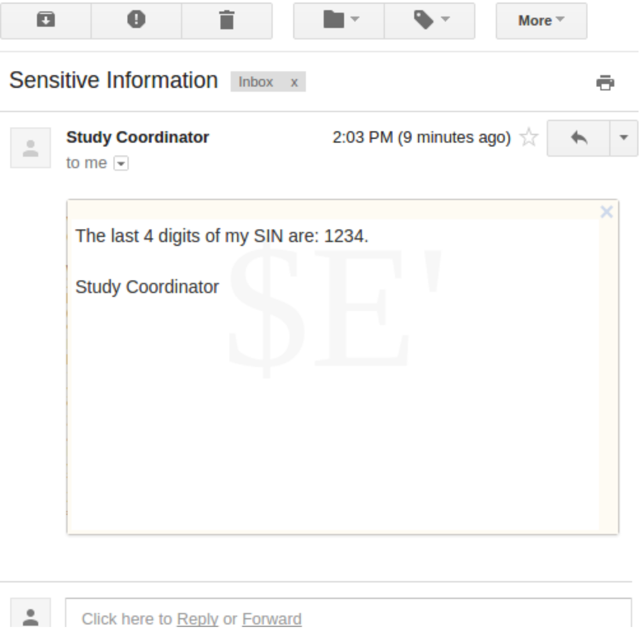
Table 3: Behaviours encountered in the performance of standalone tool tasks.

ID#	Preference	Reason for choice
P07	Integrated	I find it more convenient in that I don't have to open up another program to send the encrypted message, I can just choose whether or not to encrypt it when sending an email using gmail.
P11	Integrated	the ease of encryption options
P15	Integrated	It seems easier to use because it is integrated in the Gmail interface. However, I'd want to know a bit more about it before starting to use it...
P19	Integrated	more integrated into the user interface and makes sending an encrypted, secure message less cumbersome.
P23	Integrated	Easier to use because it can be sent right through my email.
P27	Integrated	It is very convenient as it is one go and less cumbersome than MP.
P31	Integrated	It's easier since it's integrated into gmail. I don't know if it's safer though
P35	Integrated	Easier to use
P39	Integrated	I found it more convenient because you don't need another software on. You don't need to copy and paste back and forth
P09	Integrated	integrated in gmail, not a seperate software and no need of extra copy+paste
P13	Integrated	[stand. tool] involves installing a new program whereas [int. tool] is directly through google.
P17	Integrated	The plugin is much easier to use for me, actually I don't know whether I should trust these two systems. If I don't know these systems are developed by students in our school, I will doubt the security level of these systems.
P21	Neither	Currently, I do not feel there is a need to encrypt my sensitive information. I typically do not share sensitive information via email. Also, email encryption is too troublesome, especially when no one around me has used it before (or send me invitations to use such a service to decrypt their messages).
P25	Standalone	Although it is not directly integrated into gmail, it was simpler to use and very straightforward. As well, I don't think having an 'encrypt' button at the bottom of every email I send is necessary, and might cause some confusion if accidentally pressed.
P29	Standalone	Even though it requires more steps for the encryption and decryption process, I feel as if it's more user friendly and still provides the same security.
P33	Integrated	Seems easier.
P37	Integrated	easy and simple getting protected without doing extra steps
P41	Integrated	seems a lot easier to use and add contacts
P12	Integrated	simple to use
P16	Integrated	wasn't its own program so it required less window switching. It also required less copy and pasting of the encrypted messages so would be quicker and easier to use, as I can see people accidentally not copying the entire message by missing a character and the message may become incomplete.
P20	Integrated	has a direct link in gmail. It makes it easier and less cumbersome. In terms of adding new contacts both systems require you to send an invitation. Both do not have any extra password requirement for decrypting the message. However, if [stand. tool] software has the feature that it can be downloaded by invitation only, it might make a difference. Otherwise, I don't see much difference between the two.
P24	Integrated	much more convenient to use, to the point where I wouldn't mind encrypting even everyday, non-sensitive emails with it.
P28	Integrated	seemed easier to use and less tedious
P32	Integrated	I liked the fact that it is really integrated into the browser, instead of in [stand. tool], where it is really inconvenient to switch windows just to encrypt and decrypt messages. It also seemed safer since it is an extension on the browser, compared to software on my computer.
P36	Integrated	It is a simpler process and doesn't require the constant switching between two different apps.
P40	Integrated	easier, for extremely important information, I would talk face to face or through cellphone.
P08	Standalone	I know I stated earlier that having a system like this integrated into your email would help but it acts as a different entity that is on your desktop and not integrated online. So it feels more secure to encrypt and decrypt messages separately.
P10	Integrated	simpler and faster to use as a browser extension and messages can be encrypted within Gmail
P14	Integrated	I find it is more easy to use and does the same job.
P18	Integrated	I don't have to access a separate window in order to use it
P22	Standalone	I would like both to use as one program. I like that [int. tool] is intregrated in gmail and makes it VERY easy to encrypt. But I also like that [stand. tool] could send private information over facebook, or even text messages. I think it is complex because it requires an additional program to be open, and copy and pasting to occur which I could mess up on.
P26	Standalone	I enjoyed its ease of use. It was similar to a "translation program" you would use like Google translate. This would make it easier for users that aren't familiar with using computers and foreign programs. [int. tool] was easier in the sense that it was integrated into Google. Users may not be comfortable with that since it's integrated into gmail, where [stand. tool] was a separate program not linked to gmail. This might give the sense that [int. tool] is "less safe" and more intrusive. [Int. tool] also didn't have a decryption function
P30	Integrated	I liked being able to use something that complimented the email system I currently use instead of having to learn something new then apply that to what I already use.
P34	Integrated	Like I stated before, it felt much more comfortable to use because it was better integrated into the email system. I didn't have to open up another program and do a bunch of other functions before sending an encrypted message.
P38	Integrated	It is less complicated and it encrypt it right away in the email

Table 4: The preference chosen by each participant, and the reason supplied for their choice.

C. SCREENSHOTS

This section contains screenshots supplementing Figures 2 and 3 for documenting the tools built for our study. Figures 7, 8, and 10 show various screens and prompts for our integrated encryption tool. Figures 9 and 11 show screenshots of our standalone tool. Note that we kept the branding of the tools during the user study (the integrated tool was branded as Mailvelope and the standalone tool was branded as Message Protector). We did this to identify the tools to the participants and give them memorable names to refer to during the questionnaire and interview.



Sensitive Information

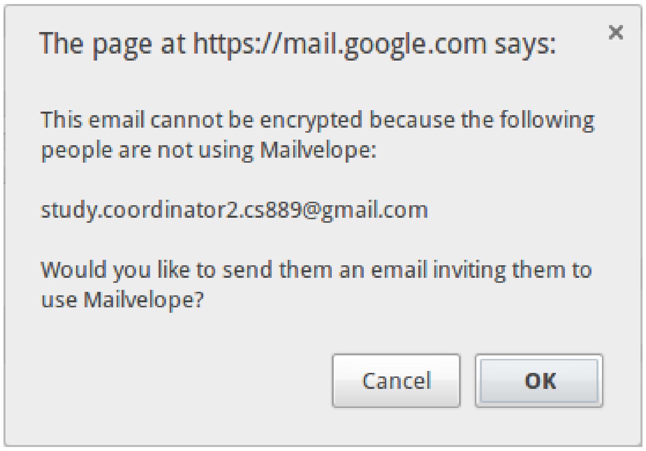
Study Coordinator
to me
2:03 PM (9 minutes ago)

The last 4 digits of my SIN are: 1234.

Study Coordinator

Click here to Reply or Forward

Figure 7: Screenshot of a decrypted message overlay in our integrated tool interface (the watermark is a security feature of Mailvelope, the underlying source code upon which we built our integrated tools)



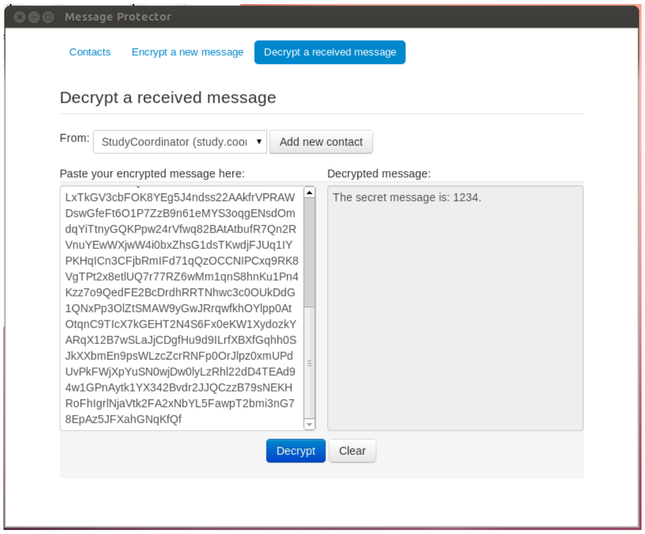
The page at https://mail.google.com says:

This email cannot be encrypted because the following people are not using Mailvelope:

study.coordinator2.cs889@gmail.com

Would you like to send them an email inviting them to use Mailvelope?

Figure 8: Prompt displayed by the integrated tool when no recipient key is found



Decrypt a received message

From: StudyCoordinator (study.coo

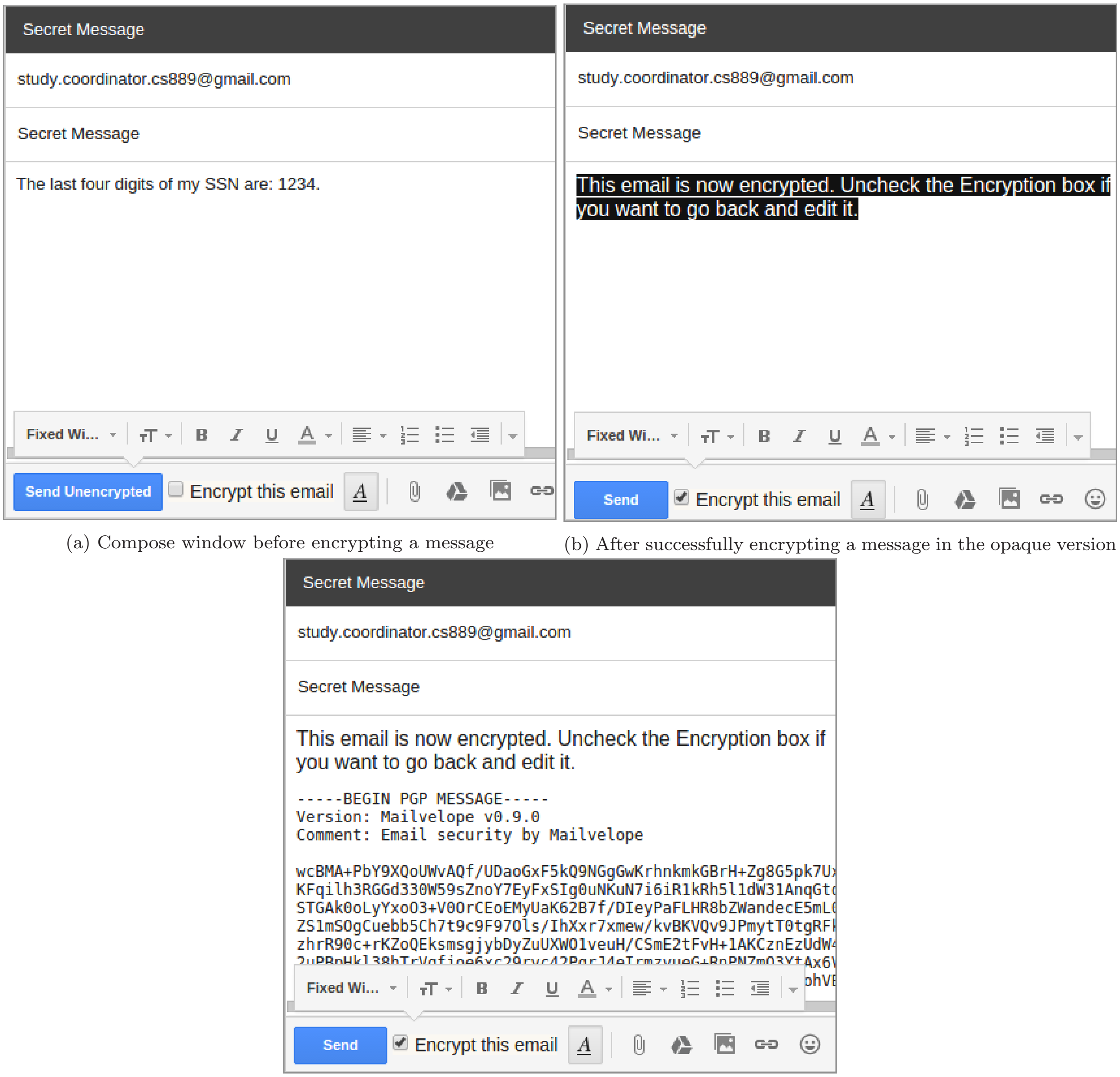
Paste your encrypted message here:

```
LxTkGV3cbFOK8YEg5J4ndss22AAkfrVPRAW
DswGfeFt6O1P7ZzB9n61eMYS3oqgENsdOm
dqYiTtnyGQKPpw24rVfwq82BAtAtbufR7Qn2R
VnuYEwWXjwW4i0bxZhsG1dsTKwdjFJUq1IY
PKHqICn3CFjbRmIFd71qQzOCCNIPCxq9RK8
VgTPt2x8etlUQ7r77RZ6wMm1qnS8hnKu1Pn4
Kzz7o9QedFE2BcDrdhRRTNhwc3c0OUkDdG
1QNxPp3OlZtSMAW9yGwJRrqwfkhOYlpp0At
OtqnC9TIcX7kGEHT2N4S6Fx0eKW1XydozkY
ARqX12B7wSLaJjCDgfHu9d9ILrfXBXfGqhh0S
JkXXbmEn9psWLzcZcrRNFp0OrJlpz0xmUPd
UvPkFWjXpYuSN0wjDw0lyLzRhl22dD4TEAd9
4w1GPnAytk1YX342Bvdr2JJQCzzB79sNEKH
RoFhIgrlNjaVtk2FA2xNbYL5FawpT2bmi3nG7
8EpAz5JFXahGNqKfQf
```

Decrypted message:

The secret message is: 1234.

Figure 9: Decrypting a message using our standalone interface



Secret Message

study.coordinator.cs889@gmail.com

Secret Message

The last four digits of my SSN are: 1234.

(a) Compose window before encrypting a message

Secret Message

study.coordinator.cs889@gmail.com

Secret Message

This email is now encrypted. Uncheck the Encryption box if you want to go back and edit it.

(b) After successfully encrypting a message in the opaque version

Secret Message

study.coordinator.cs889@gmail.com

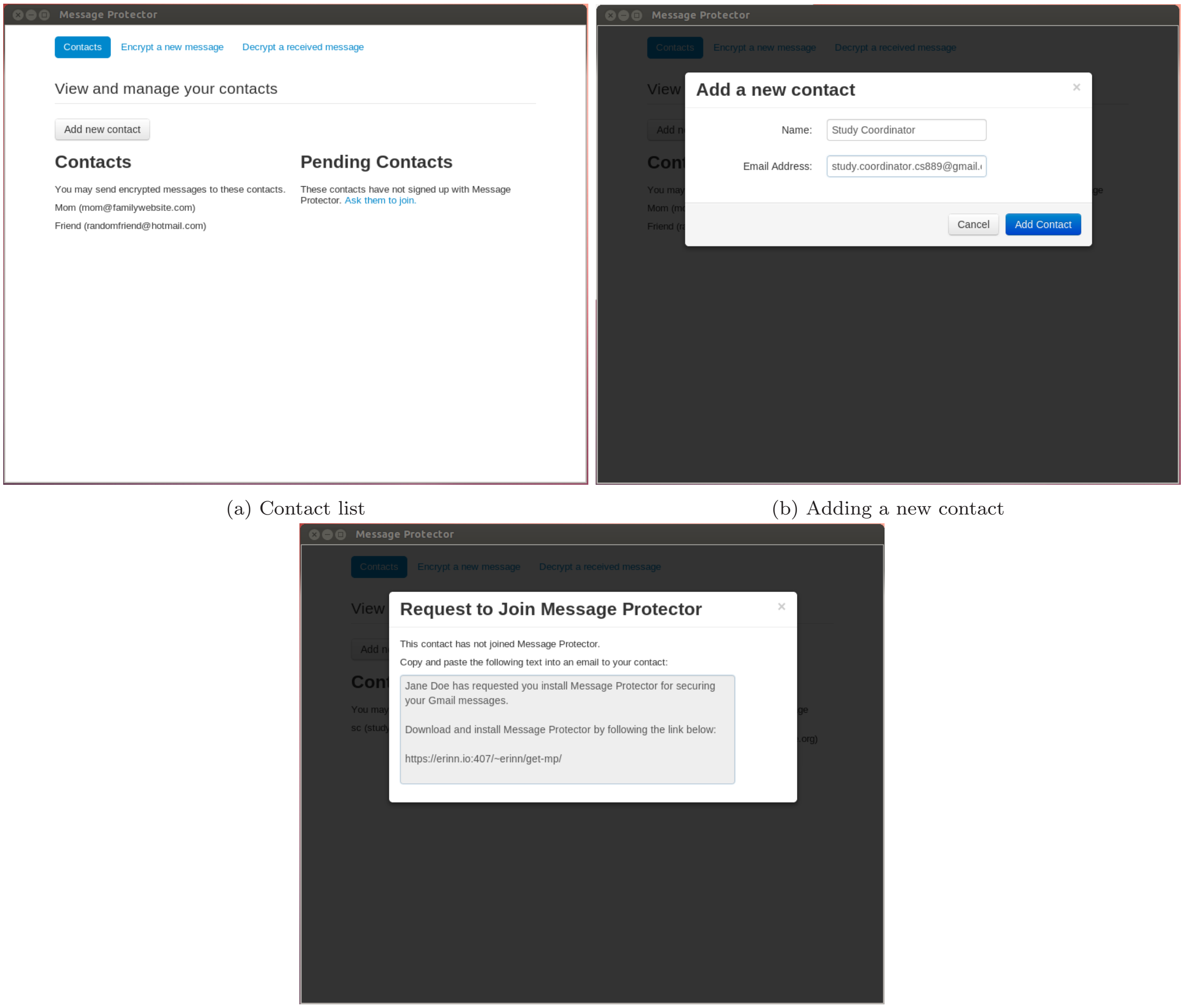
Secret Message

This email is now encrypted. Uncheck the Encryption box if you want to go back and edit it.

```
-----BEGIN PGP MESSAGE-----
Version: Mailvelope v0.9.0
Comment: Email security by Mailvelope

wcBMA+PbY9XQoUWvAQf/UDaoGxF5kQ9NGgGwKrhnkmkGBrH+Zg8G5pk7U
KFqilh3RGGd330W59sZnoY7EyFxSIg0uNKuN7i6iR1kRh5l1dW31AnqGt
STGAk0oLyYxoO3+V0OrCEoEMyUaK62B7f/DIeyPaFLHR8bZWandecE5mL
ZS1mSOgCuebb5Ch7t9c9F97Ols/IhXxr7xmew/kvBKVQv9JPmytT0tgRF
zhrR90c+rKZoQEksmsgjybDyZuUXWO1veuH/CSmE2tFvH+1AKCznEzUdW
2uPBnHkl38hTrVqfioe6xc29rvc42Pqrl4eIrmzvueG+RnPNZmO3YtAx6
ohV
```

(c) After successfully encrypting a message in the transparent version

Figure 10: Screenshots of encryption in our integrated interface



View and manage your contacts

Contacts

You may send encrypted messages to these contacts.

Mom (mom@familywebsite.com)

Friend (randomfriend@hotmail.com)

Pending Contacts

These contacts have not signed up with Message Protector. Ask them to join.

(a) Contact list

Add a new contact

Name: Study Coordinator

Email Address: study.coordinator.cs889@gmail.c

(b) Adding a new contact

Request to Join Message Protector

This contact has not joined Message Protector.

Copy and paste the following text into an email to your contact:

Jane Doe has requested you install Message Protector for securing your Gmail messages.

Download and install Message Protector by following the link below:

https://erinn.io:407/~erinn/get-mp/

(c) Prompt displayed when no recipient key is found

Figure 11: Screenshots of our standalone interface

Usability of Augmented Reality for Revealing Secret Messages to Users but Not Their Devices

Sarah J. Andrabi
UNC Chapel Hill
sandrabi@cs.unc.edu

Michael K. Reiter
UNC Chapel Hill
reiter@cs.unc.edu

Cynthia Sturton
UNC Chapel Hill
csturton@cs.unc.edu

ABSTRACT

We evaluate the possibility of a human receiving a secret message while trusting *no* device with the contents of that message, by using visual cryptography (VC) implemented with augmented-reality displays (ARDs). In a pilot user study using Google Glass and an improved study using the Epson Moverio, users were successfully able to decode VC messages using ARDs. In particular, 26 out of 30 participants in the Epson Moverio study decoded numbers and letters with 100% accuracy. Our studies also tested assumptions made in previous VC research about users' abilities to detect active modification of a ciphertext. While a majority of the participants could identify that the images were modified, fewer participants could detect *all* of the modifications in the ciphertext or the decoded plaintext.

1. INTRODUCTION

In the face of massive surveillance by nation-state-level actors (e.g., [22, 30]), including the implantation of surveillance functionality in commodity device firmware (e.g., [17]), it appears that truly private electronic communication is as challenging today as ever. At the core of this problem are the complex cryptographic operations that must be performed to encrypt and decrypt messages: These operations are too complex for humans to perform themselves, and so they must use devices to do so—perhaps devices that might have already had their privacy compromised. To achieve truly private communication, then, it would seem necessary to eliminate devices from the trusted computing base (TCB), i.e., to have humans themselves perform the cryptographic operations.

History is rife with examples of private communication performed by humans without devices, usually because capable devices were not yet available. While most of these ciphers are trivially breakable today, a notable exception is the one-time pad, which is both perfectly private and has encryption and decryption functions involving only exclusive-or operations [16]. One-time pads were a method of choice for spies in World War II for protecting communications with their home countries, performing the exclusive-or of messages manually against keys encoded on, e.g., a paper pad that they brought with them (e.g., [29]). This idea was modernized in 1994 with the invention of *visual cryptography* (VC) [26], in which keys are encoded on visual transparencies and manual exclusive-or operations are replaced with overlaying these transparencies on suitably encoded ciphertexts to reveal their contents.

The practical difficulties associated with one-time pads are well known. Most notable is that they require a quantity of key material (in the above proposals, on paper tape or transparencies) equal to the total size of all messages to be transmitted. In this paper we explore the feasibility of making the one-time pad, and specifically its implementation in VC, practical using *augmented reality* head-mounted displays (ARDs) such as Google Glass, Epson Moverio, Sony SmartEyeGlass, or Microsoft HoloLens, while still ensuring that no single device is trusted with the contents of a message to the user. We evaluate the efficacy of storing and rendering one-time pads (encoded for use in VC) in an ARD, which the user visually aligns over an encoded message rendered on another display to reveal its contents. If workable, this design should largely eliminate the problems associated with storage of one-time pads and in fact can support their generation pseudorandomly (e.g., from a secret key per sender) with little practical loss of security.

In this paper we show that VC via ARDs is in fact possible and reasonably usable, first using a small pilot study with Google Glass and then a formal user study using the Epson Moverio. For conveying messages we used individual letters and numbers encoded as images. Of the 30 participants, 26 in the formal user study decoded numbers and letters with 100% accuracy. The other four participants decoded at least 80% of the alphanumeric plaintexts accurately. The median time taken by participants to identify the decoded letters and numbers was 8.88 seconds. As indicated by our study, however, several challenges need to be addressed including the small field of view of ARDs, head tracking, and image stabilization.

Moreover, since numerous communication protocols involving VC have made assumptions about users' abilities to detect active modification of a ciphertext or the plaintext that results from it (e.g., [25, 31, 15, 9, 6]), we also evaluate users' abilities to do so. Through the Epson Moverio study, we assess whether users were able to identify invalid decoded plaintexts resulting from malicious modifications to the ciphertext. A large majority of our participants were able

Symposium on Usable Privacy and Security (SOUPS) 2015, July 22–24, 2015, Ottawa, Canada.

to detect the modifications, even though fewer were able to identify all of the changes in an image. Even in the presence of the *least* number of modifications (depending on the study) to the recovered plaintext, more than 80% of the participants detected the presence of the modifications.

We also conducted an online user study to analyze users' abilities to detect changes in ciphertexts directly. Again, a large majority of participants detected that the images were modified. We explored the extent of modification needed in order for a user to detect it, and based on our current implementation, small modifications were detected by most of the participants. Our findings indicate that augmented reality provides a promising use case for visual cryptography.

The paper is organized as follows. Section 2 presents background on visual cryptography and augmented reality. In Section 3, we present our methodology for implementing visual cryptography using augmented reality displays (ARDs). We describe our user studies in Section 4 and present the results in Section 5. Finally, we present the limitations of our work in Section 6 and the conclusion and future work in Section 7.

2. BACKGROUND

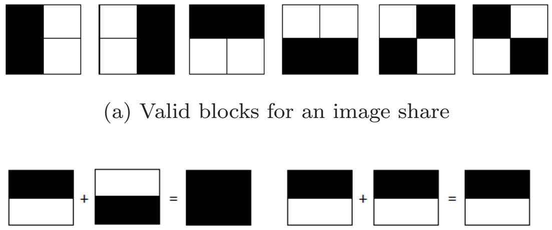
In this section we give a brief introduction to visual cryptography (Section 2.1) and augmented reality (Section 2.2). We then discuss an approach to combine these technologies that enables message recovery without trusting any individual device with the message (Section 2.3) and that forms the basis of our implementation (which is further described in Section 3).

2.1 Visual Cryptography

Visual cryptography [26] is a cryptographic secret-sharing scheme where visual information is split into multiple shares, such that one share by itself is indiscernible from random noise. (Equivalently, one of these shares constitutes a one-time pad as discussed in Section 1, and the other represents the ciphertext of the secret message.) The human visual system performs a logical OR of the shares to decode the secret message being shared.

In visual cryptography, a message share is represented by a rectangular grid of *blocks*, each block itself being a square 2×2 grid of *regions*. Each block in a message share includes two black regions and two transparent regions, as shown in Figure 1a. The ciphertext image is decoded by overlaying two image shares on top of each other. Consider the block in Figure 1b. If one share has a block with the two top regions black and the two bottom regions transparent and the second share has the reverse—two top regions transparent and two bottom regions black—then when the two shares are overlaid the resulting block will appear solid black. This is the decoded image. Recovering a white block is done in a similar manner, as shown in Figure 1c. Note that the recovered "white" block is partially black and partially transparent. This "gray" block is used to represent white.

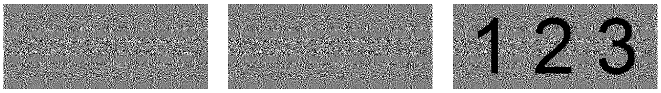
Consider Figure 2, in which the secret message is decomposed into two shares shown in Figures 2a–2b, such that on stacking the two, the user sees Figure 2c. Even though the contrast of the image has been degraded by 50% from regular black-on-white text due to the gray blocks, the human visual system can still identify the content of the secret message ("123") easily.



(a) Valid blocks for an image share

(b) Recovering a black block (c) Recovering a white block

Figure 1: Visual cryptography block representations



(a) First share (b) Second share (c) Overlaid shares

Figure 2: Example of visual cryptography

When the human visual system does the decryption, no computation is needed; it is a visual OR operation. That is, if a given region is black in either of the overlaid shares, the user sees a black region. Otherwise the user sees a white region. Even though the human eye performs a logical OR operation per region, the result is an XOR operation per block. Hence, the security of the one-time pad is retained. In particular, an individual share reveals no information about whether a specific decoded plaintext block is black or white. A disadvantage of the process is the loss in contrast of the decrypted message, which in turn affects the clarity of the recovered secret.

While an adversary in possession of only one share learns no information about the corresponding secret message, he might still compromise the integrity of the secret message if he is able to modify one of the shares in a meaningful way. A method to defend against this attack, introduced by Naor and Pinkas [25], is to use part of the secret message to convey the desired information and deliberately leave part of the secret message as a solid "color," i.e., every block consisting of two black regions and two transparent regions ("white") or every block consisting of four black regions ("black"). If the adversary modifies one of the blocks in one share by swapping the white and black regions, the result will be to change the color of the block in the decoded plaintext image. If the changed block happens to be in the dedicated non-content region, the recipient will know that one of the shares was compromised. If half of the secret message is a dedicated non-content region then the adversary has only a 50% chance of successfully modifying a share without noticeably modifying the non-content region of the decoded plaintext. One of the questions we examine in our user study is whether a user will always notice a single block of the wrong color in a non-content region of the decoded plaintext image.

The question of how to convey a message using VC such that any modification to a share by an attacker will be detectable has been well studied [6, 31, 9, 15]. These proposals all make some assumptions about the capabilities of the human visual system (HVS), which we attempt to validate in our user studies. The questions we explore are: whether a human will detect a block in a message share that does not have exactly two black regions; whether a user will notice blocks in the decoded message that have either one or three black regions; whether a user will notice blocks of the wrong color in a non-content region (as described in the previous paragraph); and whether a user will notice blocks that subtly change the semantic meaning of the decoded message.

2.2 Augmented Reality

Augmented reality (AR) systems supplement reality by combining a real scene viewed by the user and a virtual scene generated by the computer [24, 5]. The AR system superimposes a user's environment with computer-generated information, making the user's environment digitally manipulable and interactive.

The most common realization of AR is with see-through head-mounted displays (HMDs). An HMD is a device that is typically attached in close proximity to the eyes with a display system that couples a digital image with the environment [11]. HMDs enable the merging of virtual views with a user's physical environment, for example by enabling a physician to see the 3D rendering of CT images of a patient superimposed onto the patient's abdomen [4]. We refer to these augmented reality HMDs as augmented reality displays (ARDs) for brevity. ARDs have started to ship commercially, with the latest examples being Microsoft's HoloLens [23] and Google Glass [13].

The human visual system forms a core element of these near-eye ARDs, especially when considering the performance and usability of the system. The field of view (FOV) of the human eye describes the dimensions that can be seen by the eye at any given instant of time. For the human eye, this field of view is approximately 150° horizontally and 120° vertically. The binocular overlap region, within which a target is visible to both eyes, measures about 114° [11].

Many commercially available binocular ARDs like the Lumus DK-32, the Optinvent ORA-S, and the Epson Moverio BT-200/100 are limited to a FOV of less than 60°. Maimone et al. [21] developed a near-vision see-through ARD with a field of view of approximately 110° diagonally, and so far this is the maximum FOV that has been achieved with AR displays. Google Glass has a 14° monocular FOV and the Epson Moverio has an FOV of 23° [28]. The major limitation resulting from a small field of view is the inability to show finer details on ARDs.

There are several challenges in overlaying information onto a user's FOV, such as contrast sensitivity and color perception. The color perceived by users wearing an ARD is affected by the transparency of the graphics being rendered. In addition, some displays significantly diminish the brightness of the real world so that the graphics on the device's display need not be bright [11, 20]. These issues may result in changes in contrast and hues perceived by the user. For example, since black is rendered as transparent to allow the real world to be seen, dark colors may be perceived improperly, and bright colors that are intended to occlude the real-world background may not be able to do so [20].

Another challenge in head-mounted ARDs is head jitter. As a user tries to focus on an object, slight head movements translate to large movements of the displayed object. This is particularly important for our scenario, as the users are trying to align two images. However, these problems can be addressed using built-in image stabilizers, motion sensors and head tracking [8].

2.3 Using Devices to Implement VC

While our work is, to our knowledge, the first to explore using augmented reality head-mounted displays to implement VC, other works have suggested the use of specialized devices to replace the transparencies envisioned by the original VC authors. In particular, Tuyls et al. [32] suggest a transparent display device that the user holds or attaches to her computer display. To reveal a plaintext to the on looking user, the transparent display device renders one image share, and the underlying computer display renders the other. Neither the transparent display device nor the underlying computer display individually gains any information about the plaintext image that the user sees.

While the design we test replaces the transparent display device with an augmented reality display (ARD), otherwise it is conceptually similar and borrows underlying elements from the Tuyls et al. design. In particular, a sender generates and securely transmits image shares in our envisioned deployment in the same way (subject to some ARD-specific constraints, discussed in Section 3), and the receiving user can read the message without trusting either the ARD or the computer display with the contents of the message. Additionally, Tuyls et al. demonstrate how the user may respond without divulging her response to her devices.

We stress, however, that our contribution is not in the conceptual novelty of our design of a VC system using ARDs, but rather in a study of the usability of our approach for doing so. To our knowledge, we are the first to undertake such a usability study.

3. IMPLEMENTATION OF VC USING AUGMENTED REALITY

As originally envisioned, visual cryptography used printed transparencies consisting of black (opaque) and transparent regions [25, 2]. However, ARDs render images on a glass prism via light projection. Black regions are represented by the absence of illumination and are transparent, while formerly transparent regions are now rendered through the projection of white light [11, 20]. Thus, the notion of transparency and opacity in visual cryptography achieved through ARDs is inverted with respect to traditional visual cryptography using printed transparencies.

In our implementation, each share of the secret message is composed of 2×2 blocks. Each block in a share includes two illuminated (white) regions and two transparent (black) regions. When two images are overlaid, the regions combine as in Figure 3. In our initial experiments, we tried using colors other than white to render the opaque regions, but we found white was the most effective.

One share is sent to the user's ARD, and the other is sent to a display in the vicinity of the user. To superimpose them, the user views the visual share through the ARD displaying the second visual share. The two shares thus overlaid reveal the secret message. Below we describe how we designed our system to work with Google Glass and the Epson Moverio.

3.1 Google Glass

Google Glass is an Android based augmented reality device that is Wi-Fi and Bluetooth enabled. It provides a

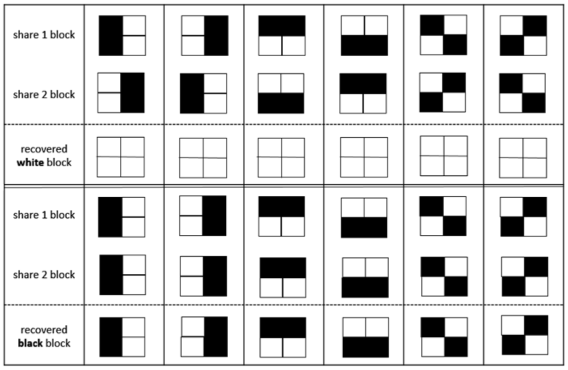


Figure 3: Construction of visual cryptography scheme for augmented reality. The secret block is obtained by superimposing the two corresponding shares, looking through an augmented-reality head-mounted display.

monocular AR experience with a small display screen for the right eye. Glass enables users to view content including text, images, and videos. It applies transformations on images that are displayed on it such that they are scaled in both the horizontal and vertical directions. To counterbalance that transformation, we scale the images to 500×600 pixels and pad them with a black background, such that our final images are 1200×1200 pixels on a desktop monitor. The display of Google Glass with a resolution of 640×360 pixels (equivalent of a 25 in (64 cm) screen from 8 ft (2.4 m) away) [13] is very small, with a field of view of 14°. We found the maximum image size for the secret message that could comfortably be displayed on Google Glass was a grid of 5×5 blocks, with each block containing two black and two white regions.

In addition to the above transformations, we added alignment markers (red borders around blocks, as shown in Figure 4) to help users align the two images. We also blacked out the areas that contained no useful information to further aid image alignment and account for head jitter.

3.2 Epson Moverio

Our second study used an Epson Moverio BT-200 Model H423A—an Android based augmented reality device, with display screens for both eyes, i.e., binocular display. It has a resolution of 960×540 pixels (equivalent to 80 inch image from 16.4 feet away) [28, 10]. The Epson Moverio has a field of view of 24° and enables 2D and 3D viewing. The 2D mode is similar to the viewing mode in Google Glass and it is the mode we used in our study. We used two image sizes containing 7×7 blocks and 9×5 blocks, respectively. These sizes were selected based on trial and error for what was a comfortable and easy image size for alignment to be done by untrained, first-time participants. The investigator for the study, because of training effects, was able to pack up to two letters or four numbers into an image and still able to align and identify the characters in the images.

Even though the information being conveyed is limited, there is a marked improvement over using Google Glass. We believe this is mostly because of the Epson's larger field of view. With a larger field of view and higher resolution more information can be conveyed through one image.

One limitation of the two headsets used, or some of the others that are commercially available, such as the Optinvent ORA-S, Recon Jet, Vuzix M100, and Lumus, is absence of any form of built-in head tracking and image stabilization. This was one of the major challenges that we faced during the Google Glass study: even slight head movement misaligned the images. We elaborate more on this as we describe the user studies.

3.3 Threats introduced by ARDs

As discussed in Section 1, our goal is to implement VC in such a way that realistic attackers can access (to read or modify) only one of the two visual shares needed to reveal the secret message. While our focus in this paper is on the usability of ARDs for this purpose, we pause to consider challenges in ensuring this property with the two ARDs that we have used in our implementations.

One type of attacker that poses challenges with either of our ARDs is a third party in proximity of the user who might photograph both the physical share (i.e., the share in the user's physical proximity) and, with high definition, the share displayed in the ARD. Since this attacker would not see these two images aligned as the intended message recipient does, he would then need to reconstruct the secret message offline using optical decoding techniques or manual effort. Such attacks, however, are not far-fetched: the possibility of observing the image displayed in Google Glass is well-known (e.g., "bystanders who see the tiny display screen above the right eye" [19] and "the screen is just visible from the outside" [12]) and similarly intricate optical decodings have been demonstrated (e.g., [18]). That said, this threat is equally present—if not more so—in traditionally envisioned deployments of VC that use, e.g., physical transparencies to reveal a secret message. Using ARDs, it is presumably easier for the user to ensure that her ARD is not being observed (e.g., by shading it with her hands) during message recovery.

A related concern is that several ARDs (e.g., Google Glass, HoloLens, Epson Moverio BT-200, and Vuzix M100) have a front-facing camera. (Several others, like the Epson Moverio BT-100 and Laster See-Thru, do not.) As such, if an ARD with a front-facing camera is compromised by malware, it could potentially use the front-facing camera to photograph the other share and combine it with the share given to the ARD to display, revealing the private message. It is therefore necessary that, for an ARD with a front-facing camera such as Glass, the camera lens be covered physically while the physical share is displayed. We did not incorporate this step into our user study but would recommend doing so in actual deployment. Similarly, if the physical share is displayed on a computer display, that computer should not have a camera facing the user that might capture the share displayed in the user's ARD (or else that camera should be physically covered).

Finally, ARDs like Google Glass are not really standalone devices; rather, they share their information with the service provider to which they are tethered—Google in the case of Glass. There is a risk that this service provider could both extract image shares from the ARD and combine them with the corresponding other image shares that it receives otherwise (e.g., sent to the user's gmail account). Addressing this risk presumably involves the sender conveying ciphertexts to the user via a different service provider than the one with which the ARD shares its contents, or to adopt an ARD that does not so freely share its contents with a service provider.

4. USER STUDIES

We conducted three user studies: a pilot user study using Google Glass, an improved formal study using the Epson Moverio ARD, and an online user study using Mechanical Turk. All our user studies were reviewed and approved by our university's IRB. In the pilot user study we gauged participants' ability to align visual shares and discern the overlaid information from noise. In the Moverio user study, participants decoded letters and numbers, and we also investigated their ability to recognize modifications to the decoded plaintexts. In the Mechanical Turk study we assessed users' ability to detect modifications to individual image shares. The pilot user study and the formal user study were conducted over a period of two weeks each. The formal user study was conducted three months after the pilot user study.

4.1 Participant Demographics

The pilot study had 31 participants: 22 male and 9 female. Out of the 31 participants, 19 were aged between 18-25, 9 between 26-35 and 3 between 36-45. Of the 31 participants, 8 wore prescription glasses. In the formal study, there were 30 participants, of which 23 were male and 7 were female. The age groups included 17 participants between 18-26, 11 between 26-35, and 2 between 36-45. Of the 30 participants, 15 wore prescription glasses. Both participant groups consisted mostly of university students, staff, and faculty members. Recruitment was done through emails to department and student group mailing lists. We recruited participants this way because participants had to come to our office location (on our university campus) for participating in the study.

Given the nature of these studies, we believe the most important bias in our participant pools is that toward younger participants. While we do not claim that these participant groups enable us to generalize our results to the general population, we are hopeful that they should be representative of populations represented by these age groups. Results would vary for other population groups and can be explored as part of future work.

For the Mechanical Turk study, we had 50 participants, with 28 male and 22 female participants. The age groups were 14 between 18-25, 20 between 26-35, 12 between 36-45, and 4 between 46-55.

4.2 Pilot User Study

4.2.1 Training

To acclimate users to the experience of using Google Glass, each user first underwent a brief training session on how to perform image navigation and alignment on the device. The training set comprised two stages: First, three photos not related to the study were displayed to demonstrate the act of navigation between pictures in Google Glass' image browser. Then, three images containing blocks (see Figure 4) were presented to the user as practice for the actual task. The users were asked to align the blocks on Glass and the monitor screen and report which blocks were white. During this training phase, the participants were provided guidance if they responded incorrectly.

In the training phase and in task 1 below, each recovered message was a Braille character. We used Braille characters because they fit well in a 5×3 block grid we used with Glass. A Braille character is a three-row, two-column grid with raised "dots" in some cells that can be felt. The locations of these dots in the grid indicate a character (e.g., see [1]). We used a white block to correspond to a raised dot and a black block to correspond to the absence of a raised dot. Our choice of Braille for this pilot study was primarily due to the simplicity of this mapping of Braille characters to VC, and of course not because we intend for blind persons (the primary users of Braille) to make use of VC or AR. The participants were not told that the recovered messages should be Braille characters and no knowledge of Braille was required.

4.2.2 Task Descriptions

After training, the participants were given three tasks. In each task, each participant was given a set of image pairs (one in Glass and one on the computer monitor) and asked to overlay and align the images on Glass and the monitor to figure out which blocks were completely white. The participant then had to note this down on a sheet of paper before moving on to the next pair of images in the set. A 3×2 grid was drawn on the paper for each image pair, and the participants marked the observed white blocks on the grid. Each participant filled out a questionnaire after completing each of the three tasks and then a final questionnaire after all tasks were completed.

Each participant was timed for each image pair presented. The timed duration included the time taken to navigate to the image on Google Glass as well as the monitor, align the two image shares, identify the white blocks, and note their locations on a sheet of paper.

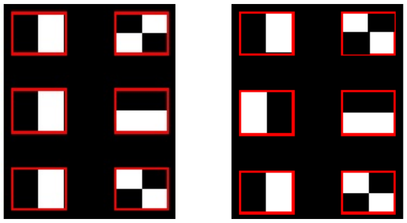

(a) First share (b) Second share

Figure 4: Image shares for Braille character "i"

Task 1: Task 1 had five image pairs; an example image pair is shown in Figure 4. The image pairs in this task were shares of randomly chosen English Braille characters. All the participants were presented a different set of image pairs. The participants had to write down which blocks in the recovered plaintext were white.

Task 2: This task was similar to the previous task but the images were shares of a 3×3 grid of blocks with at most one white block, e.g., see Figure 5. The location of the white block (if any) was chosen uniformly at random. There were five image pairs in this task, and the participants were asked to write down which block in each recovered plaintext (if any) was white by writing the white block's number, with the upper left corner being block one, the lower-right corner being block nine, and numbers incrementing along rows (like a telephone keypad). If there was no white block, the

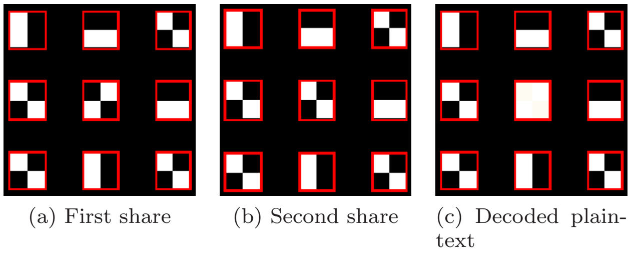

(a) First share (b) Second share (c) Decoded plaintext

Figure 5: Image shares for the number 5 in pilot study (Section 4.2)

participant was instructed to write a zero. We chose this task design because the majority of the population is familiar with this number-pad pattern, and given the limitations of using VC with Google Glass in its current form, this was a viable way to convey numbers.

Task 3: The images used in this task had the same layout as in task 2. Each participant was given three pairs of images and asked to indicate which block (if any) of the recovered plaintext was white. However, unlike in task 2, all three visual shares were displayed side-by-side at the same time on the monitor. The participants still had to navigate the image shares on Glass, however, and overlay the shares on Glass with the shares on the monitor from left to right until completing the task. We chose this task design to test the ease of alignment when more than one physical share was displayed on the monitor. For this task all the participants were given shares for the same three numbers, which (for no particular reason) was the room number of the room where the study was conducted.

4.2.3 Results

Figure 6 presents the distribution of average plaintext recovery time per participant, i.e., averaged over all image pairs in the task indicated on the horizontal axis. A timer running in the background captured the time spent per image. As soon as a participant navigated to the next image, the timer for that image started.

Each box in this plot consists of three horizontal lines, which indicate the 75th, 50th (median), and 25th percentiles of the average time per participant. Whiskers extend to cover the lowest (highest) point within 1.5× the interquartile range of the lower (upper) quartile. There was no significant decrease in time spent per image pair as a participant progressed through an individual task.

There is noticeable variation in the amount of time spent by users per image pair, ranging from a few seconds to as large as 40 seconds. For identifying the white blocks in the recovered plaintext, the users not only had to attune themselves to what they are looking for but also use a new technology—all of the users were first-time users of Google Glass.

There is no relationship between error rates and timing data. Participants who identified the decoded plaintext incorrectly may or may not have spent little or more time on those visual shares. Participants who had a higher error rate did not on average take longer than participants who had no errors. A total of 9 participants out of 31 decoded at least one plaintext incorrectly. Table 1 shows the number of participants that made the number of errors indicated in the leftmost column. A total of 403 image pairs were presented to the participants, out of which 16 occurrences were incorrectly identified. One interesting observation is that if the participant made no error in the first task, then they made no errors in the subsequent tasks.

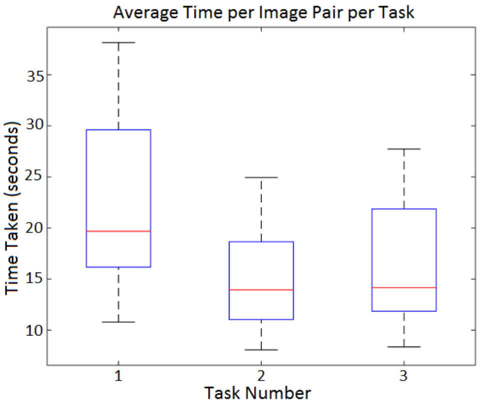


Figure 6: Box plot showing average plaintext recovery time per participant, i.e., averaged over the number of plaintext recoveries in the task on the horizontal axis, in the pilot study (Section 4.2).

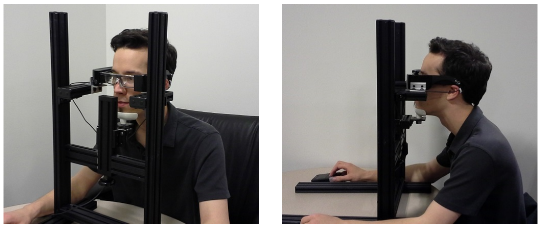
	Task 1	Task 2	Task 3
All correct	22	28	27
1 incorrect	6	3	4
2 incorrect	1	0	0
3 incorrect	0	0	N/A
4 incorrect	2	0	N/A
5 incorrect	0	0	N/A

Table 1: Number of participants per error number in pilot study (Section 4.2)

It was harder for participants wearing prescription glasses to clearly see images on Glass and align them, as indicated by them on the questionnaires. Some participants reported discomfort using Google Glass as they were left-eye dominant. We moved to the Epson Moverio in the second study to address these issues. Many participants reported that the alignment was hard because their heads would not stay completely still and they were thus unable to figure out if a particular block of the recovered plaintext was white or not. As there is no form of head tracking or image stabilization in Google Glass, this was a problem. We modified the design of the second user study based on this information. Some users also reported that they took longer on images because they had to recall which blocks were white while marking their answers down on the sheet. Based on this, we changed the design of the next study so that the users did not have to write down what they see.

4.3 Formal User Study

For the second user study we used an Epson Moverio BT-100, an Android-based ARD [10]. Unlike Google Glass, the Moverio presents content for both eyes, i.e., it is has a binocular display. However, we found the alignment to be tedious in a binocular setting; therefore, we used the Moverio only



(a) Front view

(b) Side view

Figure 7: Chin rest setup for the formal user study.

as a monocular headset, but allow the users to select either the left- or the right-eye display.

Similar to Google Glass, the Epson Moverio does not have image stabilization or head tracking capabilities to account for head jitter. To compensate for this, we provided a chin rest to the participants (see Figure 7). The chin rest stand also situated the ARD on a small platform for added support.

One visual share was displayed on the Epson Moverio and the other was displayed on a monitor. To enable image alignment, participants could move and scale the visual share displayed on the monitor. An Xbox controller was used for this purpose. The participants reported their observations (what they were asked to report varied based on the task) and the investigator noted them. At the end of the study, participants filled out a questionnaire. The questionnaire aimed to capture their confidence levels as the study progressed, as well as participant demographics.

4.3.1 Training

As in the pilot user study, the participants underwent a short training. Participants were given examples of the images they would be viewing and descriptions of the tasks. Through a presentation and practice session on the setup, participants were familiarized with the tasks. The decoded plaintext consisted of individual letters and numbers that resemble a 14-segment LED display font, as in Figure 8. After the training, participants took a 1-2 minute break. Each task was described again as participants started the tasks.

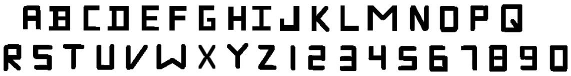
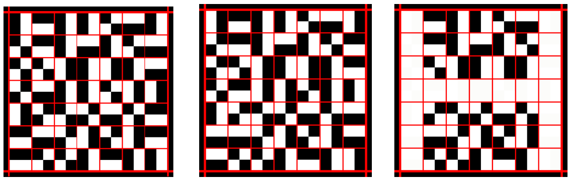


Figure 8: 14-segment LED font used for displaying letters and numbers in the study described in Section 4.3

4.3.2 Task Descriptions

There were three tasks in the study. Each task involved a set of ten image pairs. In each of the tasks, participants aligned the image pairs (visual shares) and identified the decoded plaintext character. They also reported other characteristics of the decoded plaintext for tasks 2 and 3. The participants were timed for each image pair presented; this included the time taken to navigate to the next image pair, identify the characteristics of the decoded plaintext requested in each task and report their observations. A participant had to align only the first image pair of each task; the rest of the image pairs would retain that alignment, unless the user shifted her position.

Task 1: In task 1, users simply identified the decoded characters. Each participant was given ten image pairs that encoded random characters—five numbers and five letters. Upon overlaying the two visual shares on the Epson's display and the monitor, the letter or number was revealed, as in Figure 9.



(a) First share

(b) Second share

(c) Decoded plaintext

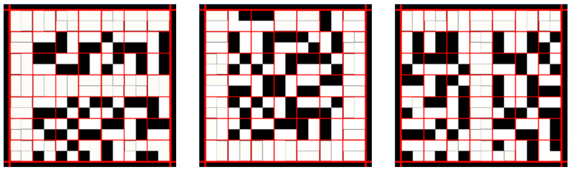
Figure 9: Image shares for letter 'H' (Section 4.3)

Task 2: In the second task, we first reiterated to the participant what constitutes a *legal* decoded plaintext image, namely one in which the blocks that form a character have all regions white (illuminated) and all other blocks have exactly two black and two white regions. All participants were given ten image pairs, where the decoded plaintext of each pair was a letter. For each image pair, we asked participants to identify the letter in the decoded plaintext, identify whether the plaintext is legal, and if not, to count the number of blocks that made it illegal—i.e., blocks that are part of the letter but not wholly white, or blocks not part of the letter that are not half-black and half-white. Here we call such blocks *nonconforming*. All decoded plaintexts in task 2 had between one and six nonconforming blocks, though we did not inform the participants that all decoded plaintexts in the task were illegal.

All participants were given the same image pairs in this task, and the plaintexts decoded from these image pairs were chosen to include "confusing" letters. For example, nonconforming blocks at particular locations can cause confusion especially between 'F' & 'E', 'O' & 'U', and 'T' & 'I'. Six of the ten image pairs provided to participants were intentionally chosen to yield one of these letters, and nonconforming blocks were positioned in their decoded plaintexts so as to maximize confusion. Examples of decoded plaintexts for these three letter pairs are shown in Figure 10.

Task 3: In task 3, each decoded plaintext had a "content" and "non-content" part, as suggested by Naor and Pinkas [25] to increase the likelihood of detecting adversarial manipulation of an image share (see Section 2.1). The content part carries the message, in this task a number. The non-content part contains no meaningful information and is made up of blocks of all the same "color", i.e., blocks that all have exactly two black and white regions (i.e., "black") or blocks that all have four white regions (i.e., "white"). For the purposes of our study, in a *legal* decoded plaintext, each block of the non-content part has exactly two black and two white regions. Thus, an illegal plaintext is simply one containing white blocks in its non-content part. These are the *nonconforming* blocks as shown in Figure 11.

The content part of a plaintext could be either on the left or the right of the plaintext image. There was a shift from

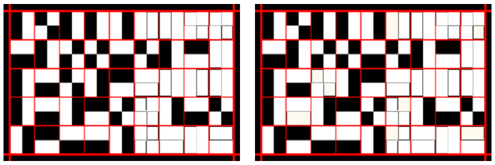


(a) 'F' with four nonconforming blocks in bottom row, or 'E' with six nonconforming blocks in bottom row

(b) 'O' with three nonconforming blocks in top row, or 'U' with three nonconforming blocks in top row

(c) 'T' with five nonconforming blocks in bottom row, or 'I' with six nonconforming blocks in bottom row

Figure 10: Decoded plaintexts for three image pairs with nonconforming blocks to maximize confusion between (a) 'F' & 'E'; (b) 'O' & 'U'; or (c) 'T' & 'I'; used in task 2 of study described in Section 4.3



(a) Legal decoded plaintext

(b) Illegal decoded plaintext

Figure 11: Decoded plaintexts obtained upon overlaying visual shares in task 3.The content part is on the right side and the non-content is on the left. (a) shows a *legal* image with non-content region containing only blocks with two black and two white regions; (b) shows an *illegal* image with malformed white-blocks in the non-content region.

right to left once during the task. The participant was told that a switch would be signaled by an all-white non-content region (a few nonconforming blocks notwithstanding); i.e., this was a signal that the next decoded plaintext would have its non-content region on the other side from its present location. If the non-content region was all-black (again, aside from a few nonconforming blocks), then the non-content region would be on the same side in the next decoded plaintext. The participants were asked to report for each decoded plaintext whether the content was on the left or the right, report the alphanumeric character they decoded, and report if the non-content part was legal and, if not, the number of nonconforming blocks it contained.

In doing so, the task evaluated users' ability to detect nonconforming blocks in a non-content area. The number of nonconforming blocks in this task ranged from 0 to 10. We also observed the minimum number of nonconforming blocks needed in the non-content part of a decoded plaintext in order for users to be able to discern the plaintext as illegal.

4.4 Mechanical Turk Study

An adversary in possession of an image share might attempt to modify blocks in the image share to change the meaning of the decoded plaintext. For example, by flipping the black and white regions in the image-share block, the adversary changes whether the corresponding block in the decoded plaintext is black or white. However, to predict the resulting block color in the decoded plaintext, he must know the color of that block in the original decoded plaintext, i.e., prior to flipping the black and white regions. If he does not know the color of the original decoded plaintext block, he can nevertheless increase the chances that it is turned to white by modifying the block in his image share to have three (or four) white regions, instead of only two. In prior implementations of VC, detecting such malformed blocks may go undetected in an individual share because of the small sizes of the blocks. However, in our implementation, detecting these malformed blocks in an image share (i.e., blocks that do not have exactly two black regions and two white regions) is feasible and can be an an important step to detecting an adversary's manipulation of that share.

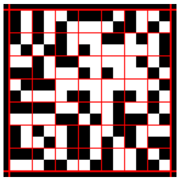


Figure 12: Image share with nine malformed block shares (Section 4.4)

We conducted an online user study on Amazon Mechanical Turk to evaluate users' ability to do so. Each participant was given the same 20 image shares in a random order and was asked to identify whether each contains any malformed blocks and, if so, how many. Note that each image was an image share, and so there was no other meaningful information in the image share. An example is shown in Figure 12. This image share has nine malformed blocks that are all next to each other in a group. The malformed blocks could also be positioned randomly in the image share.

One of the goals of the study was to observe the differences in identifying randomly positioned malformed blocks and grouped malformed blocks. Hence, the set of images given to the users was a mix of both: there were eight image shares with randomly positioned malformed blocks, eight image shares with grouped malformed blocks, one image share with one malformed block, and three image shares with no malformed blocks. We also explored the least number of malformed blocks (both grouped and random) that enabled participants to reliably discern that an image share is illegal. The presence of grouped malformed blocks is a characteristic of the attacks described by Chen et al. [6].

At the end of the study the participants filled out a questionnaire to capture demographics and their confidence as the study progressed.

5. RESULTS

This section presents the results and analysis from the formal user study (Section 5.1) and the Mechanical Turk user study (Section 5.2).

5.1 Formal User Study Results

5.1.1 Timing data

Figure 13 shows the average time for decoding plaintext taken by the users in each of the tasks. Of particular interest is the first task since it reflects the amount of time taken to identify the letters or numbers. For tasks 2 and 3, the

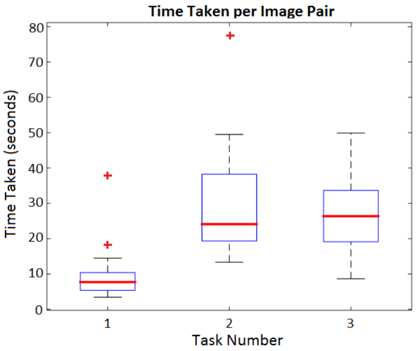


Figure 13: Boxplot showing the average time for plaintext decoding (excluding the first) per participant, in each of the three tasks as described in Section 4.3

Relation between	Correlation coefficient	p-value
Tasks 1 and 2	0.7285	5.0144×10^{-06}
Tasks 2 and 3	0.6679	5.50604×10^{-05}
Tasks 1 and 3	0.3867	0.0348

Table 2: Correlation between task times based on Pearson correlation in user study described in Section 4.3

reported times include the time taken to count the number of nonconforming blocks. The actual time to identify the characters and simply the presence of nonconforming blocks in the decoded plaintext is lower. Figure 13 does not take into consideration the time taken for the first image in each task. For the first image in each task, users spent a considerable amount of time initially aligning the image shares, yielding an outlier that dramatically skews the averages per task. The time taken to initially align the image shares ranged from 18.49 to 313.32 seconds in the first task; 11.94 to 303.16 seconds in the second task; and 27.4 to 280.79 seconds in the third task.

There is a correlation between the time taken in the different tasks on a per-user basis. Table 2 shows the Pearson correlation coefficients and corresponding p-values for each of the timing relations. Participants who took longer on either task 1 or 2 were likely to spend more time in task 2 or 3. A stronger correlation was observed between timing in consecutive tasks, i.e., between tasks 1 and 2 and between tasks 2 and 3. We also found a correlation between time taken in task 3 and image clarity (see Section 5.3.1).

5.1.2 Error Rates

In each task the participants reported the characters they observed. In task 1, 26 out of 30 participants identified all images correctly, and the remaining 4 participants identified 9 out of 10 images correctly. In task 2, 28 participants identified the letters in all the images correctly, one participant identified 9 images correctly, and one participant identified 8 images correctly. In task 3, 27 participants identified the number in the images correctly, two participants identified 8 images correctly, and one participant identified 9 images correctly. This indicates that users were able to clearly identify the letter or number being conveyed. The participants who made mistakes identifying characters in task 2 had also made errors in task 1. Two out of the three participants who made errors in identifying the numbers in task 3 also decoded the plaintexts incorrectly in tasks 1 and 2.

In tasks 2 and 3, we also evaluated users' ability to recognize *illegal* decoded plaintexts using the definitions of illegal given in Section 4.3.2. In these tasks, the participants were asked to identify whether the decoded plaintexts were illegal and, if so, report the number of observed nonconforming blocks.

Nonconforming blocks represent an attacker's active modifications observable in the decoded plaintext (see Section 2.1). In task 2, all decoded plaintexts were illegal, and in task 3, there were seven illegal plaintexts out of ten. Figure 14 shows the percentage of participant responses identifying a decoded plaintext as illegal as a function of the number of nonconforming blocks present in the plaintext. For both tasks 2 and 3, as the number of nonconforming blocks increases, more participants indicated that a decoded plaintext was illegal. Participants were able to discern that an image was illegal even when only one nonconforming block was present in the plaintext. In task 3, 97% of user responses correctly identified the *legal* plaintexts.

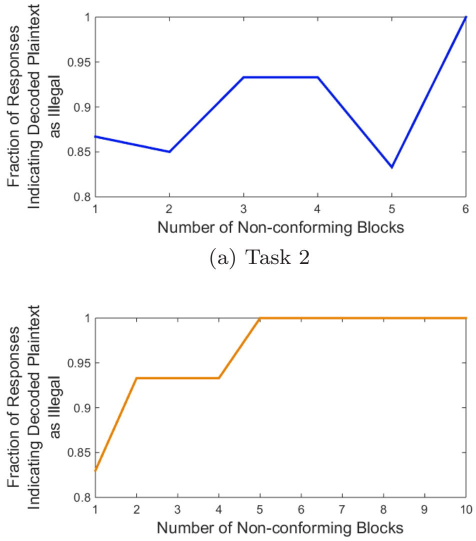
In task 2, participants were given (among others) images with the letters 'F', 'O' and 'T'. As discussed in Section 4.3, we suspected that the presence of nonconforming blocks could easily cause these letters to be confused with other letters. All participants identified these letters or plausible alternatives: specifically, participants identified each 'O' as either 'O' or 'U', each 'T' as 'T', and each 'F' as 'F', 'E', or 'P'. However, in reporting our results, we considered a participant's response as correct only if she also reported the presence of nonconforming blocks (i.e., that the plaintext was illegal).

As seen in Figure 14a, when there were five nonconforming blocks in task 2, there was a decrease in the number of responses that correctly identified a plaintext as illegal. In task 2, the only letter with five nonconforming blocks was the letter 'K'. The nonconforming blocks did not obscure the letter, nor did they cause the 'K' to closely resemble another character. This may have led participants to conclude that the image was legal.

In task 3, participants were also asked to report whether the content was on the left or the right. There was a shift from right to left once during the task, which all 30 users were able to correctly ascertain.

Figure 15 presents the number of nonconforming blocks reported versus the actual number of nonconforming blocks in the decoded plaintext. In Figures 15a–15b, the thicker red line indicates the median value. The bottom and top bars represent the 1st and the 3rd quartiles, respectively. A single line indicates that these quartiles are the same. A median value that is the same as the 1st quartile or the 3rd quartile indicates a skew in the responses.

The plots indicate that the participants erred toward reporting fewer nonconforming blocks than actually were in the plaintexts, and as the number of nonconforming blocks increased, the reported nonconforming blocks also increased. It is important to note that despite not being able to identify all the nonconforming blocks, 89.33% of decoded plaintexts in task 2 and 96% in task 3 were correctly identified as illegal by the participants. In a real-world scenario, this ability to distinguish between illegal and legal plaintexts can be considered more important than identifying the exact number of nonconforming blocks.



(a) Task 2

(b) Task 3

Figure 14: Fraction of responses indicating plaintexts as legal or illegal based on the number of nonconforming blocks present in plaintext (Section 4.3)

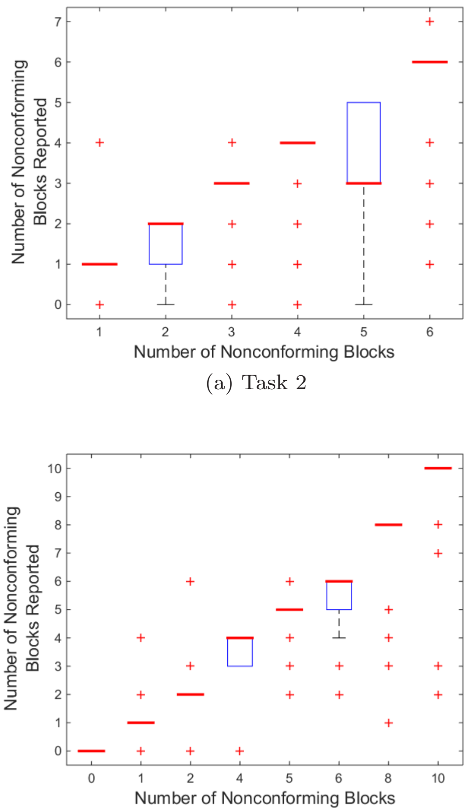
The number of participants who were able to correctly identify nonconforming blocks in task 3 is more than that in task 2 as indicated by Figure 16. This is not surprising: In task 3, participants were asked to look for white blocks in a region expected to have all black blocks (or vice versa), while in task 2, the nonconforming blocks could potentially be part of the character. This highlights the importance of using a non-content region as an aid for detecting potential modifications made by an adversary.

Based on the observations from our user study, use of visual cryptography with the aid of augmented reality displays seems promising, despite the challenges that need to be overcome. Our participants were able to discern the characters being conveyed to them, and they were also able to recognize active modification of the decoded plaintexts. It appears the assumptions of the visual cryptographic literature about the capabilities of the human visual system hold true in an augmented reality setting.

5.2 Mechanical Turk Results

In this study, participants were asked to detect malformed blocks in an image share, i.e., blocks that do not have exactly two black regions and two white regions. All 50 participants were able to correctly identify legal image shares, i.e., with no malformed blocks. However, none of the participants were able to detect illegal image shares containing only one malformed block, except for one user who reported that there were four malformed blocks in that image share. All participants correctly identified the remaining 25 illegal image shares as illegal. There was one participant who reported a large number of malformed blocks for most image shares; perhaps the participant did not clearly understand the instructions and was counting the number of regions within malformed blocks, as opposed to the number of malformed blocks.

All the participants correctly identified all image shares with grouped malformed blocks as illegal. However, not all participants detected that an image share was illegal for the case of randomly positioned malformed blocks, as indicated by Figure 17. In this case, though, as the number of malformed blocks increased, the number of participants who indicated that the image shares were illegal also increased. Mechanical Turk users typically spent less than 45 seconds per image.



(a) Task 2

(b) Task 3

Figure 15: Number of nonconforming blocks reported by participants versus actual number of nonconforming blocks present in plaintext (Section 4.3)

The number of randomly positioned malformed blocks in an image share was 3, 4, 5, 6, 7, 9, 10, or 12. The number of grouped malformed blocks in an image share was 4, 5, 6, 7, 9, 10, 12, or 13. Figure 18 shows the number of participants who reported all the malformed blocks in an image share, for a given number of malformed blocks. It shows that participants were better at detecting malformed blocks that were grouped together in the image share, as opposed to randomly positioned malformed blocks. However, there is a slight decrease in the number of participants who were able to detect all the randomly positioned malformed blocks as the number of malformed blocks increased. Figure 18 reports the number of participants who reported the exact number of malformed blocks. The number of participants who reported 12 ± 2 malformed blocks for the image share with 12 malformed blocks is 47. So even though the participants reported the incorrect number of malformed blocks, they were within a close range.

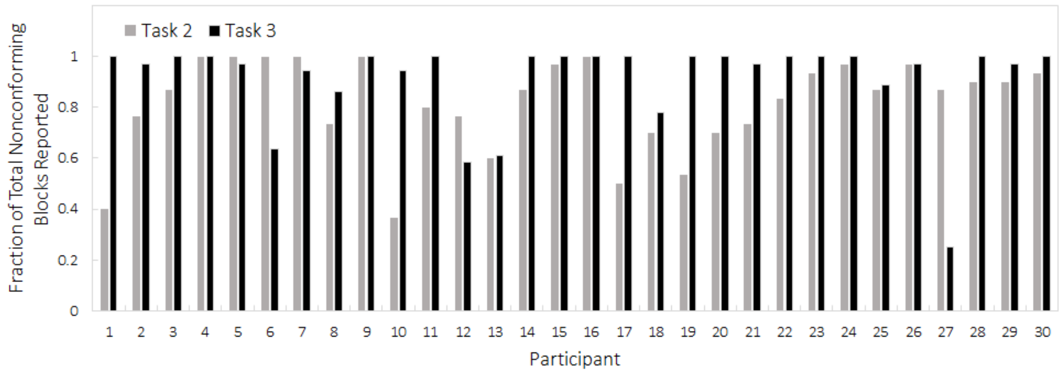


Figure 16: Fraction of nonconforming blocks reported by each participant in task 2 and 3, described in Section 4.3

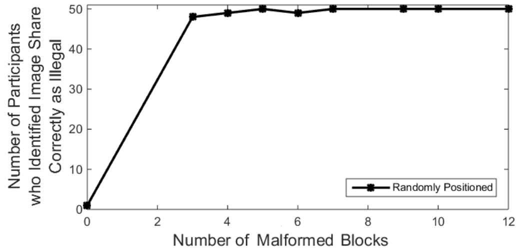


Figure 17: Number of participants who indicated that image shares with randomly positioned malformed blocks were illegal (Section 4.4)

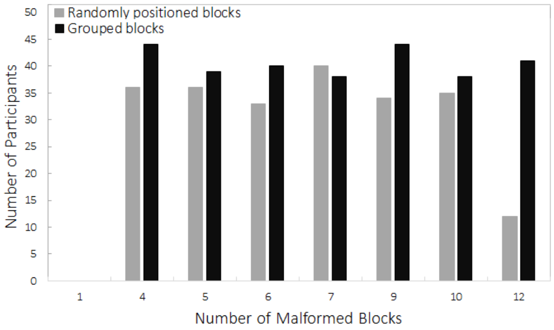
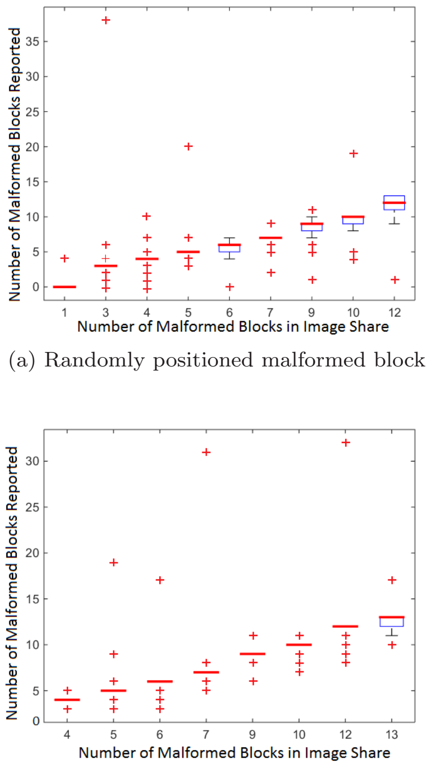


Figure 18: Number of participants who correctly identified all the malformed blocks in an image, whether the malformed blocks were randomly positioned or grouped as described in Section 4.4



(a) Randomly positioned malformed blocks

(b) Grouped malformed blocks

Figure 19: Number of malformed blocks reported versus number of malformed blocks present in the image share (Section 4.4)

Figure 19 shows the number of malformed blocks reported by participants versus the actual number of malformed blocks present in an image share, for both randomly positioned and grouped malformed blocks. For both cases, there was an increase in reported blocks as the actual number of malformed blocks increased. Figure 19 also shows that the participants erred on the side of reporting fewer malformed blocks than were actually present in the image. Again, participants could recognize illegal image shares even if they could not recognize all the malformed blocks. In order to detect an attacker's modification of one share, it would suffice for a user to recognize the image share as illegal, rather than count the exact number of malformed blocks. However, we explored the ability of participants to detect and count different number of malformed blocks, to determine the minimum number of malformed blocks needed to identify an image share that has been modified.

5.3 Questionnaire Responses

The formal user study and the Mechanical Turk survey posed questions to participants about demographics and their perceptions about their own performance during the tasks (see Appendix A for the questionnaires).

5.3.1 *Formal User Study*

We measured reported confidence, character identification in decoded plaintexts, nonconforming block recognition in decoded plaintexts, and image clarity with regard to the decision process. Participants also indicated if they used prescription glasses during the study and, if so, how glasses impacted the ease-of-use of the ARD and image alignment.

Half of our participant population wore prescription glasses during the study. During the pilot study with Google Glass, prescription glasses had proved to be a hindrance for image clarity, alignment, and ease of use. In the formal user study that was not the case. We attribute this improvement to the Epson's form factor, which makes the ARD easier to use, as compared to Google Glass, while wearing glasses. In the formal user study we found no correlation between reported image clarity and use of prescription glasses. Participants wearing prescription glasses performed neither better nor worse in terms of accuracy or timing, compared to other participants.

However, a majority of the users indicated that clarity of the images was a challenge. In the study using Epson, 11 participants indicated that the images were somewhat clear, 2 indicated that they were neither clear nor blurred, 13 indicated that the images were somewhat blurred, 2 participants indicated that the images were blurred, and only 2 participants indicated that the images were clear. These can be explained by the challenges faced by ARDs, some of which are described in Section 2.2.

Participant perception of image clarity did play a role in timing in task 3: We found a correlation between reported image clarity and the total time taken by participants in task 3 (Pearson correlation coefficient of 0.4050, p=0.026). Larger time values corresponded to lower image clarity. However, we found no such significant relationship with the times taken in tasks 1 or 2, leading us to suspect that the participants' responses to the final questionnaire may have been influenced primarily by their performance during the last task. We found no other significant relationships between timing results, accuracy, and other information gathered in the questionnaires.

5.3.2 *Mechanical Turk User Study*

Perceived confidence levels and improvement in detection of illegal images were captured by the questionnaire at the end of the user study. We found no relationships between accuracy, the information gathered in the questionnaires, and demographics.

6. LIMITATIONS

Our participant pool was composed primarily of university students and younger population groups. This could have introduced a bias in our results—our participants were the type that understand basic concepts of augmented reality and secret messaging. Our results may not generalize to the entire population. Future work should examine whether our findings would persist for wider population groups.

We conducted the experiments in a dimly lit room, which is the ideal lighting for perceiving bright white light in the ARD as occluding the background. Strongly lit environments could possibly interfere with the user's recognition of the decoded plaintext. As the augmented reality community overcomes this challenge, lighting should no longer be a limiting factor for the use of VC with ARDs.

Due to hardware limitations, we were unable to do image stabilization for reducing head jitter, and so the formal user study was conducted using a chin rest in order to minimize head jitter. Though it does not completely eliminate jitter, the chin rest seemed to substantially improve the ease of aligning visual shares. As such, we believe that head jitter is a challenge that can be overcome with augmented reality displays (ARDs) that include head tracking and image stabilization [27].

The amount of information conveyed per image is limited. Our efforts to increase the block density of the images were constrained by the small field of view and limited resolution of both of the devices. With a 9° increase in field of view, from Google Glass to the Epson Moverio, we were able to increase the block density and convey individual letters and numbers. As ARDs improve their fields of view, we expect users to be able to decode multiple characters simultaneously, which would improve the bandwidth at which multi-character messages could be decoded. That participants invested a median of roughly 8 seconds to decode and recognize a plaintext character in our formal user study suggests that such advances (and user training) will be necessary for messaging via our approach to become practical for any but the most sensitive messages.

7. CONCLUSION AND FUTURE WORK

Using one-time pads for highly sensitive communications (e.g., in intelligence or diplomatic contexts) has a long history. In this paper we sought to modernize this technique by coupling visual cryptography (VC) with augmented reality displays (ARDs). Specifically, we explored how VC and ARDs can be leveraged to enable a user to receive a secret message without revealing that message to her ARD (or any other device acting in isolation). Our evaluation focused on the ability of users to decode VC-encrypted plaintexts using their ARDs. Through an initial pilot study and subsequent formal study, we demonstrated that users were able to effectively leverage VC via ARDs to decode a single character at a time, provided that head jitter can be addressed—which we did using a chin rest in our second study, but which should be addressable using image stabilization—and that the block density was not too high. Future studies involving ARDs with image stabilization would be useful.

We also measured the ability of users to detect active modification of a VC share by an adversary, both in our formal user study (by detecting nonconforming blocks in decoded plaintexts) and in a Mechanical Turk study (by detecting malformed blocks in image shares). Our studies showed that detecting such active attacks is feasible for most users, though not perfectly. To our knowledge, we are the first to verify the capabilities that are assumed of the human visual system by past work on modification detection in VC.

The limited bandwidth of VC using ARDs unfortunately would appear to limit its use to only the most sensitive contexts. However, there are various other VC schemes (e.g., [2, 3, 7, 14, 33]) that might offer different usability characteristics when used with ARDs. Exploring these possibilities, as well as new devices such as HoloLens and augmented-reality contact lenses, appear to be fruitful directions for new research.

Acknowledgements

We are grateful to Prof. Henry Fuchs for useful discussions and for lending us the Epson Moverio device. We express our gratitude to Jim Mahaney for helping build the chin rest for our studies. We are also grateful to Andrew Maimone, True Price, and Kishore Rathinavel for useful discussions and their assistance with documenting the studies.

8. REFERENCES

[1] American Foundation for the Blind. Braille Alphabet and Numbers. http://braillebug.afb.org/braille_print.asp/.

[2] G. Ateniese, C. Blundo, A. De Santis, and D. R. Stinson. Extended capabilities for visual cryptography. *Theoretical Computer Science*, 250(1):143–161, 2001.

[3] B. Borchert. Segment-based visual cryptography. 2007.

[4] L. D. Brown and H. Hua. Magic lenses for augmented virtual environments. *Computer Graphics and Applications, IEEE*, 26(4):64–73, 2006.

[5] T. P. Caudell and D. W. Mizell. Augmented reality: An application of heads-up display technology to manual manufacturing processes. In *System Sciences, 1992. Proceedings of the 25th Hawaii International Conference on*, volume 2, pages 659–669. IEEE, 1992.

[6] Y.-C. Chen, D.-S. Tsai, and G. Horng. A new authentication based cheating prevention scheme in naor–shamir's visual cryptography. *Journal of Visual Communication and Image Representation*, 23(8):1225–1233, 2012.

[7] S. Cimato, A. De Santis, A. L. Ferrara, and B. Masucci. Ideal contrast visual cryptography schemes with reversing. *Information Processing Letters*, 93(4):199–206, 2005.

[8] P. Daponte, L. De Vito, F. Picariello, and M. Riccio. State of the art and future developments of the augmented reality for measurement applications. *Measurement*, 57:53–70, 2014.

[9] R. De Prisco and A. De Santis. Cheating immune threshold visual secret sharing. *The Computer Journal*, 53(9):1485–1496, 2010.

[10] Epson. Epson Moverio BT 100. http://www.epson.com/cgi-bin/Store/jsp/Product.do?sku=V11H423020.

[11] B. Furht. *Handbook of augmented reality*, volume 71. Springer, 2011.

[12] S. Gibbs. Google Glass review: useful - but overpriced and socially awkward. http://www.theguardian.com/technology/2014/dec/03/google-glass-review-curiously-useful-overpriced-socially-awkward, Dec. 2014.

[13] Google. Google Glass. http://support.google.com/glass/answer/3064128?hl=en&ref_topic=3063354.

[14] Y.-C. Hou. Visual cryptography for color images. *Pattern Recognition*, 36(7):1619–1629, 2003.

[15] C.-M. Hu and W.-G. Tzeng. Cheating prevention in visual cryptography. *Image Processing, IEEE Transactions on*, 16(1):36–45, 2007.

[16] D. Kahn. *The Codebreakers: The Story of Secret Writing*. The New American Library, Inc., New York, NY, 1973.

[17] Kaspersky Labs' Global Research & Analysis Team. Equation: The death star of the malware galaxy. http://securelist.com/blog/research/68750/equation-the-death-star-of-malware-galaxy/, Feb. 2015.

[18] B. Laxton, K. Wang, and S. Savage. Reconsidering physical key secrecy: Teleduplication via optical decoding. In *15th ACM Conference on Computer and Communications Security*, Oct. 2008.

[19] M. Liedtke. Review: 1st peek through Google Glass impresses. http://www.scmp.com/lifestyle/technology/article/1295459/review-1st-peek-through-google-glass-impresses, Aug. 2013.

[20] M. A. Livingston, J. L. Gabbard, J. E. Swan II, C. M. Sibley, and J. H. Barrow. Basic perception in head-worn augmented reality displays. In *Human Factors in Augmented Reality Environments*, pages 35–65. Springer, 2013.

[21] A. Maimone, D. Lanman, K. Rathinavel, K. Keller, D. Luebke, and H. Fuchs. Pinlight displays: wide field of view augmented reality eyeglasses using defocused point light sources. In *ACM SIGGRAPH 2014 Emerging Technologies*, page 20. ACM, 2014.

[22] Mandiant. APT1: Exposing one of China's cyber espionage units. http://intelreport.mandiant.com/, 2013.

[23] Microsoft. Microsoft HoloLens. http://www.microsoft.com/microsoft-hololens/en-us.

[24] P. Milgram, H. Takemura, A. Utsumi, and F. Kishino. Augmented reality: A class of displays on the reality-virtuality continuum. In *Photonics for Industrial Applications*, pages 282–292. International Society for Optics and Photonics, 1995.

[25] M. Naor and B. Pinkas. Visual authentication and identification. In *Advances in Cryptology–CRYPTO'97*, pages 322–336. Springer, 1997.

[26] M. Naor and A. Shamir. Visual cryptography. In *Advances in Cryptology–EUROCRYPT'94*, pages 1–12. Springer, 1995.

[27] Nels Dzyre. Ten Forthcoming Augmented Reality and Smart Glasses You Can Buy. http://www.hongkiat.com/blog/augmented-reality-smart-glasses/.

[28] Optivent. Augmented Reality HMD Benchmarks. http://optinvent.com/HUD-HMD-benchmark#benchmarkTable.

[29] M. J. Ranum. One-time-pad (Vernam's cipher) frequently asked questions. http://www.ranum.com/security/computer_security/papers/otp-faq/, 1995.

[30] M. Schwartz. Who can control N.S.A. surveillance? *The New Yorker*, Jan. 2015.

[31] D.-S. Tsai, T.-H. Chen, and G. Horng. A cheating prevention scheme for binary visual cryptography with homogeneous secret images. *Pattern Recognition*, 40(8):2356–2366, 2007.

[32] P. Tuyls, T. Kevenaar, G.-J. Schrijen, T. Staring, and M. van Dijk. Visual crypto displays enabling secure communications. In *Security in Pervasive Computing*, pages 271–284. Springer, 2004.

[33] Z. Zhou, G. R. Arce, and G. Di Crescenzo. Halftone visual cryptography. *Image Processing, IEEE Transactions on*, 15(8):2441–2453, 2006.

APPENDIX

A. USER STUDY QUESTIONNAIRES

The questionnaires presented to the participants for the formal user study using Epson Moverio and the online Mechanical Turk study are presented in this section. In the paper we referred to modifications of blocks as *'nonconforming'* in the formal user study and as *'malformed'* in the online user study. However, in the questionnaires and while explaining the tasks to the participants in the study, we used the term *'illegal'* to refer to the same. The terms *nonconforming* and *malformed* are used in the text to distinguish between the blocks as seen in the decoded plaintext and blocks as seen in one image share.

A.1 Formal User Study Questionnaire

Please answer the following questions

1. My confidence in my responses as I progressed through the tasks became:
   Higher
   Somewhat Higher
   Neither Higher nor Lower
   Somewhat Lower
   Lower
2. As I progressed through the study, my ability to identify characters/numbers became:
   Better
   Somewhat Better
   Neither Better nor Worse
   Somewhat Worse
   Worse
3. As I progressed through the study, my ability to detect illegal images became:
   Better
   Somewhat Better
   Neither Better nor Worse
   Somewhat Worse
   Worse
4. The clarity of the overlaid images was:
   Clear
   Somewhat Clear
   Neither Clear nor Blurred
   Somewhat Blurred
   Blurred
5. Had you used any AR technology before taking part in this study?
   Yes No
6. How would you describe your interaction with technology in your day-to-day life?
   Low Moderate High
7. Do you wear prescription glasses?
   Yes No
   (a) If yes, wearing prescription glasses made using the AR glasses:
   Easy
   Somewhat Easy
   Neither Easy nor Difficult
   Somewhat Difficult
   Difficult
   (b) Wearing prescription glasses made aligning the images:
   Easy
   Somewhat Easy
   Neither Easy nor Difficult
   Somewhat Difficult
   Difficult
8. Gender:
   Female Male Other
9. Age Range:
   18-25
   26-35
   36-45
   46-55
   56 and above

A.2 Mechanical Turk Questionnaire

Please answer the following questions:

1. My confidence in my responses as I progressed through the tasks became:
   Higher
   Somewhat Higher
   Neither Higher nor Lower
   Somewhat Lower
   Lower
2. As I progressed through the study, my ability to detect illegal images became:
   Better
   Somewhat Better
   Neither Better nor Worse
   Somewhat Worse
   Worse
3. How would you describe your interaction with technology in your day-to-day life?
   Low Moderate High
4. Please select your gender:
   Female Male Other
5. Please select your age range:
   18-25
   26-35
   36-45
   46-55
   56 and above

Unpacking security policy compliance: The motivators and barriers of employees' security behaviors

John M Blythe
PaCT Lab
Department of Psychology
Northumbria University, UK
john.m.blythe@northumbria.ac.uk

Lynne Coventry
PaCT Lab
Department of Psychology
Northumbria University, UK
lynne.coventry@northumbria.ac.uk

Linda Little
PaCT Lab
Department of Psychology
Northumbria University, UK
l.little@northumbria.ac.uk

ABSTRACT

The body of research that focuses on employees' Information Security Policy compliance is problematic as it treats compliance as a single behavior. This study explored the underlying behavioral context of information security in the workplace, exploring how individual and organizational factors influence the interplay of the motivations and barriers of security behaviors. Investigating factors that had previously been explored in security research, 20 employees from two organizations were interviewed and the data was analyzed using framework analysis. The analysis indicated that there were seven themes pertinent to information security: Response Evaluation, Threat Evaluation, Knowledge, Experience, Security Responsibility, Personal and Work Boundaries, and Security Behavior. The findings suggest that these differ by security behavior and by the nature of the behavior (e.g. on- and offline). Conclusions are discussed highlighting barriers to security actions and implications for future research and workplace practice.

1.1 INTRODUCTION

1.2 Employees and Information Security

Recently, attention has been drawn to the accidental disclosure of sensitive information and the role employees play in both its protection and leakage. In the UK, the governance of sensitive data belonging to living individuals is under the jurisdiction of the Data Protection Act (DPA; 1998) and governed by the Information Commissioner's Office (ICO). The ICO can sanction organizations up to £500, 000 for breaching the DPA as the leakage of sensitive data can cause harm and distress to individuals, including reputational and financial damages. The information stored by organizations is not restricted to living individuals as organizations also store information sensitive to their business operation, e.g. their intellectual property. Leakages of this sensitive information can negatively affect businesses' operation and reputation.

Despite the many negative consequences resulting from information disclosure, the prevalence of security breaches is high. For example, the PWC 2014 Information Security Breaches Survey found that 81% of large organizations and 60% of small businesses experienced a security breach in the previous year [1]. This survey indicates that breach rate is high but the severity of these breaches is wide-ranging. More severe cases can have repercussions to organizations; for example, in 2011 the Sony PlayStation Network was hacked leaking the personal information of its gamers. Alongside service disruption and damage to Sony's reputation, they were also fined £250, 000 by the ICO [28] for breaching the DPA (1998).

Employees are a mixed blessing when it comes to information security. They act as both a major cause of breaches and as the last line of defense. Research indicates that 46% of data breaches in the UK are due to insider negligence [32] and erroneous behavior when handling information [54]. To protect their organization's systems and data, employees must follow a number of security procedures to counteract security threats. These may include using strong passwords, encryption, anti-malware software and installing software updates. The specific responsibilities will differ by organization and be dictated within the Information Security Policy (ISP). These policies detail security actions employees are expected to take, some of which may be easier to follow than others.

Security procedures such as antivirus updates are now being automated to reduce the burden on employees [24]. However, other procedures such as password design are the direct responsibility of the employee. The degree to which an employee behaves securely may differ depending upon the level of effort required. Required effort is one of the many factors that influence employees' behavior.

A number of theories of behavior have identified different factors that influence behavior. In this paper we will review these factors and identify whether or not there is support for them in the security literature. We then present the findings of a qualitative study investigating these factors in two different research institutions.

1.3 Security Research Paradigms

Previous research into ISP compliance has been largely underpinned using models from behavior change literature to identify influencers of security behavior. These include Protection Motivation Theory, the Theory of Planned Behavior and the Health Belief Model. Studies exploring the validity of such models, which do focus on single behaviors tend to focus on private use of technology rather than workplace use [e.g. 22, 40].

The use of this "compliance paradigm" is criticized for its operationalization of security behavior as a single behavior referred to as 'compliance' [9]. ISPs dictate many different security behaviors (See appendix A for summary of ISP topics

Symposium on Usable Privacy and Security (SOUPS) 2015, July 22-24, 2015, Ottawa, Canada.

identified). Furthermore, there is little consensus on the content of security policies between organizations. This approach assumes employees' awareness of the content of these policies and finally when questioned about ISP compliance, different people may adopt different frames of reference depending on what is most salient to them at the time. These issues raise concerns about the validity of quantitative, survey research on policy compliance conducted across multiple organizations. Such research is often interested in exploring what motivates compliance behavior, but what influences compliance for one behavior might not influence it with another. For example, self-efficacy might be important to motivate compliance with password behaviors but not important for downloading software updates.

By reducing compliance to a single behavior it therefore limits our understanding of what influences individual security behaviors. Behavior change research acknowledges that motivation of behavior differs by behavior and context [20]. It is important within a work context to explore specific security behaviors rather than focusing solely on compliance with ISPs.

1.4 What influences secure behavior?

Models from behavior change are useful to understand the processes that underpin security behaviors. These can aid the design of interventions to promote secure behavior based upon the strength of the relationships between the theoretical constructs and the security behavior of interest. The two most frequently used theories are the Theory of Planned Behavior (TPB) [e.g. 12, 30], which identifies a link between attitudes and behavior, and Protection Motivation Theory (PMT) [e.g. 25, 33] which is a risk-perception theory exploring an individual's threat and response appraisal and their motivation to protect themselves.

The use of theoretical models facilitates the identification of factors that lead to employees' compliance with their organization's ISP or why consumers engage in a specific security behavior. In this section, the factors that have been consistently explored in research on security in the workplace and in home-users are discussed.

Self-efficacy is *an individual's beliefs about their competence to cope with a task and exercise influence over the events that affect their lives* [5]. In a security context, employees who have high self-efficacy are more likely to follow security procedures, as they are more effective in learning how to follow them and believe they are able to perform the required behavior. Self-efficacy is within many behavior change theories including PMT and social learning theory. Self-efficacy is consistently shown to influence security policy compliance [12, 25, 29, 30, 40, 55]. Furthermore, support has been found for a relationship between self-efficacy and virus protection behaviours [36], using a personal firewall [39], being cautious with emails that have attachments [40], and anti-spyware adoption [22, 37, 51] for consumers.

Social influence *is the extent to which an individual's behavior is influenced by what relevant others (e.g. colleagues) expect him/her to do and the extent to which they believe others are performing the behavior.* In a security context, employees are more likely to behave securely if those around them behave securely and expect such behavior of others. Employees' work environment and the individuals within this environment are therefore important drivers of security actions. The role of social influence is consistently shown to relate to compliance intention [12, 24, 25, 29, 30].

Attitude is *the individual's positive or negative feelings toward engaging in a specified behavior,* in other words towards behaving securely or complying with the ISP. The TPB argues that attitude is a predictor of behavior, alongside subjective norms and perceived behavioral control (a form of self-efficacy) [3]. The notion is that a positive attitude toward behaving securely influences intentions to behave securely. The influence of attitude on compliance intention has been consistently supported [12, 25, 29, 41]. Support has been found for a relationship between attitude and anti-spyware adoption [16], online privacy protective strategies [13, 59] and firewall adoption [35]. This suggests that attitude may be an important antecedent of security behavior.

Research has also explored individuals' threat and response evaluations in the context of security which stems largely from PMT [45]. The theory argues that individuals are motivated to protect themselves based upon their threat and coping appraisal. An individual's threat appraisal assesses the perceived susceptibility to the threat and the severity of the consequences. The coping appraisal is their evaluation of the response to the situation and consists of response efficacy and self-efficacy.

Perceived susceptibility is *an individual's assessment of the probability of events happening to them*. Individuals that have a sense that security attacks are unlikely, may not engage in security practices. On the other hand feeling susceptible to security attacks may result in protective behavior. The role of perceived susceptibility on compliance intention [30, 49] and use of anti-virus software by consumers [36] is supported. The relationship between perceived susceptibility and anti-spyware usage is not always supported [14, 22]. Recent research found perceived susceptibility did not play a role in employees' security breach concerns [25].

Perceived severity is *the assessment of the seriousness of a security threat and its associated consequences*. If an employee perceives a threat to the information resources of their organization to be severe, they are more likely to engage in security actions and adopt secure behaviors [12]. The relationship between perceived severity and secure behavior is not always supported. Support was found for the relationship to information security compliance [25, 49, 55]. However, other research found that perceived severity was not supported, attributing this to differences in the conceptualization of severity in previous studies [30]. The support for a relationship between severity and anti-spyware adoption [14, 22] has been found but its role in being cautious with emails that have attachments [40] and anti-virus protection has remained unsupported [36]. This further highlights that factors do not play the same role in all security behaviors.

Whilst some research [41] supported the role of susceptibility and severity on compliance intention, they combine these constructs so it is difficult to disentangle the effects.

Response efficacy is *the belief in the benefits of the behavior* [45] i.e. that a specific security behavior will reduce security breaches. On the other hand, if an individual has less belief in the efficacy of the behavior, they are less likely to adopt it. Response efficacy, which is part of PMT, has received less attention in research compared to other factors. The research that exists supports the relationship between response efficacy and ISP compliance [30], attitude toward security policies [25] and intention to adopt anti-spyware software [14, 22, 33]. Recent studies found a negative relationship with ISP compliance [55] or no relationship [49].

Response costs refer to *beliefs about how costly performing the recommended security behavior will be. These costs include money, time, and the effort expended*. If an individual perceives that a considerable cost is associated with a behavior, they will be unlikely to follow through with it. Conversely, if a small cost is incurred, the behavior may be adopted. The compliance budget [8] supports the role of response costs, they found that individuals and organizations place different values on the cost and benefits of different behaviors within ISPs. They argue that an employee's compliance or non-compliance is determined by the perceived costs and benefits of it. Mixed findings are reported in the literature; a negative relationship with ISP compliance has been found [25, 55] whereas other research has found no relationship [30]. Mixed findings are also reported for anti-spyware adoption [14, 22].

Despite the identification of factors that influence security behaviors, there is a lack of research that has explored these factors qualitatively, and how they may be moderated at the individual-level and within the organizational context. In other words, we are interested in what may cause high or low levels of these researched factors in the workplace. Appendix C provides an overview of the literature-driven framework to be explored in the current study.

1.5 Methods in Security Research

Quantitative methods have been primarily adopted in security research such as questionnaire studies that adopt regression models to investigate the degree to which factors influence ISP compliance [e.g. 24, 30], or security behaviors [e.g. 40].

Behavioral intention is seen as most proximate to behavior and is viewed as the best predictor of behavior [56]. Intention is the individual's motivation to undertake the desired behavior. Most existing research explores intention as it's easier to measure (self-reports) than actual behavior (objective measure). With the exception of some studies [58] which obtained objective security data about employees, there is over-reliance on this subjective measure. Research has indicated that intention only accounts for a third of the variance in actual behavior [48]. Research needs to focus on actual behaviors rather than focusing on intention to act.

Qualitative methods have been used to explore security behavior but have received less attention. This research has adopted a number of techniques including one-to-one interviews [4, 8, 53] and diary studies [31]. The lack of adoption of qualitative methods might be due to the potentially intrusive nature of information security research and concerns for business reputation of recruited organizations [34]. These concerns may be heightened due to rises in the number of security breaches in recent years and the imposition of fines on organizations by bodies such as the ICO.

Qualitative studies are useful to explore the motivators of and barriers to information security behaviors. Exploratory and inductive in nature, they aim to generate data pertinent to a research question that is not necessarily confounded by a particular theory or paradigm. However, there has been little research using a deductive approach. Deductive approaches within behavior change literature are quite common. Elicitation studies are one form of a deductive approach. These are useful for ensuring that beliefs and attitudes are data-driven from the population rather than pre-determined by previous research and the research team's preconceptions [18]. They can be used prior to questionnaire development [3, 38]. As the current study is interested in the interplay of factors that are part of these behavior change models for security behaviors, a deductive approach was considered more appropriate as it made space to understand how these factors may differ for different security behaviors and allowed additional themes to emerge that may have not been identified within previous research.

Behavior change models have been used to categorize qualitative data as they allow the exploration of the constructs of the theory with a target group [e.g. 46] and as a framework to analyze existing qualitative data in finance-related security behavior [15]. Apart from some research [15], this approach has remained relatively untapped in the information security domain.

1.6 Research Aims and Research questions

The present study aims to explore what influences secure and insecure practice within the workplace by understanding employees' attitudes, beliefs and security behavior. This study adopts a deductive approach to elicit behavioral determinants which have been previously explored in IS research. The following research questions are to be addressed:

> **RQ^1**. What are the influencers of employees' secure and insecure behavior and how might they differ across behaviors?
>
> **RQ^2**. What are the potential barriers to security behaviors?

2. METHOD

2.1 Approach

This study used a semi-structured qualitative approach and employed framework analysis to elicit factors that influence security behaviors through one-to-one interviews. Interviews were chosen over focus groups as the topic of security was deemed sensitive due to its links to employees' job performance.

The vignettes formed the focus for the interviews. 16 vignettes were developed for the current study covering the security behavioral categories identified from a review of ISPs collated from organizations (see appendix A). Vignettes were used as a tool to help engage participants with cyber security discussion in interviews. The nature of this research requires the disclosure of insecure practice and honest discussion from employees, the social desirability of this behavior and because it is directly linked to job performance may mean that this information is difficult to elicit from employees. Vignettes are versatile and can be used for a number of purposes including icebreakers to build rapport with participants, elicit attitudes and beliefs about a topic, and investigate topics that are sensitive to respondents [7]. They have been used for a variety of sensitive issues [26], and with vulnerable groups [6] in research.

Following advice from previous research, the vignettes were designed to remain relatively mundane and avoid unusual events and characters, whilst also appearing realistic [6, 19]. They also provided enough contextual information to enable a clear understanding of the situation but were ambiguous enough to ensure that multiple solutions exist [57]. The vignettes were designed around common security incidents related to the eleven categories identified from the ISPs (appendix A). Additional vignettes were provided for categories, which had many sub-categories. Common security incidents were identified through security provider's reports (e.g. McAfee, PwC), news reports, and the researcher's knowledge and experience. The vignettes focused upon low expertise behaviors, what research [53] has defined as "naïve mistakes" rather than focusing on malicious behaviors. The wording of vignettes was particularly important to ensure that they did not influence the respondent [57] so we designed the vignettes to avoid the consequences of the character's action (as we were interested in assessing perceived severity). The vignettes therefore remained ambiguous in whether the behavior and situation portrayed was secure or insecure. By avoiding the consequences of the characters action, we would be able to assess participants' perceptions of the consequences. This approach is emphasized in research [47] which argues that vignettes should have unresolved issues and finish at the high of tension in the story. The vignettes were neutral and covered behaviors people may not perceive as insecure but are known to be risky from a security perspective.

2.2 Participants

A purposeful sample of 20 participants was recruited from two organizations from the North East of the UK. We initially only had access to interview 10 participants from organization 2. We had the intention of interviewing more however we found that during data analysis that the same comments were emerging which suggested that interviewing more participants would not have led to further insight. The final sample size was adequate for framework analysis [43] and we were fortunately granted access to external companies, despite the known difficulties of sample access with this research topic in qualitative research [34]. All participants met the following criteria: (1) currently in full time employment, (2) used a computer for work on a daily basis and (3) dealt with sensitive information classified under the DPA or information sensitive to their company's intellectual property.

2.2.1 *Organization 1: A University*

5 males and 5 females took part, aged 25-49 years (mean 33.5, SD=9.07). Job tenure ranged from 9 months to 15 years with an average of 3.78 (SD=4.25) years. 4 participants had permanent contracts whilst 6 had temporary. All participants used a computer for more than 4 hours daily. Only 1 participant had read the ISP. All participants used personally-owned devices in the workplace and 9 conducted work tasks on their personally-owned devices. 7 participants also stored personal data on their work devices.

2.2.2 *Organization 2: Industry Research Group*

4 males and 6 females aged between 26-57 years (mean 39.10, SD=10.61). Job tenure ranged from 5 months to 27 years with an average of 11.12 (SD=10.89) years. 8 participants had permanent contracts whilst 2 had temporary. 9 participants used the computer for more than 4 hours daily whilst 1 used the computer for 3-4 hours. 9 participants had read the ISP: 2 had read the policy in the last 1-6 months, 2 had read the policy 6-12 months ago, and 5 in more than 12 months ago. All participants used personally-owned devices in the workplace and 6 conducted work tasks on these. 7 participants stored personal data on their work devices.

2.3 Procedure and Interview Guide

The study received approval from the faculty ethics board. Participants who met the criteria for participation were recruited using internal emails in the participating organizations. Participants were interviewed individually, in a private room at their organization and upon arrival were asked to read an information sheet covering all aspects of the investigation, including the purpose of the study and what they were required to do. They then provided written informed consent. Upon study commencement, participants were first required to complete a demographic questionnaire. They then took part in a semi-structured interview lasting 45-60 minutes. The interview was designed to be semi-structured to allow exploration of the initial framework and key issues and themes pertinent to the research question, while also allowing flexibility to probe unexpected topics raised by the participant [27]. An interview guide (see appendix B) was developed to elicit the behavioral influencers, which have been previously investigated in security research.

Participants were first introduced to a topic area (from the review of ISPs - see appendix A for full list of topic areas covered) in which the researcher provided a short description of the topic to ensure that the broad scope of information security was covered within the interviews. Participants were then presented with a vignette related to individual behaviors from the topic area. The vignettes were used to provide a safe way to open discussion around security for each topic and to encourage honest disclosure from participants. Upon presentation, participants were asked to imagine, drawing on his or her own experience, how they would react in that scenario. Following this, discussion centered on how participants currently behave in the workplace in relation to the ISP areas. At this point, the interview guide was used to elicit behavioral influencers for the behaviors discussed. We were also interested in potential factors that were not covered by the previous research and as such, further discussion for potential factors or reasons for their behavior not covered by the interview guide was encouraged.

Upon completion of the study, participants were presented with a debrief sheet which fully explained the purpose of the investigation and re-emphasized participants right to withdraw their data. Participants were all entered into a prize draw to win a £50 Amazon voucher.

3. ANALYSIS

The data was transcribed verbatim and analyzed in NVivo 9 using the principles of framework analysis [44]. The five-step procedure was used [52]: (i) the researcher is immersed in the data by transcribing and re-reading transcripts; (ii) identify emergent themes from the data. The current study identified these a priori from previous research, which formed the basis for the initial framework. However, new themes were allowed to emerge that were unaccounted for by the *a priori* framework and allowed the data to dictate the themes [44]. (iii) The data was then indexed in correspondence to the themes within the framework. (iv) Charts are used to arrange the data that was previously indexed in the third stage. The use of charts and maps allowed the data to be classified under headings that relate to the thematic framework. (v) The final stage, mapping and interpretation, involved the development of a schematic diagram from the analysis to guide the interpretation of the data. It was important that in the final stage that any conclusions drawn from the data echoed the underlying attitudes, beliefs and values of the participants [52]. Upon completion, two other researchers conducted a mini-audit of the analysis done by the lead researcher who were given the initial coding, quotes and identified any emerging themes for stages 2 and 3 of the framework analysis. Upon data completion, the two researchers also checked the final themes and associated quotes.

4. THEMES

Seven themes emerged from the framework analysis of the data. Appendix C provides a visual comparison of the initial and final framework. From the initial framework, self-efficacy, attitude and social pressures were not present however knowledge, experience, personal and work boundaries and security responsibility did emerge from the framework analysis.

Appendix D provides visualizations for each of these themes and Table 1 provides an overview of these themes.

Table 1. Emergent themes from the framework analysis

Theme	Brief description
Response Evaluation	Assessment of security behaviors as characterized by response efficacy, perceived benefits & response costs
Threat Evaluation	Appraisal of the threats to information security as influenced by individual threat models, susceptibility, severity & information sensitivity appraisal
Knowledge	Knowledge of security risks and security actions & the sources that contribute to this
Experience	Previous experience of security including security breaches & work experience
Security Responsibility	Employees perception of who is responsible for security in their workplace
Personal & Work Boundaries	Boundaries between personal & work life
Security Behavior	The actions employees take to ensure information security, categorized as high, medium or low security hygiene

Overall, we found no major differences between participants from each organization. The findings will therefore be discussed together; however any identified differences will be explained.

4.1.1 Response evaluation

Prior to undertaking a security action, employees evaluate the response and its associated outcomes. This is referred to as response evaluation, which is characterized by response efficacy, perceived benefits and response costs.

4.1.1.1 Response costs

Findings suggest that employees make a decision about whether to behave securely based upon an appraisal of the costs associated with the behavior. The major cost is the degree to which it impacts upon job productivity as there appears to be a "productivity threshold" regarding security actions. When the productivity threshold is reached, it can lead to a number of behavioral outcomes. For instance, the employee may circumvent the security process or disregard the security behavior. This was apparent for behaviors relating to information access such as password restrictions on information or accessing documents stored on servers. Furthermore, tasks such as restarting the work computer for security updates were also seen as impacting upon productivity. Employees recognise the disturbance these prompts for restart cause to their workflow and will subsequently postpone the task until a period of low activity or until the end of the working day.

"I will postpone it, postponing security updates happens a lot because they usually time them at really inconvenient times.. it's like well do you want me to do my job?...." (P14, Org2)

This security vs. productivity imbalance is also evident in software acquisition procedures. Organizations often place restrictions on the software employees can install on their work machines, requiring administration rights and authorization for the installation of new software. There were organizational differences in the current study with regards to how the companies mandate software acquisition. The university has a very restrictive policy in which employees do not have administration rights and must seek IT services to approve and install additional software. The industry research group had a less restrictive system allowing employees to freely install software. Both organizations had the option of allowing employees to install authorized licensed software from the company network. However, the lack of installation restriction within the research institution meant that employees did not consider the licensing agreements of certain software and would download software (such as freeware) without consultation. The official procedures for software acquisition were considered "time consuming" and requiring budget approval indicating monetary costs associated with acquiring legitimate software. Employees assumed that they would not gain budget approval and had developed a "don't bother" attitude with regards to official procedures which leads to risky software acquisition.

"because I know it is going to end up as a no anyway I just don't bother with that.. just save yourself the grief and go and get the free thing, that does the job equally well without the hassle.." (P14, Org1)

Correct software acquisition had the largest response cost – reduced productivity as it directly affects employees "*doing their job*". Monetary costs typically referred to the acquisition of software for personal devices (such as purchasing anti-virus).

Cognitive demands were another major cost which occurred as a result of using passwords. Employees have a number of passwords to remember and different password requirements are set for different systems, resulting in high cognitive demand.

"Well passwords.. after many years using computers the passwords just get longer and more complicated to remember, most of them are just randomly generated letters and numbers which can make them hard to remember especially if you.. well especially if you have to change them" (P6, Org1)

Not all security behaviors have response costs, as some actions require minimal time and effort by the user. Specifically the security behaviors of locking the computer, keeping a clear screen and desk policy, and checking physical environments when working in public locations were seen as having minimal costs. Employees identified that although these behaviors have smaller costs, a "habit" was required to ensure they follow through with the action.

".. there is no real effort on my part and I mean ultimately it is CTRL ALT DEL and you have locked your computer and that's all it is.. so it's not exactly an effort from my perspective.. that's probably it.. it doesn't delay me or put a burden on what I am

doing generally.... I would be a little bit more resistant if there was a lot more effort for me to do stuff…" (P14, Org2)

Previous research has mixed findings with regards to response costs and security behaviors [14, 22, 25, 30]. The current study suggests that each security behavior may have a different set of response costs that are not equally as costly as suggested by the ISP compliance paradigm. These differences in response costs for each security behavior may account for the mixed results in the security literature. The findings also support the "compliance budget" which suggests that individuals' choice to comply or not comply is determined by the perceived costs and benefits [8].

4.1.1.2 Perceived benefits

Overall, employees' understood the benefits of security behaviors in terms of protection of information and technology from malicious others, and maintaining confidentiality of data.

"advantages are that you can keep your information secure.. you can be confident that you're taking responsibility" (P2, Org1)

There was also an overall perception of "layers of security" in which the individual security actions help contribute to the overall picture of information security.

"It's like having a burglary, if you leave your door open it's like inviting someone in but if you put extra locks on, it's deterring them so I think the stronger your password is, the more of a deterrent it is to people.." (P8, Org1)

Employees also gain reassurance that their actions are aiding information security and they feel safer in what they are doing.

"I like it (anti-virus) because I think it's important, it gives you an element of security that what you are using is safe… so you don't have to worry as much.." (P18, Org2)

"..well I think having it there, whether its effective or not just makes me feel just a little bit safer.." (P1, Org1)

4.1.1.3 Response efficacy

The findings indicated that employees struggle to evaluate the effectiveness of security actions as they lack awareness and feedback of the result of their behavior.

"I don't know, if you password protected it whether somebody could still access it, I don't know. I guess they probably could" (P4, Org1)

Feedback appears to be playing a major role when employees evaluate the effectiveness of a security behavior. Employees don't receive information about their efforts so they are unaware of the utility of the security action. This indicates an "action-feedback" gap in employees' information security efforts.

"They say that if you don't notice something has gone wrong that is a sign of effectiveness, that's what they say so I am gonna go with I think it is working (anti-virus software)" (P14, Org2)

Furthermore, employees' response efficacy is capped as there was an overall "sense of insecurity" in that they believe hackers or the IT savvy will always be able get access, undermining the effectiveness of their efforts. However, they do perceive their efforts as effective against the average person or criminal.

"I think it's (encryption) effective.. if someone really wants to find out what is on there.. they will find out.. if they are a hacker.. but it's enough to stop.. like if Joe picked it up and put it into his computer and it said you can't read this file because it is password protected or encrypted in some way.. it may be enough to stop him and just hand it and say I have found this.. so again I think it is a good enough deterrent and as I say if someone for whatever reason really wanted what was on that stick.. I am sure they could find ways of cracking the encryption but it is a good enough deterrent for 90% of the population.." (P19, Org2)

Perceived benefits and response efficacy are types of outcome expectancies. Outcome expectancy is present in many of the theories of behavior. An individual's perceived benefit of security behaviors has received little research within security. Research has investigated users' perceived benefits of email security behavior, using the health belief model, on security behavior and supported the relationship [40]. However, this conceptualization refers to a user's perceived effectiveness of the behavior or "response efficacy". Perceived benefits in the current study, refers to individual's estimation of the advantages of engaging in security behaviors which may be distinct from an individual's efficacious perceptions.

At the end of the session, participants were asked to pick three security behaviors that they perceived to be most important for information security. The findings indicated that access control behaviors were perceived to be most important for security (n=19; such as using strong passwords and changing passwords regularly), followed by offline security behaviors (n=9, such as locking computer or using locked cabinets) and an awareness and responsibility of security (n=7, such as personal responsibility and treating information confidentially). Using security software (n=6) and security with removable media (n=4) were also seen as important. Internet (n=3) and email (n=2) security, company procedures (n-2), business continuity practices (n=1) and personal usage (n=1) were less prevalent. The findings indicate that whilst employees struggle to evaluate security actions, they do place more importance on some security behaviors over others, particularly behaviors related to access control.

The role of response efficacy has received little attention in research to date. Previous research has supported the relationship between response efficacy and factors such as intention to comply with security policies [30], attitude toward security policies [25], and intention to adopt anti-spyware software [14, 22, 33]. However, recent research has found contrasting findings [49, 55]. The current study highlights a potential barrier to high response efficacy, as employees cannot evaluate their security efforts as they lack feedback on their performance. However, they did indicate which behaviors they think are most effective for security with those relating to access controls having most perceived utility. Protection motivation theory argues that response efficacy is part of a person's coping appraisal and that higher levels of response efficacy will increase the likelihood of engaging in the behavior. This study suggests that employees do not receive feedback or information regarding security actions and the effectiveness of these actions. Response efficacy may therefore be a potential barrier to security behavior within the workplace.

4.1.2 Threat Evaluation

A number of factors that affect threat appraisal were identified.

4.1.2.1 Information Sensitivity Appraisal

Employees felt that the information they work with has different levels of sensitivity. However, perceptions of low data sensitivity were more prevalent in this sample. Their appraisal seemed to be based on an assessment of the "value" of the information. This entailed a comparison to data with a perceived higher value such as health-related and financial-related information.

"Again, vulnerable in the respect that I could probably do more but at the same time, I am not sure what other people could do with the stuff that I leave lying around, it's not highly confidential or anything like that... I haven't got peoples' bank details or anything like that.." (P9, Org1)

"I think you have got to think of a better way of giving yourself a reminder than having that exposed especially if it has got patient.. at that level healthcare that's.. you couldn't take any chances with that sort of thing so.." (P12, Org2)

Furthermore, employees' appraisal involved consideration of the information's "audience" and their preconceptions of who can use the data.

"..there is no objective value to this information that somebody has given us.. because to the vast majority of people it means absolutely nothing.. it's pointless and they would not be bothered even if they were found out"(P2, Org1)

These findings support research that found that employee's perceptions of information sensitivity interacted with their perceptions of organizational security [2], rated information about individuals as more sensitive than commercially sensitive information and placed security as a higher priority on some information. This study demonstrates this appraisal through employees' evaluation of the information's value and audience.

4.1.2.2 Susceptibility

Perceptions of susceptibility to security threats appeared to be an important factor in the employees' behavior. The perception varied between employees and the nature of the threat - offline or online.

Offline threats to information and systems involve physical attempts to infiltrate the information security of organizations, which can include the attempts of outsiders or malicious employees. Perceived susceptibility to these kinds of threats appears to be low amongst most employees. Individuals perceive that offline threats will be malicious others acting in a more opportunistic manner rather than pre-meditated. They appear to hold an optimism bias with offline threats, believing they are not at risk of being a victim and comparing the likelihood of a physical threat to other employees or other organizations.

"Yeah the physical security I feel fairly protected.. I would say also because of the likelihood of people who surround me to come and search through my files is just next to zero so yeah I feel very secure" (P3, Org1)

"so in that respect it's probably absolutely safe 99.99% of the time to leave completely personal information all over your computer and leave it unlocked because the majority of people that come into contact with it will not be interested and not want access to it and not want to do anything with it.. so it's only to protect for that minority of times.. for that possibility that somebody might want it and want access to it.." (P2, Org1)

With regards to online threats, the employees perceived themselves to be highly susceptible. There appeared to be an overall sense of insecurity or learned helplessness when it comes to behavior online. This is particularly related to employees' response efficacy. Individuals' have an estimation of the effectiveness of different types of security behaviors and practices, however they feel that *"hackers can still get access"* and the *"IT savvy can still bypass security"*. Employees understand the importance of security behaviors but feel that their efforts can be circumvented regardless.

"I have no idea.. probably they are (passwords) effective if you are going to protect yourself against somebody.. if you wanna kind of see security from the person next to you however in terms of people whose job it is to break passwords.. probably not very effective and I do realize that there are people out there whose vocation is to break peoples' passwords and virus peoples' computers…" (P3, Org1)

"For somebody like me I think your password would be enough to bar me from accessing your information, logging into your computer but I think somebody who had good sound IT knowledge could probably bypass them and get into other peoples' information" (P16, Org2)

The relationship between levels of susceptibility and engagement in security behaviors has mixed support in the literature. Its relationship with ISP compliance intention has consistently been supported [30, 49] as has its role in anti-virus software usage [36]. A potential reason for the lack of support in previous studies is that their conceptualization of threats is often non-specific and they do not refer to types of threat [e.g. 55]. This study demonstrates that an individual's threat assessment differs depending upon an online or offline threat, with online having higher perceived vulnerability amongst employees. Previous studies do not make this distinction when assessing perceptions of susceptibility. Perceived susceptibility to online threats is closely linked with response efficacy, i.e. they do not believe they are protected even if they behave securely.

4.1.2.3 Threat models

Employees appear to have a variety of security threat models. This is dependent on their knowledge of security risks, their perceptions of appropriate security actions and perceived likelihood of threats. For example, there appears to be a large discrepancy in attitudes towards writing down passwords. Some employees perceive this as being highly insecure and would not engage in this behavior, suggesting that they are more concerned with physical threats than online threats in password security.

"I am quite conscious that someone can find a scrap of paper that I have written with important company stuff on so I don't do that.. even for my personal stuff I don't do it" (P11, Org2)

Some employees may perceive this as being insecure but determine the likelihood of an online threat as greater than an offline threat.

"I just have like a note.. well.. I have a note with all passwords for all the different places where I need stuff, like online because there is too many passwords to remember so I need to have them written down somewhere.." (P1, Org1)

Other differences were notable in threat perceptions of working remotely and allowing unauthorized users to use work devices, locking work computers, and using encryption on removable media.

4.1.2.4 Severity

There was disparity in perceived severity of security breaches and of security non-compliance across different domains. Employees were mainly aware of the consequences to their organization's reputation and the potential implications of this. For example, competitors getting hold of their company's intellectual property and breaching government legislation.

"again other than the competitive threat that we are developing something that we don't want the competition to know about and they get access to that information... you know something like that I guess would be of value to the competition so that they would then have time to put a counter strategy together"(P16, Org2)

Employees were highly aware of the impact to technology from a security breach. This was primarily the consequences of downloading a virus or other malicious software.

"I suppose technically it could affect the whole university system which would cause massive outrage and whatever, so I think you would get into a lot of trouble for doing stuff like that and I think it would have large consequences" (P9, Org1)

Perceptions of personal consequences were mixed; employees were not aware of how their company would react if they caused a breach in security. Employees assumed it might lead to disciplinary action or impact their own and companies' productivity. Employees seemed to consider the consequences to others less although did mention dissatisfied service users and distressed service users.

"I am aware of the kind of potential problems that you could cause, and the stress you could cause people if any information was disclosed about a particular person but I don't know if I did something that caused a problem within the university systems I don't know what action would be taken" (P7, Org1)

Previous research has focused on the role of perceived severity in ISP compliance [25, 49], and anti-spyware adoption [14, 22]. The role of perceived severity on anti-virus adoption [36], being cautious with emails that have attachments [40] and other ISP literature [30] is unclear. Our findings suggest there are different levels to an individual's perceived consequences or perceived severity. These are consequences to the organization, technology, 3rd parties and to the self. Within these levels, knowledge of the consequences also differs with less awareness of consequences to others and to oneself. This suggests that an individual's perceived severity is not one overall construct but may comprise of different types of severity implications. This may account for the differences in existing research.

4.1.3 Experience

Experience related to individuals experiences of security beaches and previous work experience.

4.1.3.1 Security breach experience

The current study suggests that previous experience appears to be important for current behavior. Previous job roles and experiences of security threats (including viruses and phishing emails) appear to promote awareness and secure behavior. An employee's experience of security breaches can lead to different courses of action depending upon their evaluation of an effective response to the breach. Employees' reported "security overreactions" in which they undertake inappropriate continuity behavior or take a "scattergun approach" to dealing with the breach by engaging in multiple behaviors to ensure recovery and continuity (e.g. deleting all contacts and changing all passwords).

"I mean once.. something must have happened to my email address, my yahoo email address because people were just getting emails just saying "try this money making scheme" so as soon as I got that.. I deleted everyone off my contact lists because I had them somewhere else and changed my passwords and things like that.." (P2, Org1)

Other reported "security overreactions" were non-use of accounts and concluding that devices should be thrown out following a virus infection.

"I could see that it is not a right file and I have no idea why I clicked on it and the computer is now very slow and unusable so we are going to be binning it or selling it for parts.. no reason for that and it shouldn't be happening.. and we know that we should never disable the anti-virus" (P3, Org1)

These experiences typically refer to personal experiences; however work-related experience is also important for secure behavior especially when it impacts on employees' productivity. For example, an employee's organization experienced a virus breach leading to implications that affected the whole business operation.

"this is not some pen pusher saying don't use pen drives.. It's actually really serious and that was a good lesson for me and I think a lot of people don't understand the importance of things like that but because I have got experience of what happens.. of what could go wrong.. when it goes bad.. when it goes wrong it goes wrong really badly.." (P15, Org2)

4.1.3.2 Work experience

Organizations differ in their approaches to information security and subsequently their methods to promote security awareness and practices amongst employees. This is known as the "security culture" of an organization, which are the shared values and assumptions regarding information security. An organizations' culture is idiosyncratic so there will be differences in the levels of security culture across companies. Employees discussed transfer of their behavior from previous organizations; this appears to be more evident in employees who come from organizations with a higher security culture than their current employer.

"Again from my previous job there was.. it was a very secretive company and there was a lot of examples where there was competitor espionage and things like that.. it was a very regular occurrence and a very serious thing so security was.. it was like Fort Knox over there most of the time so it just got drilled into you to lock your computer work station so that is just something that I brought with me to this job.. I notice that a lot of people don't lock their work stations here" (P11, Org2)

However, not all behaviors are transferred, there appears to be a threshold where employees will not transfer the behavior if it requires too much effort on their part. For example, strong password enforcements in previous companies do not lead employees to adopt a strong password management practice in their current job if it is not enforced.

"I have had the same password for the last 6 and a half years ... I know I should change that, in my previous employer we got sent a

reminder to change the password,…every three months we had to change our password… I know I should change it but I just don't have the memory space to do that.. I would forget what I had changed it to" (P9, Org1)

Experience has received little investigation in previous research and has largely been supported in terms of anti-spyware usage [51], adoption of online privacy protections [59], and adoption of virus protection behavior [36]. These findings suggest that previous breach experience is important for current behavior. Furthermore, employees' experiences of security in previous jobs are also important and potential transferability of behavior has not been formally explored in employee security behavior.

4.1.4 Security-related knowledge

The theme of security knowledge comprises of sources of knowledge and knowledge of specific domains (i.e. security risks and security actions).

4.1.4.1 Security risks

This study revealed that knowledge of security risks is diverse and varies depending upon security behaviors and security threats. Awareness of risks specific to poor password management is most prevalent and indicates that employees are able to identify the risks associated with: using poor passwords, not changing passwords, disclosure of passwords, recycling passwords and writing passwords down. Furthermore knowledge of risks associated with employees having administrative rights, risks when working remotely, viruses, and social engineering tactics such as phishing emails were also high. Knowledge of risks associated with mobile devices, removable media and physical security was mixed, with mobile devices in particular an area where employees lack awareness of the risks of using mobile devices and the potential vulnerability of these devices.

4.1.4.2 Security actions

Employees' knowledge of security actions was also mixed, particularly with regards to those that are formally set in their organizations' ISP. Analysis revealed differences in employees' knowledge of the security policy and its associated procedures between the two recruited companies. Information from the demographic questionnaire indicated that in the academic institution only 1 employee had read the policy compared to the other organization in which 8 had read their companies' policy. Whilst reading the policy does not indicate compliance to it or awareness of the entire content, it does appear to be a source of reference for some employees when determining appropriate security actions. Those who are unaware of their ISPs rely on their own awareness of appropriate security actions when behaving with information and technology. Consequently, they report relying on other sources of knowledge to inform appropriate security actions (such as recommendations from fellow employees).

In terms of security actions, encryption for removable media and work devices was the security action in which employees lacked most awareness of and sometimes there was clear confusion between the differences between encryption and password protection. Other security actions employees appeared to be knowledgeable of were those associated with authenticating users, physical security of information and technology, and the prevention of malicious software. Two-factor verification for account access (e.g. cloud storage) was mentioned less and could be a potential behavior that requires further awareness.

4.1.4.3 Sources of Knowledge

Employees sourced security information from individuals within their workplace or social circle whom they regard as having "IT expertise". In the workplace this was employees from the IT department or colleagues/friends with IT expertise.

".. I think it's pretty good.. I have got windows laptops and I have got a mac and.. I have done research on the different virus software that you can use which is freely available.. I only use the freeware stuff.. and I have asked my friends as well who are quite up on computers and what not and I make sure that I use kind of the same ones that they do.." (P1, Org1)

To a lesser extent, fellow colleagues and line management were sources of knowledge and this most commonly related to the receiving of suspicious emails or files, in which case they would seek information from their immediate peers before contacting "IT expertise" sources. Other sources of knowledge reported were company procedures such as the information security policy or professional codes of conducts, which cover aspects relating to the integrity of information and its security. For example, one employee has to sign non-disclosure agreements (NDA) with service users and this influences her behavior.

"I probably used to leave my computer unlocked more.. but in the job that I do now we have to sign non-disclosure agreements so if you are working with a university on certain things or different companies you have to sign NDAs and there have been some projects which have been deemed as pretty secret I guess so you have to sign them and say that you won't talk to anybody about them.. you won't.. and as part of signing them it says when you leave your desk you must lock your PC.. you will adhere to this and stuff so I am very aware of doing that.." (P19, Org2)

The media was another source of information such as reports about hacking to consumers and organizations and their associated consequences such as identity theft and fraud-related experience (individual) and network disruption and reputation (organizational). Media reports relating to security risks and their implications were also noted, such as government bodies losing unencrypted USB sticks with sensitive information on them.

"Well.. so far it's not too bad other than there has been a few cases where we have seen.. Facebook or LinkedIn passwords being cracked so the information that I have got on Facebook isn't particularly of interest but of course then when you go into online banking and everything that's when it starts to get a bit scary.." (P17, Org2)

4.1.5 Personal and work boundaries

An important factor influencing secure and insecure behaviors is the degree to which individuals engage in personal activities on their work devices and the boundaries they have between home and the workplace. Those who reported strong boundaries between home and work limit the personal usage they conduct (e.g. using work email for work-use only and limiting personal browsing).

"Well actually when I am at work I just do work and usually the sites and places that I visit on the web are educational resources.. I don't really surf the web and stuff and don't just click on

random links... I just stick to work related things and like I assume those kind of resources are pretty clear" (P12, Org2)

These strong boundaries extend to outside the physical workplace and relate to the use of work devices for personal usage when working remotely. Employees with strong personal boundaries said they use work devices solely for work purposes and don't allow unauthorized users (e.g. family, friends) to use them.

"Don't let anyone else use the computer. No one would want to use the computer anyway but I don't let anyone else use it... I don't like leave it in anyone else's care.. it's always kind of, under my own care because it's not my computer to pass around" (P8, Org1)

These individuals also demonstrate a preference for using work-issued devices over their personal devices for work tasks. They may therefore be less likely to engage in BYOD activities.

"*Try not use personal devices.. that is as close as it gets.. I just view it as a work one, it's just that I am using it with two different works.. I don't use.. I think it's important in my mind having that line for a couple of reasons.. the information that is coming out of work, I don't want it stored on my home stuff for any trace of it..*" (P4, Org1)

The role of technology in employees' work/life balance is well documented in organizational psychology literature. Ubiquitous access to the workplace can enhance individual productivity but can also inflate individual's stress levels leading to job burnout [42]. A strong work life balance may also be important for security. Limiting working remotely is important for security as it can reduce security risks associated with working outside of the workplace. Individuals with a high work/life balance limit doing work tasks outside of the workplace.

".. once I leave work that is me done but for serious work.. I know for example my boss and other people they have work laptops and they can work from home.. they get special equipment where they can do that.. it's not really applicable to me.." (P14, Org2)

Employees report feelings of high psychological ownership of their personal devices and limit work-related information.

"Yeah I don't even know if it is a security conscious thing.. I think it is more just.. work/life balance of this is my phone.. I don't want to contaminate it with work stuff... yeah it's mine, it's not the company's" (P19, Org2)

Individuals with blurred boundaries between personal and work usage reported being less restrictive in their boundaries and engage in personal tasks on work devices. For example, email usage for work and personal.

"I kind of do receive emails from my friends at work coz they also work here but I don't receive emails from my friends who don't work here on that account but at the same time I also have it set up so that I do receive my Gmail stuff to that computer as well so it sort of kind of blurs the boundaries a little bit" (P6, Org1)

When working remotely these boundaries are more blurred, employees may use work-issued devices for personal usage and allow others to use the work devices.

"I have done it myself if my nieces have been up and there is only one laptop.. like my own personal one and someone wants to do something else then I would give them the work laptop to do it.."(P19, Org2)

Employees reporting less distinctive boundaries between home and the workplace consequently have a lower work/life balance, they prefer ubiquitous access to work information so may use their own personal devices to stay connected to work. These employees also engage in more personal risky tasks on their work machines and disclose their own sensitive information such as discussed by the following employee who uses online banking on their work computer as they rely on the security of their organization and assume that it is more secure than their own devices.

"Because everything on mine (home computer) is what I have put onto it or set up to work on it or adjusted the settings and I don't really understand what I am doing with stuff like that so you assume that because you get an email from IT services periodically that goes to all users that says that we have identified a machine which is running malware on the network and they will give you the work station name of it and you eventually track it down, you assume that because it's a corporate computer system that there is some money and some resource and expertise at keeping it safe.." (P13, Org2)

The use of personal devices in the workplace or BYOD (Bring Your Own Device) can bring many advantages for businesses including enhanced employee productivity, satisfaction and mobility [10]. Despite this, BYOD also leaves organizations open to information breaches. Despite calls for organizations to implement more stringent BYOD security strategies [10], there is little research exploring employee attitudes towards BYOD, the factors that influence this form of behavior and the role of personal device ownership on information security. This study sheds some light on security behaviors and BYOD activities relating to work/life boundaries.

4.1.6 Security responsibility

Employees rely heavily on "security experts" in their company to maintain their systems, particularly for anti-virus, encryption, and installing updates. Employees recognize that it is their responsibility to handle passwords and protect data.

"To be honest I assume that if that's what the company tell us to use then somebody in the technology area has decided that it is secure enough and that our firewalls are there and whatever" (P16, Org2)

Relating to the prevention of viruses and other malicious software, employees appear to rely heavily on their organization with assumptions that "*somebody else is taking care of it*" and relying on the expertise of IT to ensure that they are protected.

"Yeah actually I haven't checked what it is and how it works and whether I should do something about myself or if it's something that just works in the background.. I'm hoping that it's just something that's in the background and then its updated automatically.. I haven't checked so far, I always just assume that's updated centrally from the IT services" (P10, Org1)

In adoption of new security practices, diffusion of responsibility was apparent. Employees would only adopt a new security behavior if the company enforced it, diffusing responsibility to the organization to force them.

"Yeah I would be quite happy to do it if the company came out and said every USB stick that you put in has to be encrypted and yeah I would do it.. again it becomes that another hurdle to get through in the productivity of work but I can understand that reasoning for it.." (P19, Org2)

This diffusion of responsibility was not just limited to the organizations that the employees work for but to service and product providers they use for work tasks. For example, there was a general perception that Apple products are more secure so you do not need to add any additional security - you can rely on Apple for the security.

"I have got a mac at home so as far as I know I don't need any security on it.. it has got its own inbuilt" (P12, Org2)

The current study supports the findings of existing research [17] which found that individuals delegate responsibility to one of four modalities: technology, individuals, organizations and institutions. However, its relationship to specific security behaviors in existing quantitative studies has remained relatively unexplored.

4.1.7 Security behavior

Security behavior refers to an employee's ability to engage in appropriate and effective security actions. Three aspects to security behavior were identified and employees categorized accordingly, referred to as "security hygiene", which indicates the effectiveness of the security actions employees undertake. The previous themes affect the degree to which an individual engages in high, medium or low security hygiene. Security hygiene is determined by prevention strategies and security citizenship.

4.1.8 Prevention strategies

Prevention strategies are behaviors that contribute towards information security in the workplace and aim to prevent security breaches. For example, not downloading suspicious attachments, not clicking on suspicious links online, adopting strong passwords, locking computers, encrypting removable media and non-disclosure of sensitive information to name a few.

Employees with high security hygiene take appropriate action and take fewer risks with their security behavior. They rely less on their organization for security and have a more proactive stance towards security. They can also correctly identify whether a physical or cyber security deterrent is most suitable for the security threat. For example, they will adopt encryption on removable media rather than rely on keeping it on oneself.

"Yeah I use a USB stick with encryption and it's just a bit of a reassurance because having in the past, I haven't lost a USB stick but I have not been able to find it for a few hours, dunno where I have put it and so feel a lot more comfortable now where there is using a USB stick with actual encryption on and knowing that if it did disappear then, you know, there wouldn't be staff information going into the wrong hands.." (P4, Org1)

Those with medium security hygiene may take appropriate action and know which security actions are most suitable but engage in more risks with their behavior such as creating less strong passwords and then writing it down or locking the desk cabinet but leaving the key located within the vicinity. They are less proactive in their stance towards information security and rely more on their organization for security.

"I put them in the filling cabinet but I didn't actually lock it but they were out of sight so I suppose that is as far as I went.. I didn't lock but I do remember going I shouldn't just.. because they are so easy.. it's not like a computer or a laptop that you would be seeing walking out with, the mobile phones were just too easy to pick up so yeah I put them out of sight but I don't think I actually locked them" (P10, Org2)

Employees with low security hygiene, lack awareness of appropriate security actions and engage in inappropriate security behaviors. They rely heavily on "security defaults" such as using the default security password and relying on the computer to auto-lock when leaving their desk. They are more reactive towards security needs and rely on security enforcement by their organization for their security behavior. They lack awareness of appropriate security actions for physical or cyber security threats and as such, they may engage in non-technical deterrents when a cyber-security deterrent would be more beneficial. For example, relying on physically securing a USB rather than using encryption.

"however the advantages are that I am much more consciously aware because 15-20 times a day I need to pick my keys up and I would notice if the USB.. because the USB stick is attached to a.. like a lanyard thing that goes around your neck so if that was missing I would be really consciously aware of it.." (P2, Org1)

Their behaviors are considered more negligent as they may be aware of security actions but fail to perform the behavior.

"I have kind of blurred the lines a bit by having a laptop, it mostly stays at home but when I do take it to work, it's sensible to have a password on but I just don't for ease of access" (P6, Org1)

4.1.8.1 Security citizenship

This refers to actions individuals engage in which aid the organization in business continuity and recovery. Individuals with high security hygiene seemed to engage in practices such as backing up data and informing colleagues of security issues.

"Well.. the phishing thing.. they are all set up.. I don't mess around with them, I just leave it as it is.. if I see anything dodgy I have emailed like IT before and made them aware of it and sent them the email" (P1, Org1)

Individuals with low security hygiene, on the other hand, rely more on their organization for business continuity practices and take less responsibility and action to aid the organization.

"No.. that's the one thing that I am really a bit confused about, I don't know if there are like official procedures for backing up or if I should do it myself.." (P20, Org2)

5. CONCLUSIONS

Overall seven themes emerged through the use of this deductive approach that explains why employees engage in security actions. The findings of the study suggest that the following relationships between the factors may be present (see appendix C for graphical overview of the initial and final framework). This study suggests that employees' security behaviors are influenced by their security knowledge and prior experience. Prior to carrying out the behavior, employees undergo threat and response evaluations. Knowledge and prior experience also influence these evaluations. Additionally, their perceptions of responsibility and boundaries between personal and work influence behavior. Finally, the interplay of all these factors influences the degree to which employees engage in security behaviors. This study indicates that there are different levels of security behavior characterized by prevention strategies and security citizenship.

The use of the deductive approach incorporated factors from many behavior change theories which allowed the comparison of the final framework with existing theory. The final framework suggests an extended PMT model with other security-contextual factors that may be able to explain additional variance in behavior if it was to be explored quantitatively and with regression analysis. By exploring these constructs qualitatively, we were able to explore what leads to high or low levels in these constructs and the individual, system and organizational components that may influence different perceptions. In doing this, it has provided better clarity of the use of PMT in security and may explain the disparate findings for a number of PMT constructs (severity and response costs).

The current study has provided a number of contributions to the security research area and organizational practice. Firstly, the findings demonstrate that ISP compliance is complicated as different security behaviors are motivated by different factors and to different degrees. Where possible, future research should move away from using an ISP compliance paradigm and focus on individual security behaviors. Likewise, organizational campaigns would benefit more from targeting specific security behaviors.

Secondly, response efficacy was shown to be a potential barrier to some security behaviors, response efficacy is low because employees lack feedback on how effective their security behavior is at reducing threats. Systems rarely provide enough feedback or positive reinforcement to users on their *proactive* security behavior although sometimes provide information on their *reactive* behavior (e.g. weak password or non-updated system). Systems need to provide more feedback on their efforts and provide information on the effectiveness of these for prevention of security threats. Furthermore, employees perceive that their security efforts may be in vain as they don't receive reinforcement from their organization/management to keep up their behavior. Research shows the importance of management feedback on employee performance [23] and the importance of positive reinforcement in shaping behavior [50]. One approach may be for organizations to include security behavior as part of the performance appraisal of employees. As security is part of an employee's job role, it should be given more focus and feedback from the attention of management during day-to-day business operation and more specifically, as part of their employees' performance appraisal.

Thirdly, the current study showed that employees undergo an information sensitivity assessment, evaluating the sensitivity based upon their perceptions of the value of the information and the audience for it. The study highlights differences in individuals' threat evaluation; employees' perceived susceptibility differs depending upon off- and online threats. Within information security research, off- and online threats are often given equal weighting or not specified. However, this study suggests that research needs to consider these as two separate information security issues (on- vs offline) and campaigns need to focus on communicating susceptibility to these threats differently to employees and being specific when framing susceptibility questions. More work is required to provide concrete definitions of sensitivity levels, rather than it being determined in relation to other types of information.

Fourthly, security responsibility was an emergent theme which suggested that employees perceived different responsibilities for security tasks, some of which they accept responsibility for and others they diffuse the responsibility onto their organization. Organizations need to be more transparent to employees with regards to what they are expected to do and what is within their remit. Organizational policies dictate these responsibilities however they need to be embedded within the culture of the organization. Finally, employees' personal/work boundaries may help explain risky behavior in the workplace and adoption of BYOD has implications for these boundaries. These boundaries need to be explored further.

The initial deductive framework included the factors social pressures, attitude and self-efficacy however these did not emerge within the final framework. Attitude emerged more broadly across the other constructs rather than as a separate construct. For example, security responsibility and personal/work boundaries have attitudinal components within them. For social pressures, when discussing security behavior, employees didn't appear to be concerned about the behavior of others and of their line management, with regards to their motivations for behaving securely. However, this factor may play more of a larger component within the security culture of both of the organizations. Previous research has explored the role of security culture, which is the shared beliefs, norms, values and learned ways that have developed through the organization's history [11] and are captured in the mission statements and the vision of the organization as they are the values they wish to be known for. A poor security culture is one where security is not built into these shared assumptions and is not part of "*the way things are done around here"*. In the absence of a security culture, individual-level motivational factors may play more of an important role as information security is at the level of the employee rather than driven top-down and across the organization. This may account for the lack of discussion around social pressures in the two participating companies.

Self-efficacy proved difficult to assess within an interview context and this could be due to difficulties in tapping into an individual's perceived capabilities of engaging in security tasks. Self-efficacy may play a latent but difficult to assess role due to impression management in organizations [21]. Employees may wish to maintain the perception that they are competent in their job roles so may not wish to disclose information that may negatively affect these perceptions (i.e. an inability to undertake security actions).

The use of a deductive elicitation approach proved a useful application for exploring the factors that influence security behavior. Refinement of the initial framework through the qualitative data allowed the emergent factors to be driven fully from the data set but also allowed comparison with the behavioral determinants identified *a priori* from the existing literature. Furthermore by using this approach it allowed exploration of theoretical constructs with target populations ensuring that behavioral motivators are data-driven rather than pre-determined by the research. This is important for behavior change as it allows the data from the qualitative interviews to be used for questionnaire and intervention development in future research.

6. REFERENCES

[1] 2014 Information security breaches survey: Full technical report: 2014. *http://www.pwc.co.uk/audit-assurance/publications/2014-information-security-breaches-survey.jhtml.*

[2] Adams, A. and Sasse, M.A. 1999. Users are not the enemy. *Communications of the ACM*. 42, 12 (1999), 41–46.

[3] Ajzen, I. 1991. The theory of planned behavior. *Organizational Behavior and Human Decision Processes*. 50, 2 (Dec. 1991), 179–211.

[4] Albrechtsen, E. 2007. A qualitative study of users' view on information security. *Computers & Security*. 26, 4 (Jun. 2007), 276–289.

[5] Bandura, A. 1977. Self-efficacy: toward a unifying theory of behavioral change. *Psychological review*. (1977).

[6] Barter, C. and Renold, E. 2000. "I wanna tell you a story": exploring the application of vignettes in qualitative research with children and young people. *International Journal of Social Research Methodology*. 3, 4 (2000), 307–323.

[7] Barter, C. and Renold, E. 1999. The use of vignettes in qualitative research. *Social research update*. 25, 9 (1999), 1–6.

[8] Beautement, A., Sasse, M. and Wonham, M. 2009. The compliance budget: managing security behaviour in organisations. *In Proceedings of the 2008 workshop on New security paradigms* (2009), 47–58.

[9] Blythe, J.M. 2013. Cyber security in the workplace: Understanding and promoting behaviour change. *Proceedings of CHItaly 2013 Doctoral Consortium* (2013), 92–101.

[10] Bring your own device: Agility through consistent delivery: 2012. *http://www.pwc.com/en_US/us/increasing-it-effectiveness/assets/byod-1-25-2012.pdf.*

[11] Brown, A. 1998. *Organisational Culture*. Pitman Publishing.

[12] Bulgurcu, B., Cavusoglu, H. and Benbasat, I. 2010. Information security policy compliance: an empirical study of rationality-based beliefs and information security awareness. *MIS Quarterly*. 34, 3 (2010), 523–548.

[13] Burns, S. and Roberts, L. 2013. Applying the Theory of Planned Behaviour to predicting online safety behaviour. *Crime Prevention and Community Safety*. 15, 1 (Feb. 2013), 48–64.

[14] Chenoweth, T., Minch, R. and Gattiker, T. 2009. Application of protection motivation theory to adoption of protective technologies. *Proceedings of the 42nd Hawaii International Conference on System Sciences* (2009), 1–10.

[15] Davinson, N. and Sillence, E. 2014. Using the health belief model to explore users' perceptions of "being safe and secure" in the world of technology mediated financial transactions. *International Journal of Human-Computer Studies*. 72, 2 (Feb. 2014), 154–168.

[16] Dinev, T. and Hu, Q. 2007. The Centrality of Awareness in the Formation of User Behavioral Intention toward Protective Information Technologies. *Journal of the Association for Information System*. 8, 7 (2007), 386–408.

[17] Dourish, P., Grinter, R.E., De La Flor, J.D. and Joseph, M. 2004. Security in the wild: User strategies for managing security as an everyday, practical problem. *Personal and Ubiquitous Computing*. 8, 6 (2004), 391–401.

[18] Downs, D.S. and Hausenblas, H.A. 2005. Elicitation studies and the theory of planned behavior: a systematic review of exercise beliefs. *Psychology of Sport and Exercise*. 6, 1 (Jan. 2005), 1–31.

[19] Finch, J. 1987. The vignette technique in survey research. *Sociology*. 21, 1 (1987), 105–114.

[20] Fishbein, M. and Cappella, J. 2006. The role of theory in developing effective health communications. *Journal of Communication*. 56, (2006), 1–17.

[21] Gardner, W. and Martinko, M. 1988. Impression management in organizations. *Journal of management*. 14, 2 (1988), 321–338.

[22] Gurung, A., Luo, X. and Liao, Q. 2009. Consumer motivations in taking action against spyware: an empirical investigation. *Information Management & Computer Security*. 17, 3 (2009), 276–289.

[23] Hackman, J. and Oldham, G. 1976. Motivation through the design of work: Test of a theory. *Organizational Behavior and Human Performance*. 16, 2 (1976), 250–279.

[24] Herath, T. and Rao, H.R. 2009. Encouraging information security behaviors in organizations: Role of penalties, pressures and perceived effectiveness. *Decision Support Systems*. 47, 2 (May 2009), 154–165.

[25] Herath, T. and Rao, H.R. 2009. Protection motivation and deterrence: a framework for security policy compliance in organisations. *European Journal of Information Systems*. 18, 2 (Apr. 2009), 106–125.

[26] Hughes, R. 1998. Considering the vignette technique and its application to a study of drug injecting and HIV risk and safer behaviour. *Sociology of Health and Illness*. 20, (1998), 381–400.

[27] Hutchinson, S. and Wilson, H.S. 1992. Validity threats in scheduled semistructured research interviews. *Nursing Research*. 41, 2 (1992), 117–119.

[28] ICO 2013. Sony Fined £250, 000 after millions of UK gamers details compromised. *https://ico.org.uk/about-the-ico/news-and-events/news-and-blogs/2013/01/sony-fined-250-000-after-millions-of-uk-gamers-details-compromised/.*

[29] Ifinedo, P. 2014. Information systems security policy compliance: An empirical study of the effects of socialisation, influence, and cognition. *Information and Management*. 51, 1 (2014), 69–79.

[30] Ifinedo, P. 2011. Understanding information systems security policy compliance: An integration of the theory of planned behavior and the protection motivation theory. *Computers & Security*. 31, 1 (Nov. 2011), 83–95.

[31] Inglesant, P. and Sasse, M. 2010. The true cost of unusable password policies: password use in the wild. *In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (2010), 383–392.

[32] Institute, P. 2010. *2009 annual study: UK cost of a data breach. Ponemon Institute*.

[33] Johnston, A.C. and Warkentin, M. 2010. Fear appeals and information security behavior: An empircal study. *MIS Quarterly*. 34, 3 (2010), 549–566.

[34] Kotulic, A.G. and Clark, J.G. 2004. Why there aren't more information security research studies. *Information & Management*. 41, 5 (May 2004), 597–607.

[35] Kumar, N., Mohan, K. and Holowczak, R. 2008. Locking the door but leaving the computer vulnerable: Factors inhibiting home users' adoption of software firewalls. *Decision Support Systems*. 46, 1 (Dec. 2008), 254–264.

[36] Lee, D., Larose, R. and Rifon, N. 2008. Keeping our network safe: a model of online protection behaviour. *Behaviour & Information Technology*. 27, 5 (Sep. 2008), 445–454.

[37] Lee, Y. and Kozar, K. 2008. An empirical investigation of anti-spyware software adoption: A multitheoretical perspective. *Information & Management*. 45, 2 (2008), 109–119.

[38] Montaño, D.E. and Kasprzyk, D. 2008. Theory of reasoned action, theory of planned behavior, and the integrated behavioral model. *Health behavior and health education: theory, research, and practice*. K. Glanz, B. Rimer, and K. Viswanath, eds. Jossey Bass. 67–96.

[39] Ng, B. and Rahim, M. 2005. A socio-behavioral study of home computer users' intention to practice security. *PACIS 2005 Proceedings* (2005), 234–247.

[40] Ng, B.-Y., Kankanhalli, A. and Xu, Y. 2009. Studying users' computer security behavior: A health belief perspective. *Decision Support Systems*. 46, 4 (Mar. 2009), 815–825.

[41] Pahnila, S., Siponen, M. and Mahmood, A. 2007. Employees'behavior towards is security policy compliance. *Proceedings of the 40th Annual Hawaii International Conference on System Sciences* (2007), 156b–156b.

[42] Peeters, M., Montgomery, A., Bakker, A. and Schaufeli, W. 2005. Balancing Work and Home: How Job and Home Demands Are Related to Burnout. *International Journal of Stress Management*. 12, 1 (2005), 43–61.

[43] Ritchie, J., Lewis, J. and Elam, G. 2003. *Qualitative Research Practice: A Guide for Social Science Students and Researchers*. Sage: London; Thousand Oaks; New Delhi.

[44] Ritchie, J., Spencer, L., Bryman, A. and Burgess, R. 1994. Analysing qualitative data. *London: Routledge*. (1994).

[45] Rogers, R.W. 1975. A protection motivation theory of fear appeals and attitude change. *Journal of Psychology*. 91, (1975), 93–114.

[46] Searle, A., Vedhara, K., Norman, P., Frost, A. and Harrad, R. 2000. Compliance with eye patching in children and its psychosocial effects: a qualitative application of protection motivation theory. *Psychology, Health & Medicine*. 5, 1 (2000), 43–54.

[47] Seguin, C.A. and Ambrosio, A. 2002. Multicultural vignettes for teacher preparation. *Multicultural Perspectives*. 4, 4 (2002), 10–16.

[48] Sheeran, P. 2002. Intention — Behavior Relations : A Conceptual and Empirical Review. *European Review of Social Psychology*. 12, 1 (2002), 1–36.

[49] Siponen, M., Mahmood, M.A. and Pahnila, S. 2014. Employees' adherence to information security policies: An exploratory field study. *Information & Management*. 51, 2 (Dec. 2014), 217–224.

[50] Skinner, B. and Ferster, C. 1997. *Schedules of reinforcement*. Massachusetts: Copley Publishing Group.

[51] Sriramachandramurthy, R., Balasubramanian, S.K. and Hodis, M.A. 2009. Spyware and adware: how do internet users defend themselves? *American Journal of Business*. 24, 2 (2009), 41–52.

[52] Srivastava, A. and Thomson, S. 2009. Framework analysis: a qualitative methodology for applied policy research. *Joaag*. (2009).

[53] Stanton, J., Stam, K., Mastrangelo, P. and Jolton, J. 2005. Analysis of end user security behaviors. *Computers & Security*. 24, 2 (Mar. 2005), 124–133.

[54] Thomson, K., Solms, R. von and Louw, L. 2006. Cultivating an organizational information security culture. *Computer Fraud & Security*. (2006).

[55] Vance, A., Siponen, M. and Pahnila, S. 2012. Motivating IS security compliance: Insights from Habit and Protection Motivation Theory. *Information & Management*. 49, 3-4 (May 2012), 190–198.

[56] Vries, H. de, Dijkstra, M. and Kuhlman, P. 1988. Self-efficacy: the third factor besides attitude and subjective norm as a predictor of behavioural intentions. *Health education research*. (1988).

[57] Wason, K., Polonsky, M. and Hyman, M. 2002. Designing vignette studies in marketing. *Australasian Marketing Journal (AMJ)*. 10, 3 (2002), 41–58.

[58] Workman, M., Bommer, W. and Straub, D. 2008. Security lapses and the omission of information security measures: A threat control model and empirical test. *Computers in Human Behavior*. 24, 6 (Sep. 2008), 2799–2816.

[59] Yao, M.Z. and Linz, D.G. 2008. Predicting self-protections of online privacy. *Cyberpsychology & behavior : the impact of the Internet, multimedia and virtual reality on behavior and society*. 11, 5 (Oct. 2008), 615–7.

7. APPENDICES

7.1 Appendix A- Security behavioral categories and example vignettes

Category	*Description*	*Vignette*
Remote working	Actions for working on mobile devices and in external locations	Miles is a merchandiser for a large menswear store and constantly travels to other stores within the local area. One of the benefits of Miles's job is that he is given a company laptop as he is constantly mobile. Miles has a 15 year old daughter, who he lets use his laptop when he doesn't need it as his laptop is of much better quality than his daughter's PC. Mile's daughter uses the laptop for playing computer games, however she often disables the anti-virus software as it slows down her favorite game.
Removable media	Portable storage devices that can be connected to and removed from a computer (e.g. USB sticks)	Mary works as a Lecturer at the local university, she has an important presentation at a national conference in London, 300 miles away from her home. Due to the long train journey and therefore intermittent internet connection, Mary decides to store her work on a USB stick so that she can continue working on the train from her laptop. The documents stored on the device include assignment results, presentation notes and an excel document listing the names and addresses of the students enrolled on one of her classes. After exiting the train and arriving at the conference location, she realizes that she has lost the USB stick.
User access management	How access controls are allocated and managed e.g. passwords	Matthew is staying late to work on an important assignment which is due the next day, Matthew has limited security access to confidential information stored on a company password-protected server but he requires a certain document to finish this report. Normally, Matthew would have to get authorization from the information owner who accesses the file for Matthew but instead the owner gave Matthew their password to access the server so that he could do it himself.
Prevention of malicious software	Actions to prevent malicious software	The updates for the anti-virus on Laura's work computer are controlled by her organization; however she has to occasionally restart her computer to allow the updates to install. Laura is regularly prompted by the anti-virus software to restart the computer however Laura keeps postponing this task as she is too busy to wait for her computer to restart and for her to re-open the documents she was working on.
Breaches of security	Steps for recovering and reporting security incidences	Chris is about to go on a two weeks holiday from work and on his last day his computer starts acting strangely. For example, the cursor on his computer screen would start to move around on its own and new files would appear on his desktop. Chris only realizes that something peculiar is going on later that day, rather than reporting it to IT, he decides to switch off his computer and deal with the issue on his return.
Physical security	Strategies to physically protect infrastructures, information and information resources	Kimberley works as a secretary in a busy open plan office. Kimberley's work computer has access to a number of highly confidential documents. She is normally stationed at her desk however at lunch she leaves to have her break in the staff room. During this time, Kimberley leaves her computer unlocked.
Information control	Responsibility in protection, storage and processing of information	Lee is disposing of old records which contain sensitive information about clients. His office has two bins for disposing of waste: one for confidential waste and the other for general waste. The confidential waste bin is full so Lee puts the old records in the general waste bin.
Software & Systems	Software and system acquisition, installation and maintenance	Anna requires the latest photo editing software for one of her work tasks, the department has no budget to purchase any new software, however Anna knows a website where she can download an unofficial version of the software. Her work computer allows Anna to download and install it.
Acceptable usage	Appropriate usage of information systems, email and the internet	Beth is a call centre employee and during her work breaks she uses her work computer for personal use. She has just booked a holiday to Tenerife which required her to enter her personal information and credit card details.
Continuity planning	Outlines prevention and recovery from internal and external threats	Michelle's work computer is run by Windows Vista, however she prefers to use her own personal laptop which has Windows 8 installed as its operating system. She brings her laptop into work on a daily basis and does all her work tasks on her laptop. However, Michelle does not back up the data that is stored on her personal laptop.
Compliance with legislation	Compliance to legislation acts such as the Data Protection Act (1998)	Sam is a medical doctor and part of this job role requires him to write notes about patients during his sessions which contain sensitive and personal information that is covered under the DPA (1998). Sam often leaves his notes on his desk in his office. Whilst Sam has an office to himself, other staff such as the cleaners can gain access when required.

7.2 Appendix B: Interview guide

Interview opening:

- Focus of session explained to participant
- Participant provided with an information sheet and informed consent granted from participant
- Emphasize that participants responses will not be shared with their management/company

Participant to complete demographic questionnaire

For each topic area for the policy categories:

- Provide description of category (e.g. for user access management - *Businesses have a number of computer systems to store and process data which employees use. Users have to identify themselves with a user ID and a password to gain access. Employees may have restrictions on their user access to both computer and information*)
- Present participant with vignette
- Ask participant to imagine, drawing on his or her own experience, how they would react in that scenario
- *Optional questions*
    - What advice would you give? / What should they (the character) be doing to protect themselves?

<Researcher to then go back to the topic area>

- Within your workplace, how do you maintain security when/with <topic area>
- Which security behaviors do you perform? / How do you ensure data security?
- What security behaviors do you not perform? / What do you find difficult to do?

For behaviors discussed by participants, the following elicitation questions were used

Determinant	Example elicitation questions
Self-efficacy	If you want to perform these behaviors, how certain are you that you can?
Experiential Attitude	What do you like/dislike about these behaviors?
Instrumental Attitude	What are the advantages and disadvantages of performing these behaviors?
Social pressures	Who would encourage/ discourage you to perform these behaviors?
Response efficacy	How effective do you think these behaviors are in reducing threats and why?
Response cost	What are the costs in terms of monetary, time and effort in performing these behaviors?
Perceived susceptibility	How vulnerable to a threat are you by not performing these behaviors?
Perceived severity	What are the potential consequences of not performing these behaviors?

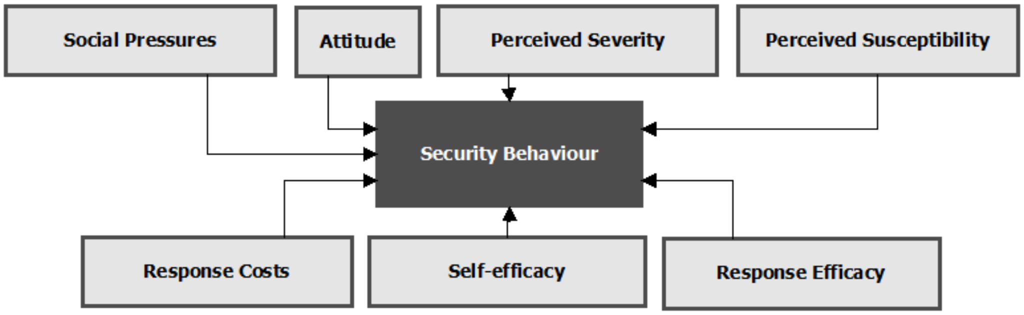
Closing questions

- Anything else that you feel you contribute to security that hasn't been discussed?
- What are the top three security behaviors you think are most important?

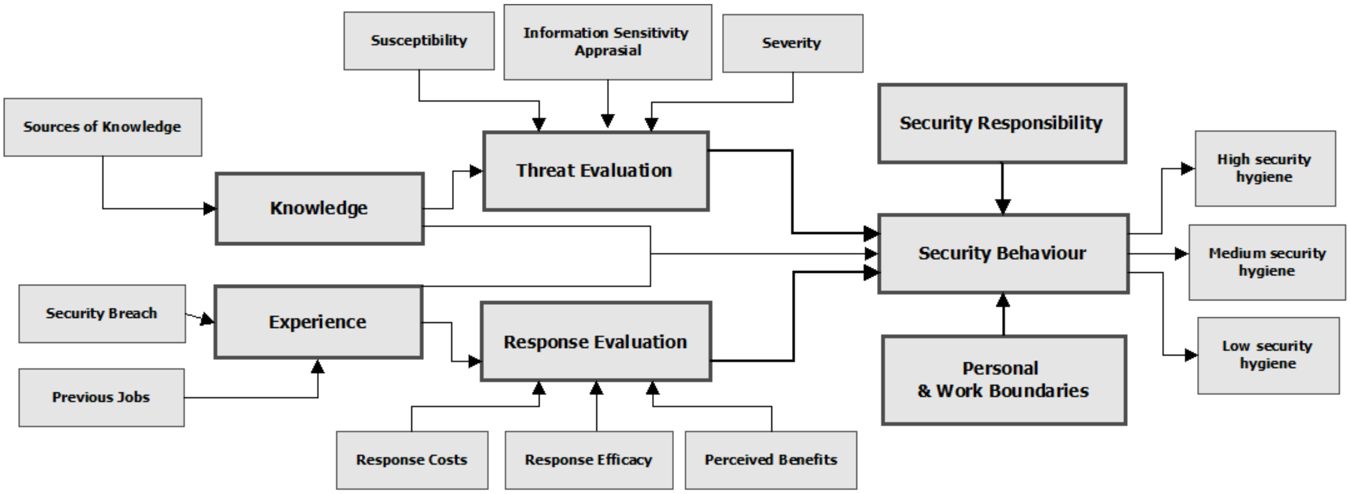
<Participant provided with debrief sheet and thanked for their participation>

7.3 Appendix C: Initial and Final framework

7.3.1 Initial framework based on literature

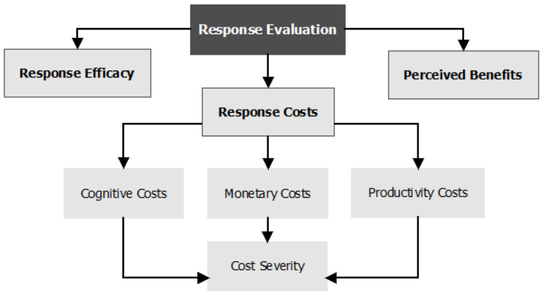


7.3.1.1 Final data-driven framework from framework analysis

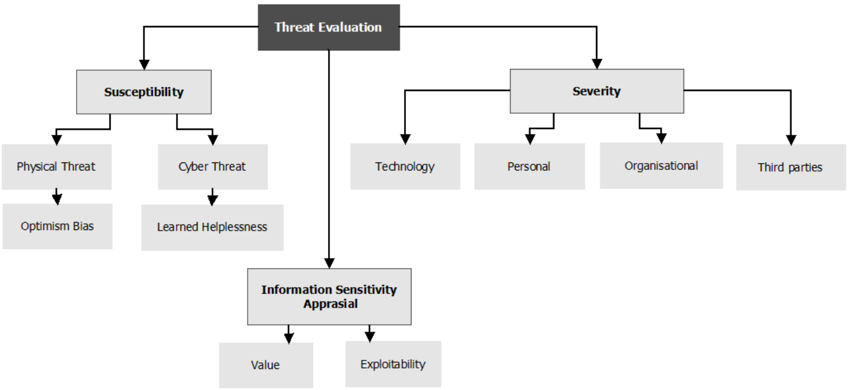


7.4 Appendix D: Theme Visualizations

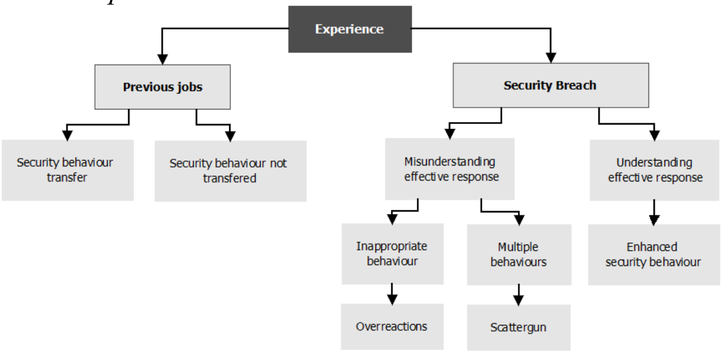
7.4.1 Response Evaluation



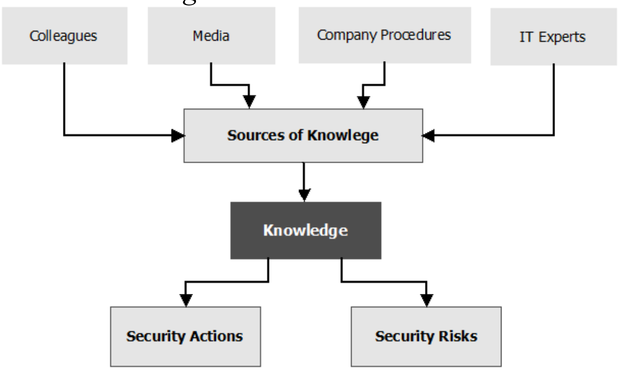
7.4.2 Threat Evaluation



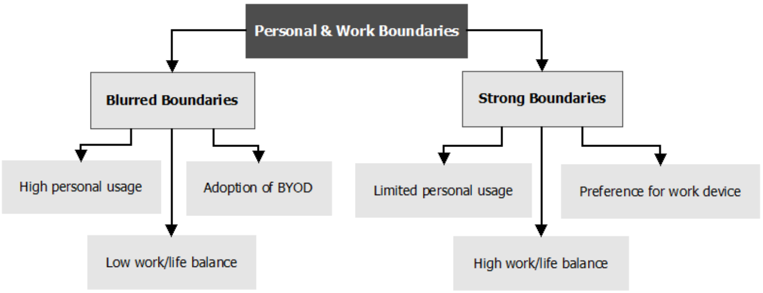
7.4.3 Experience


Experience
Previous jobs
Security Breach
Security behaviour transfer
Security behaviour not transfered
Misunderstanding effective response
Understanding effective response
Inappropriate behaviour
Multiple behaviours
Enhanced security behaviour
Overreactions
Scattergun

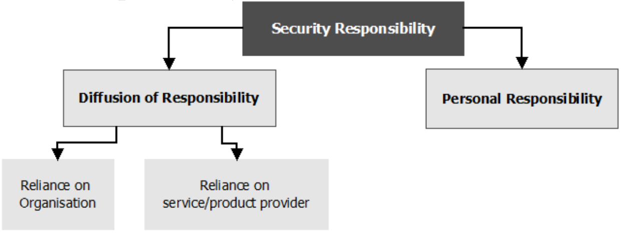
7.4.4 Knowledge



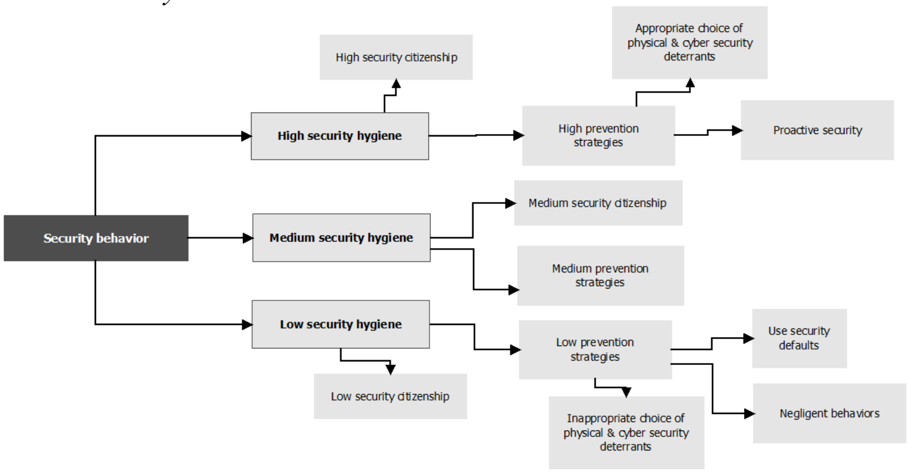
7.4.5 Personal and Work Boundaries



7.4.6 Responsibility



7.4.7 Security behavior


Appropriate choice of physical & cyber security deterrants
High security citizenship
High security hygiene
High prevention strategies
Proactive security
Security behavior
Medium security citizenship
Medium security hygiene
Medium prevention strategies
Low security hygiene
Low prevention strategies
Use security defaults
Low security citizenship
Negligent behaviors
Inappropriate choice of physical & cyber security deterrants

“I Added ‘!’ at the End to Make It Secure”: Observing Password Creation in the Lab

Blase Ur, Fumiko Noma, Jonathan Bees, Sean M. Segreti, Richard Shay,
Lujo Bauer, Nicolas Christin, Lorrie Faith Cranor
Carnegie Mellon University
{bur, fnoma, jbees, ssegreti, rshay, lbauer, nicolasc, lorrie}@cmu.edu

ABSTRACT

Users often make passwords that are easy for attackers to guess. Prior studies have documented features that lead to easily guessed passwords, but have not probed why users craft weak passwords. To understand the genesis of common password patterns and uncover average users’ misconceptions about password strength, we conducted a qualitative interview study. In our lab, 49 participants each created passwords for fictitious banking, email, and news website accounts while thinking aloud. We then interviewed them about their general strategies and inspirations. Most participants had a well-defined process for creating passwords. In some cases, participants consciously made weak passwords. In other cases, however, weak passwords resulted from misconceptions, such as the belief that adding “!” to the end of a password instantly makes it secure or that words that are difficult to spell are more secure than easy-to-spell words. Participants commonly anticipated only very targeted attacks, believing that using a birthday or name is secure if those data are not on Facebook. In contrast, some participants made secure passwords using unpredictable phrases or non-standard capitalization. Based on our data, we identify aspects of password creation ripe for improved guidance or automated intervention.

1. INTRODUCTION

Despite decades of research investigating passwords, many users still make passwords that are easy for attackers to guess [9, 22, 35, 62]. Predictable passwords continue to cause problems, as evidenced by the recent release of celebrities’ private photos obtained in part through a password-guessing attack on Apple’s iCloud [11, 37]. While most everyone would prefer a world without the burden of remembering a portfolio of passwords [18, 53], passwords are familiar, easy to implement, and do not require that users carry anything. As a result, passwords are unlikely to disappear entirely in the near future [7]. Although expecting users to remember complex and distinct passwords for dozens of accounts is absurd, single-sign-on systems, software password managers, and biometrics [4] promise to reduce this burden. Passwords also remain useful for frequently accessed accounts, as master passwords for password managers, and as an integral part of two-factor authentication.

Researchers have identified common, predictable choices that result in easy-to-guess passwords [9, 22, 35, 62]. While some users may be making informed cost-benefit analyses and creating weak passwords for low-value accounts, other users may have misconceptions about what makes a good password. Existing security advice [10, 28, 36, 49, 73] and real-time feedback [1, 12, 15, 32, 59, 63] may be insufficient in disabusing users of these misconceptions.

To understand where users fall short in their attempts to create passwords, we conducted the first qualitative laboratory study of the process of password creation. Whereas analyses of large sets of passwords can reveal common patterns, a qualitative study is better suited to discern precisely why these patterns appear because researchers can probe the rationale behind behaviors through context-based follow-up questions. Prior lab studies of passwords have focused on password management [2, 20, 27, 29, 52, 55], how users cope with password-composition requirements [45, 66], novel password systems [19], and the external validity of password studies [16]. In this paper, we report on the first lab study focusing exclusively on how users craft and compose passwords step-by-step.

We conducted in-person lab sessions with 49 participants, each of whom created passwords for a banking website, news website, and email account in a think-aloud, role-playing scenario. We also explored participants’ general strategies and inspirations. This enabled us to pinpoint participants’ misconceptions and identify strategies that seem both usable and secure against large-scale guessing attacks, such as an offline attack [6, 31, 70].

We found that most participants had a well-defined process for creating passwords. Commonly, participants either had a base word or a systematic human “algorithm” for generating passwords based on the site. While many strategies led to predictable passwords, some participants successfully mixed unrelated words or crafted unique phrases to create more secure passwords. Some participants desired passwords of different security levels across the three websites, yet nearly half did not, indicating that some people may routinely waste effort creating and remembering strong passwords for low-value accounts. Participants struggled to create passwords that matched their desired security levels, sometimes creating strong passwords that they inteded to be weak, and vice versa.

Participants were concerned primarily with targeted attacks on their passwords, rather than large-scale, automated attacks. As a result, some participants believed the (common) name of their pets or birthdays would be strong passwords because they had not posted that information on their Facebook page, not accounting for the types of automated guessing attacks often seen in the wild when sites like LinkedIn [9], eHarmony [57], Gawker [5], or Adobe [43] had their password databases compromised.

We identified numerous other security misconceptions. Most participants knew that dictionary words make bad passwords, yet

Symposium on Usable Privacy and Security (SOUPS) 2015, July 22–24, 2015, Ottawa, Canada.

others incorrectly expected common keyboard patterns (e.g., "qwerty" or "1qaz2wsx") to be a secure replacement. Some participants had learned that phrases make secure passwords, yet chose obvious phrases (e.g., "ilove*SiteName*"). Commonly, participants believed that adding a digit or symbol to the end of a password would make it secure, whereas such an action is very predictable. Other participants conflated difficulty for users with difficulty for attackers, such as thinking that words that are hard to spell are secure.

In contrast, some participants employed strategies that resulted in strong passwords. These strategies included combining unrelated words or developing unique phrases. Whereas many participants insecurely capitalized the first letter of their password in deference to the rules of grammar, others employed non-standard capitalization to make far stronger passwords. Whereas some participants ill-advisedly used the website name as a core part of their password, others used songs and concepts they associate with the site. These related concepts would be far less obvious to attackers.

Many misconceptions we identified might derive from misinterpretations of well-meaning security advice. For example, some participants seem to have misconstrued the idea that "a strong password should contain letters, digits, and symbols" as the false statement "*any* password that contains letters, digits, and symbols is secure." Similarly, the admonition to avoid dictionary words in passwords does not mention birthdays or keyboard patterns, which some participants incorrectly believed to be secure. Building on our results, we discuss aspects of abstract password guidance and data-driven tools that could help users create better passwords by avoiding the misconceptions we observed in this study.

We next discuss related work in Section 2. Then, we present our methodology in Section 3. We present our findings in Section 4, discuss their implications in Section 5, and conclude in Section 6.

2. RELATED WORK

Password-based authentication remains ubiquitous for online accounts [7]. Even if passwords are replaced with devices that do not rely on human memory [41, 53], the deployment of such systems and subsequent decline of passwords would be gradual. Even recent multi-step authentication systems, such as two-factor authentication systems from Google [26] and Microsoft [40], tend to retain passwords as one part of the approach.

The literature on passwords is vast; below, we briefly discuss the most relevant prior work. However, prior studies of password characteristics focus post-facto on passwords that have already been created, in contrast to our qualitative focus on passwords in the process of being created. Prior studies with a similar qualitative approach have generally examined complementary topics, such as password management and novel password systems.

2.1 Analyses of Password Characteristics

Many password databases have been leaked in recent years [5, 9, 43, 57]. Both the popular press and academics have mined these password corpora to identify common passwords characteristics. For example, popular media reported on the leaked set of RockYou passwords, noting the most common password was "123456" [62]. Researchers found that RockYou passwords commonly included digit sequences, names, and phrases about love [69].

Researchers have also focused on the semantic content of passwords [60, 64]. Historically, researchers have found that some of the most prevalent semantic themes in passwords include names and locations [39], as well as dates and years [65]. Researchers have also noted love, animals, and money as common semantic themes [64]. While two-word Amazon payphrases are not as predictable as general English text, common themes include music, television, and sports [8]. Combining multiple words and substituting characters are also common strategies [30].

Other studies have entailed collecting passwords created under controlled conditions in online studies. For example, our group has used this technique to study password-composition policies [31, 38, 51] and password-strength meters [59]. While controlled experiments can be used to collect some behavioral metrics, our qualitative methods allow us to collect far more explanatory data.

We also aim to understand password characteristics. However, qualitatively observing password creation as it happens, rather than after the fact, lets us not just learn *what* users do, but also *why*.

2.2 Laboratory Studies

Other laboratory studies have focused on complementary aspects of the password ecosystem. These aspects have included password-management practices [2, 20, 27, 29, 52, 55] and how users respond to password-creation requirements [45, 66].

Researchers have studied how users recall multiple passwords. Their participants learned six passwords each, including text passwords and graphical passwords. Participants were asked to authenticate two weeks later [13]. Researchers have also explored automatically increasing password strength. Participants created passwords in the lab, and the system added random characters, which participants could shuffle until arriving at a configuration they liked. The authors found that inserting two random characters increased security, yet adding more characters hurt usability [19].

More recently, researchers interviewed 27 participants about their strategies for password management and usage. Participants had an average of 27 accounts and five passwords. They often made tradeoffs between following password advice and expending too much effort [55]. While our methods resemble those of prior lab studies, we are the first to focus on how users create passwords.

3. METHODOLOGY

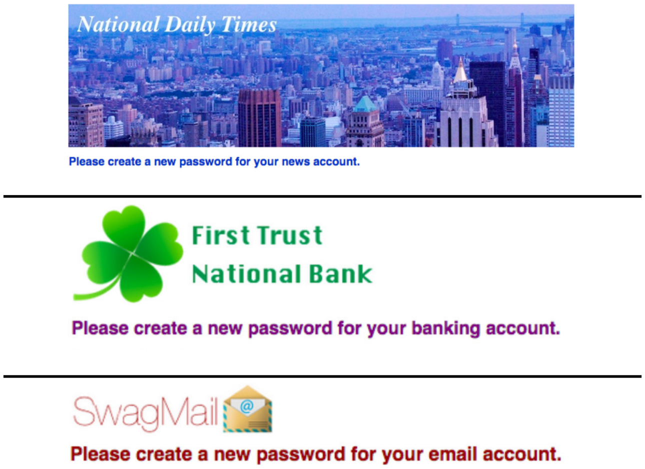
To uncover precisely how average users construct passwords, we conducted face-to-face interviews in our lab. Participants created passwords for three different types of accounts we hypothesized would elicit different security levels. Each participant created all three passwords under a single password-composition policy that we randomly assigned from three possibilities. Participants engaged in a think-aloud process while creating each password and answered follow-up questions about their processes, decisions, and general habits related to password creation. The study was approved by the Carnegie Mellon University IRB.

3.1 Recruitment and Logistics

We recruited participants for a study on passwords through ads on our local Craigslist and flyers at public places in and around Carnegie Mellon University's Pittsburgh campus. Each session was designed to last between 45 minutes and one hour. We compensated participants $25 for the session. The study took place in a room in our laboratory with either one or two moderators. Participants used a laptop from our lab for the study. We audio-recorded the interviews and subsequently transcribed them.

3.2 Study Protocol

We began the study with demographics questions. We then asked participants to create passwords for three websites while thinking aloud. Next, we asked participants about their general password-creation approach and strategies. Finally, we had participants recall each of their three passwords. The text below provides more detail about each step, and the appendix contains the full interview script.



Please create a new password for your news account.

Please create a new password for your banking account.

Please create a new password for your email account.

Figure 1: The design of the news (top), banking (middle), and email (bottom) sites for which participants made passwords.

Our demographic questions included age, gender, and occupation. We also asked about familiarity with different computer devices and Internet usage in order to understand the context in which participants created and recalled passwords. In order to introduce participants to the technique of thinking aloud, we next had them perform a warm-up activity in which they thought aloud while crafting a slogan for a bumper sticker.

We then asked participants to create passwords on three different websites, which we assumed would be of different value to participants. These were mock-up websites that we created for the purpose of this study. The three sites, presented in randomized order in the study, were a news website ("National Daily Times"), a banking website ("First Trust National Bank"), and an online email website ("SwagMail"). Figure 1 shows each site's visual design.

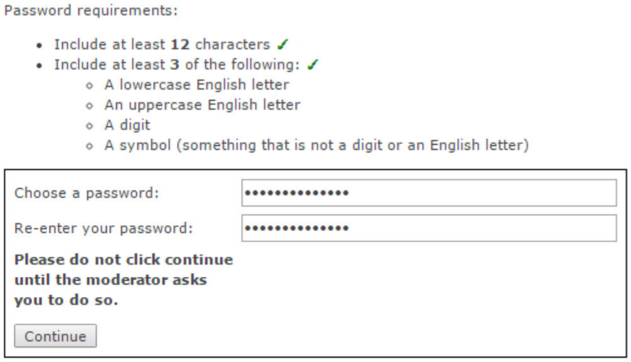
We hypothesized that participants would view the password for the news website as having minimal value, whereas the banking and email account passwords would be of higher value. That is, participants would find those accounts more important to protect. Because participants each created three passwords, we could examine the passwords' similarity. Previous research documented that users often reuse passwords verbatim or with minor, predictable modifications [14, 17, 20, 72].

We asked participants to role-play and "pretend that [they] are actually creating new passwords to sign up for new services" and act as if they will "need to use those passwords again to log in to the account [they] sign up for." Furthermore, so that we could understand precisely where in the process of password creation participants came up with different ideas, as well as in what order, we had participants think aloud when creating their password.

Each participant created passwords for all three accounts under a single password-composition policy assigned round-robin from the following three possibilities:

- **1class6**: passwords must include at least 6 characters;
- **2class8**: passwords must include at least 8 characters, among which are at least 2 of the following: a lowercase letter, an uppercase letter, a digit, a symbol;
- **3class12**: passwords must include at least 12 characters, among which are at least 3 of the following: a lowercase letter, an uppercase letter, a digit, a symbol.

As participants met each requirement, a checkmark appeared next to the requirement, as shown in Figure 2. Participants needed to re-enter their password correctly before proceeding. We chose the 1class6 and 2class8 policies to represent minimal and typical password-composition policies, respectively. We chose 3class12 as a policy that has relatively complex requirements, yet prior research studies have found to be reasonably usable [51]. We expect policies that require longer passwords to see increasing adoption in the real world given the vulnerability of passwords containing eight or fewer characters [24, 54]. We chose to have participants create all three passwords under a single composition policy because we were more interested in how a single participant's behavior differed across sites of potentially different value, as opposed to how a participant's behavior changed across password-composition policies.



Password requirements:

- Include at least **12** characters ✓
- Include at least **3** of the following: ✓
  - A lowercase English letter
  - An uppercase English letter
  - A digit
  - A symbol (something that is not a digit or an English letter)

Choose a password: ••••••••••••••

Re-enter your password: ••••••••••••••

Please do not click continue until the moderator asks you to do so.

Figure 2: As participants created a password, checkmarks indicated which requirements they had completed. The password appeared as asterisks.

We then asked participants about their general strategies for creating passwords and whether the strategies they employed in the study resembled their usual behavior. We excluded from further analysis behaviors they said were atypical. We also asked whether and how they make modifications if they reuse a password, and whether an account of theirs had ever been compromised.

The final part of our study tested password recall. First, to distract participants so that they would think about something other than their passwords for a few minutes, we asked participants to count backward from 100 in increments of seven. Then, we asked participants to log on using each of their three study passwords. We gave each participant up to five attempts to do so, simulating the rate-limiting that many websites use to prevent online attacks.

3.3 Analysis of Password Security

To inform our qualitative analyses of password-creation behaviors, we needed an objective metric of password security. We therefore measured each password's guessability, or how quickly an attacker would guess that password in a large-scale guessing attack [6, 31, 70], using the software tool Hashcat [54]. This tool is widely used by attackers [22, 23, 24, 34, 44] and, relative to other guessing approaches, is generally successful at guessing a large fraction of target password sets in the configuration we used [61]. We made 100 trillion (10^{14}) guesses against participants' passwords, which represents about 6 hours of guessing on a single modern GPU (AMD R9 290x) for passwords stored unsalted using the NTLM hash function, 3 weeks for passwords stored using SHA256, and 904 years for passwords stored using SHA512crypt.

Hashcat takes as input a word list and a set of mangling rules, or transformations (e.g., "add a 1 at the end" or "change every A to @") to apply to word list entries. It is impossible to model every attacker or to study all word lists. We thus chose settings and training data that prior work found to represent a reasonable step beyond Hashcat's default configuration [61]. Our word list comprised large sets of leaked passwords and natural-language dictionaries. The passwords were taken from breaches of MySpace [48], RockYou [62], and Yahoo! [21]. We used dictionaries found effective in past studies [31, 70]: all single words in the Google Web corpus [25]; the UNIX dictionary; and a 250,000 word inflection dictionary [50]. The combined set of passwords and dictionaries contained 19.4 million unique entries, ordered by descending frequency. The mangling rules comprised the "generated2" set included with oclHashcat and a Hashcat translation of rules originally released by Trustwave SpiderLabs [58] for the tool John the Ripper.

Although this approach simulates a large-scale guessing attack, it does not simulate an attacker who knows personal details of the user. Therefore, members of our research team manually examined the study passwords alongside participants' think-aloud transcriptions. If the password was derived primarily from a date significant to the participant or the name of a participant's family member or pet, we marked the password as vulnerable to a targeted attack. Similarly, if the password was mostly derived from the name of the website on which the participant was making a password, we marked it as vulnerable to an attack targeted to that site. Automated cracking methods do not natively support these sorts of highly targeted attacks, necessitating this limited manual analysis.

3.4 Qualitative Analysis

Because our objective was to gain a nuanced perspective on how users craft passwords, we relied heavily on qualitative methods. Rather than approach the study with well-defined hypotheses or very targeted research questions, we instead chose to let participants' strategies and misconceptions emerge from the data.

To that end, one member of the research team first tagged each self-contained thought, representing a distinct password-creation strategy or behavior, mentioned by any participant either during the think-aloud portion of password creation or in response to an interview question. For example, one of the tagged thoughts was, "swap the g for a $ because gold is something related to money." We identified 546 thoughts across our 49 participants.

The members of the research team then collaboratively analyzed these thoughts in a process derived from affinity diagramming [3]. The members of the research team began with each of the 546 thoughts, as well as the corresponding password, printed out on an individual piece of paper. We then iteratively grouped these thoughts into distinct clusters, continuously refining, collapsing, and separating clusters. These clusters represented thoughts the team felt related closely to each other. At the end of our full-group session, we had grouped these 546 thoughts into 18 initial clusters, with themes such as using the website itself as inspiration for a password or adding random characters to a password.

While these clusters represented closely related behaviors, they conflated secure and insecure actions. To separate successful strategies from security misconceptions, two members of the research team went back through all quotes in each cluster and discussed whether that particular behavior would be beneficial for security, negatively impact security, or whether the security impact was uncertain. As a result, we split some clusters to distinguish between secure variants of a strategy and those that were likely predictable by attackers, transforming the initial 18 clusters into 25 clusters.

Finally, within each of the 25 clusters, we performed an additional round of affinity diagramming to further disambiguate distinct behaviors from each other. For instance, within the broad category of "use words inspired by the website," we created distinct sub-clusters of "passwords derived from the website name," "words a participant associates with the site," "phrases a participant associates with the site," "songs the participant associates with the site," "people the participant associates with the site," "emotions the participant associates with the site," and "descriptions of the website's visual design/logos." This process resulted in 122 distinct behaviors that we report within the context of the 25 broad themes.

In addition to our formative analysis of strategies for creating passwords, we had more targeted research questions related to how participants approach creating and managing passwords. We based these questions on a combination of prior work and our own expectations and experiences. These targeted questions covered the security levels participants desired for different websites, the reuse of whole passwords or elements thereof [14, 72], the order in which participants would think of different chunks of their password [60], and how participants manage passwords [18, 55]. We tagged each instance of a participant discussing or exhibiting behaviors related to these areas. Using the same group process we used to analyze creation strategies, we again clustered these behaviors.

Throughout the paper, we focus on reporting the theme captured by each cluster of behaviors and providing relevant quotes where illustrative. In a few cases, we report the frequency of different behaviors to provide a better sense of our data; these frequencies are not intended to suggest that any quantitative analyses of our data are appropriate. To protect participants, some of whom might have used their real passwords in the study, we adopt Fahl et al.'s suggestion and report in this paper sanitized passwords that replace potentially personalized information with analogous content [16].

3.5 Limitations

Our study suffers from limitations typical of small-scale, qualitative studies. We used a small sample that is not representative of any larger population. For instance, more participants than average have technical backgrounds. Despite these limitations, qualitative studies offer rich insight into not just *what* users do, but *why*. Password characteristics have been very widely studied post-facto, yet the moment-to-moment decisions of password creation had not previously been studied in such depth.

A lab study can only capture a sliver of the many ways in which people use passwords, limiting ecological validity. For example, we had participants create three passwords in succession, whereas password creation for different sites is often spread out over time. Furthermore, we test password recall during the same lab session it was created, albeit following a distraction task. In contrast, users need to recall passwords very frequently for some accounts, yet infrequently for others. Similarly, some users log into accounts using different devices or using password managers, which we do not test. However, only two of the 49 participants reported that they normally use password managers.

Our participants made passwords for a study, not a real account. As a result, they had little incentive to make the passwords hard to guess or easy to remember. To gauge the generalizability of different types of password studies, Fahl et al. compared students' actual university single-sign-on passwords with passwords the same students created for an online or lab study [16]. They found passwords from lab studies to be acceptable proxies for real passwords.

4. RESULTS

Our participants generally wished to create strong passwords, at least for some accounts; they just did not always know how to do so. Even worse, they sometimes wrongly believed their choices were contributing to a strong password even when these choices were actually making the password more predictable. In this section, we discuss the passwords participants actually made, alongside their considerations and micro-decisions along the way.

Table 1: The average length and number of character classes in unique passwords participants created.

Policy	#	Length (characters)			# Classes			
		Median	Mean	σ	1	2	3	4
1class6	37	10	10.1	3.5	6	12	8	11
2class8	47	9	9.9	2.2	–	7	9	31
3class12	47	13	14.4	3.5	–	–	17	30

We begin by describing our 49 participants in Section 4.1. We then briefly summarize the characteristics and guessability of the passwords they created in Section 4.2. Even though many participants made passwords that exceeded the minimum requirements of their assigned password-composition policy, roughly half of the passwords were vulnerable to an automated guesing attack or to a targeted attack. Next, in Section 4.3, we describe participants' desired security level for each of the three sites for which they were creating passwords. Unfortunately, the value participants assigned to accounts diverges from what a security researcher might expect.

The main contributions of this paper rest in the qualitative analyses we detail in the subsequent sections. In Section 4.4, we explore participants' security considerations, as well as their abstract, broad approaches for generating a password. We found that participants try to create passwords to match their perceived value of different accounts. We also found that some participants reused passwords or base elements verbatim across sites. We highlight general approaches and human algorithms participants used to craft a password. Some approaches, such as generating a unique phrase, appear secure and also memorable to participants. Sadly, other participants unwittingly employed very predictable approaches.

Despite their desire to create secure passwords, many participants struggled to distinguish approaches that increase password security from those that make a password easier to guess. In Section 4.5, we delve into participants' low-level strategies and micro-decisions. Subtle differences often separated choices that increased security from those that made passwords predictable. For example, basing a password on a song or visual image the participant associates with the website for which he or she is creating a password is far better than using a password like "ilove*SiteName*!" Many of participants' misconceptions can be viewed as twisted interpretations of advice about how to create a strong password.

4.1 Participants

We interviewed 49 participants, 21 male and 28 female. Their ages ranged from 19 to 63. Young participants were overrepresented relative to the general population as the mean age was 31 and the median 24. Of the 49 participants, 24 were students, 13 of whom studied a technical discipline like engineering. Of the non-student participants, 16 were employed in a variety of occupations, while the other 9 were currently unemployed or retired. All participants used text passwords regularly and were frequent Internet users. To preserve anonymity, we refer to each participant as PN.

4.2 Password Characteristics and Security

The 49 participants each created 3 passwords, resulting in a data set of 147 passwords, of which 131 were unique. No participant created the same password as any other participant, but 13 participants reused a password verbatim across two or three of the three accounts. When we report password characteristics and guessability in this subsection, we report on unique passwords, counting a password that a participant reused multiple times only once.

Table 2: The number of passwords created under each policy that were vulnerable to a *general* attack of 10^{14} guesses using Hashcat, as well as the number manually identified as vulnerable to a *site-specific* attack using the website name, or a *user-specific* attack. We also present the number that appear *secure* against all three attacks.

		Vulnerable to attacks			
			Site-	User-	
Policy	#	General	specific	specific	Secure
1class6	37	21	0	0	16
2class8	47	19	2	3	23
3class12	47	10	8	3	26

The quantitative metrics we report in this subsection are not intended to suggest generalizability, which would be inappropriate for a small-scale, qualitative study. Instead, we present these numbers to give a broad sense of the passwords our participants created.

Participants often significantly exceeded the requirements specified by their assigned password-composition policy, as shown in Table 1. For example, the median length of a 1class6 password was 10 characters, rather than 6, and 84% of 1class6 passwords included multiple character classes despite the lack of any character-class requirement. Although 2class8 passwords were only required to contain characters from two distinct character classes, 66% of these passwords contained all four character classes.

Across password-composition policies, 38% of the passwords participants created were guessed within 10^{14} guesses in the automated guessing attack using Hashcat. Table 2 gives an overview of how many passwords created under each composition policy were vulnerable to attack. Sanitized examples of passwords vulnerable to this automated guessing are *Tyrone1975* (1class6), *Gandalf*8* (2class8), and *Triptrip1963* (3class12). In contrast, sanitized examples of passwords that were not guessed include *5cupsoftoys* (1class6), *AfNaHiLoco* (2class8), and *7301Poplarblvd$* (3class12). Using lists of common passwords, six passwords were trivially cracked, including three 1class6 passwords (*gabriel*, *password*, and *qwerty*), two 2class8 passwords (*1Qazxsw2* and *Password1!*), and one 3class12 password (*Newspaper123*). None of the other passwords were among the most commonly used passwords [9, 35].

Our automated, large-scale Hashcat attack did not specifically focus on site-specific information, such as the name of the site on which an account was being created. We manually evaluated vulnerability to site-specific attacks, considering a password to be vulnerable if the name (e.g., "First Trust Bank" or "1sttrust") or function of the site (e.g., "email" or "breakingnews") was the majority of the password. We marked ten additional passwords (e.g., *1234SwagMail@* and *nationaldailytimesP@ss2*) as vulnerable.

In addition to general attacks, passwords can also be guessed in attacks targeted to a user's personal information. We manually examined passwords not guessed by Hashcat alongside participants' explanations to determine whether a password would be vulnerable to a user-specific attack. We marked passwords vulnerable if the name of the participant, immediate family member, or pet, or a date or geographic location of well-known significance to the participant, formed the majority of the password. We marked six additional passwords (e.g., structured *Firstname.Lastname715* and *hOMETOWN!123*) as vulnerable.

4.3 Security Level of Each Site

On the assumption that some or all of the participants would create fundamentally different types of passwords based on their desired security for an account, we had participants create passwords for three types of accounts we expected would be of different value: a news site, a banking site, and an email account. We hypothesized that most participants would consider the account on the news site effectively worthless, yet attribute more value to the other accounts.

In stark contrast to our hypothesis, 21 of the 49 participants (43%) considered all three accounts to be of about equal value. Although access to a user's email account can often be used to reset the passwords to his or her other accounts, many participants shared P21's opinion that an "email [account] is not important."

Many of these participants felt the password for the banking account was similarly not much more important than the password for a news site. Although consumer financial protections in the United States would minimize or completely mitigate financial harms of an online banking account compromise, and while additional security features (e.g., security questions) might also help to secure a banking account, few participants explicitly mentioned these factors. P22 was one of the few who did, saying, "Email usually gets the highest security because even if they break into the bank account, the bank often requires you to send something like a special code they sent to your email." Other participants noted that they did not have much money in their bank account and thus did not care about that account. For instance, P48 noted, "As a college student, I don't have a lot of money to worry about." This is despite the fact that identity theft can be ruinous to one's future creditworthiness.

Some of these participants who viewed accounts to be of equal value reused the same passwords verbatim across these sites. Other participants used an identical password-generation technique across these three accounts. P34 was an example of the latter approach, saying he did not "want the same password as with email" for his banking password or the news site. As a result, he cycled through the names of his three brothers, appending "24!" to each to arrive at *Joey24!*, *Johnny24!*, and *Jimmy24!* as his passwords.

Seven participants (14%) felt that their news account was low-value and that their email and banking accounts were of equal (high) value. In some cases, participants reused passwords across accounts they considered to be of the same security level. P44 explained, "I use the same password [as banking] with email because I don't want to remember many passwords." P32 similarly said, "One thing I do a lot is use the same password that are for things about the same security purposes." While such a strategy might be prudent for low-value accounts [18, 42], it may open important accounts to attack. These seven participants felt the news account was worthless. P28, who considered the news site a "junk website," said her approach for "junk websites will be something that's just easy to remember. If it happens to get stolen it won't make that much of a difference."

Another 11 participants (22%) considered the news account and email account low-value, yet felt the banking account was important. As P30 explained, "[Email] is not important to me." Similarly, P41 said, "I don't care about the security of [the news] account," and also said she is "not too concerned about email getting hacked."

The remaining 10 participants (20%) considered all three accounts to be of different value. All thought the news site was lowest value. Eight of the 10 participants considered their banking account more valuable than their email account, while the other two felt their email account was more valuable. P3 said he used "an easier password" for the news account because it "does not have financial" implications, but wanted email "to be a little bit secure."

Many of these participants felt strong pressure to create a secure password. For example, P23 wanted her banking password "to be very secure because if there's any security risk then I would be losing a great deal of money." Similarly, P18 explained, "[Creating a banking password] stresses me out, banking more than even [health] lab results, because I think it's a combination of the fear of identity theft, and draining an account, and relaxing too much, and constantly watching it. I know I want a really strong password. Thinking through how I want to create that is tough."

Sadly, many participants struggled to craft passwords whose actual guessability matched their desired security level. For example, P6 was one of the ten participants who assigned different values to all three accounts. Unfortunately, the only one of her three passwords that was not guessed was her password for the news account, which she intended to be the *least* secure password. Her password for the news account combined dictionary words from two languages with unpredictable capitalization, yet she expected the password to be predictable because it contained dictionary words. In an attempt to craft more secure passwords for the other two accounts, she used permutations of her name and her birthday. Both of these passwords that were intended to be secure were guessed.

Overall, 57% of participants did differentiate across accounts regarding the desired security of their password. Unfortunately, as we detail in the subsequent sections, many of the behaviors these participants thought improved security dramatically had at most a modest impact. The remaining 43% of participants did not differentiate across accounts, potentially resulting in them inefficiently expending their finite memory for passwords [18] on passwords for low-value accounts and thus limiting their ability to remember strong passwords when password strength actually matters.

4.4 General Approach to Password Creation

Of the 49 participants, 43 (88%) said they had a well-defined process for creating passwords that they put into action during the study. In this section, we delve into the approaches we observed.

While password reuse has been studied previously [14, 55], understanding how our participants reused passwords provides essential context. First, we discuss the 13 participants whose general strategy centered on verbatim reuse of a single password. We then discuss the 10 participants who had a base keyword that they reused across sites, making small modifications per password. We also outline other participants' general algorithms for crafting a password, including the order in which they chose different elements to combine into a password. Finally, we briefly mention how password-management strategies [55], impacted participants' expectations and approaches.

4.4.1 Password Reuse

Password reuse is a major threat to password security because the compromise of one account can lead to the subsequent compromise of other accounts for which the user has chosen the same username and password [14]. We found that even when participants expressed a desire to behave securely, they still reused passwords. Only three participants said that, in the abstract, they would never reuse passwords. The other 46 participants said they generally reused passwords. While reusing the same throwaway password for low-value accounts can be an efficient coping strategy, reuse across high-value accounts is risky [18, 42].

In our study, three participants (6%) created a single password and reused it across all three sites because they worried about their ability to remember multiple distinct passwords. Notably, P26, one of the three participants to reuse a single password for all three accounts, had said, "For online banking and things, I try to make my passwords a little more secure," yet reused a password consisting of an obvious keyboard pattern ("1Qazxsw2") for all three sites.

Ten other participants (20%) created only two distinct passwords across the three accounts. Three of these participants had the same password for the email account and banking site, six participants shared a password for the news site and email account, and one participant used the same password for the news site and banking site. These participants usually failed to see this behavior as potentially problematic. For instance, P2 said he saw "no security downside" in using the same password for the news and email accounts.

Some participants knew in the abstract that password reuse is a poor idea, yet did so anyway. Often, participants found remembering distinct passwords to be too difficult. For example, P6 said it is "very difficult to remember all different passwords." Therefore, she has five distinct passwords that she reuses. P19 reused passwords for accounts she does not think are "that important" and "anything that does not have anything to do with my credit card."

Other participants noted that they had never experienced problems due to password reuse. P1 explained, "I know [password reuse] is a terrible idea, but it does not keep me awake at night....I have never seen any negative consequences." Similarly, P9 said she "usually uses the same password for many things," but is not concerned "because [she has] been using the same password for a long time" and had yet to experience a problem. P45 felt he "should be" worried about consequences of password reuse, yet does not worry.

A number of participants said they did not feel password reuse was a problem because the password they reuse is strong. P2 said he reuses passwords "all the time...if the password is a good one." P35 said similarly, "My [reused] password is not easily guessed," while P49 explained, "No one can guess my [reused] password." Unfortunately, if any site on which that password is reused is compromised and the system administrators do not follow industry best practices (i.e., passwords are salted and hashed using a slow hash funtion like bcrypt [46]), these participants may have multiple accounts compromised if the attacker guesses the password in an offline attack. Notably, during the study, two of these three quoted participants made passwords that they believed were strong, yet were guessed in the general attack using Hashcat.

4.4.2 Element Reuse

Although they did not reuse passwords wholesale, ten participants usually had a long substring in common across their passwords, while eight additional participants sometimes used a common string across passwords. While reusing a base element can still result in strong passwords if modifications and additions to the shared elements are non-trivial, predictable modifications to a base element are common [14, 72]. In those cases, if a single password is compromised, the rest will follow quickly.

In an example of element reuse, P27 used the street name from a former address as his starting point. His email, news, and banking passwords in the study were thus *cedarville1*, *cedarville2*, and *cedarville3a*, respectively. He explained, "I would just go one number up....That way, if I'm having a problem remembering [a password], at least I'll have a base and figure it out from there." He said he would just try increasing digits until landing on the correct password. Notably, he desired for his banking password to be more secure than the others, yet did not achieve this goal. He said, "For a bank, for a little more security, I click it up one number...and add something like another letter." His belief that adding a single "a" on the end makes the password more secure seems misguided. Unfortunately, Hashcat rules to append single characters onto the end of the password are common in lists of mangling rules [34, 54, 58].

P10 also began with a common substring, a mnemonic. His passwords were thus *ATdim12nd#*, *ATdim12sw#*, and *ATdim12ft#*. The two letters that varied represented the names of the sites (*N*ational *D*aily Times, *Sw*agmail, and *F*irst *T*rust). P15 adopted a similar strategy, using "1234" as his starting point and appending a variant of the site name. His passwords were *1234Nat'lDailyTimes*, *1234SwagMail@*, and *1234FirstTrustNat'lBank*.

A common misconception was that making minor or incremental additions to common substrings would result in secure passwords. For example, P37 said she did not care about security for the news site or email account. Therefore, she used *Tyrone* and *gabriel* as those passwords, respectively, drawing on names of family members. However, "because security is required for a bank account," she added Tyrone's birth year, resulting in *Tyrone1975* for the banking site. Unfortunately, the common technique of appending a recent year to a password does not make it secure against a trawling, large-scale attack, let alone against an attacker who researches the user's family members. Similarly, P44 uses the same 8-letter base word across all of his passwords (both in the study and in daily life). For the news account he deemed unimportant, he appended "123." Instead, for the two accounts he wanted to be secure, he appended "1974," his birth year, falling into the same trap as P37. Lists of mangling rules used by attackers frequently include rules that append years to entries in the word list [34, 54, 58].

Other participants expected that character substitutions would similarly transform their typical base password into something more secure. For example, P45 took his shared news and email password, *ninjakick44ninjakick!*, removed the repetition, and instead performed character substitions to arrive at *n1nj@k1ck!* as his banking password. He explained, "I want my banking password to be extra secure... I replaced letters with symbols and numbers to make it secure." Unfortunately, such substitutions are very predictable [60] and provide uncertain security benefit [67].

Similarly, P49 turned her shared news and email password *Elephant0215!*, which she expected to be weak, into the banking password *@El3phant4225*, which she expected to be strong. In particular, she performed a predictable character subsitution, yet also replaced her birthday (02/15), which she expected to be "linked to my bank account," with her favorite four-digit sequence. She also changed the ending exclamation point into a leading "@" because it reminds her of Twitter. While neither password was guessed by Hashcat and both are thus at least moderately secure, these types of modifications would not in general reliably transform an otherwise weak password into a strong one.

Similarly, P22 "[does] not worry about security for the news account," so he used "two unrelated words (jungle and salmon) followed by digits" to create *junglesalmon711*. In the spectrum of passwords we observed for this study, "junglesalmon" is actually a relatively strong starting point because of the combinatorics of combining two unrelated words, yet P22 assumed it was weak because it contained dictionary words. To make a banking password, which he hoped would be strong, he prepended "R" and appended "@$" to the string he believed was insecure, expecting the resultant *Rjunglesalmon711@$* to be secure. As with P49, neither password was guessed by Hashcat; both are fairly secure. However, P22's misconception that minor changes to a password he believes is weak can make the password secure is troubling, especially when the capital letters and symbols are in predictable positions [38].

4.4.3 Algorithm for Password Creation

Most participants had an algorithm they always used to make a password. Rather than relying on reuse, these approaches followed conceptual patterns. For example, P3 always used a word, a year, and an emoticon. For low-security accounts, the word was a place she had visited. For high-security accounts, she used a "magic" word because she "want[s] it to be secure." Her resultant passwords were *Croatia2011:-p*, *Patagonia2014:-)*, and *HocusPocus;-)2003*. P7 always used words she associated with the site followed by a single digit and an exclamation point. P19 stuck with a consistent order, "usually one capital, followed by lowercase letters, and a number or symbol."

In some cases, participants unwittingly created some passwords they believed to be strong and others they believed to be weak, yet the supposedly strong passwords were easier to guess. P35 was one such participant. As an English teacher, she liked to use "longer word[s]. It's what teachers expect from students, which is what made me think of it." Her algorithm was to add a digit onto a single long word. Her email and banking passwords, which she expected to be secure, were *Likelihood4* and *Deliberation9*. Both were guessed by Hashcat since "likelihood" and "deliberation" are in attackers' dictionaries, even if not in students' lexicons.

In contrast, she said, "I do not care about the security" for the news site and wanted to write "I journal." Because journaling is associated with newspapers, she said, "I think using a word related to the site would lessen the security." To make the word a little longer, she changed "journal" to "journalistic," resulting in the password *Ijournalistic?8*. This password, intended to be the weakest, was actually the strongest. While the participant associated the phrase "Ijournal" with a news site, this association is far from obvious.

In other cases, participants made passwords of similar strength despite expecting some to be far more secure than others. For example, none of P18's three passwords were guessed by Hashcat, yet she believed her news password was relatively weak and "simple" and her email and banking passwords were "really strong." Her approach to password creation was to pick words of significance to her and to write them in mnemonic form, usually followed by a symbol and some digits. The letters in her news password, *tdVc$567*, stood for "the da Vinci Code," while the letters in her email password, *Tjks&987*, represented the first names of her "siblings [and dog] in birth order." Her banking password, *_EmiLt345*, which she deemed to be the strongest, was based on the name of a "friend who lives out of the country," which is why she felt that password was strongest. However, all three passwords were strong; none were guessed by Hashcat.

Most participants said they developed their password-creation algorithm on their own, but 12 participants (24%) had read articles giving advice about creating secure passwords or attended an organizational security-training class. Unfortunately, both types of participants fell victim to misconceptions about security. P36, for example, had attended security training provided by her university and was taught to use phrases in passwords. As a result, in the study, she decided to use *ilove1sttrust!* as the password for the "First Trust National Bank." Although the participant believed this to be a "secure" password because it contained a phrase, "ilove" is a very common substring in passwords, and the name of the site, even slightly modified, is very predictable. The institutional security training, while correct in intent, fell short in helping P36 create secure passwords.

4.4.4 Order in Which Elements Are Chosen

One of the most common password structures when multiple character classes are required is a series of letters followed by a digit and a symbol [70, 71]. P44's approach was common among our participants: "I always put a capital letter at the beginning and numbers at the end."

Our think-aloud protocol let us unpack the order in which participants chose different elements of their passwords. We analyzed the order in which participants discussed each element of their password during its creation. The vast majority of participants first thought of a word, followed by digits and symbols. Their final passwords reflected this order.

For example, among passwords for banking accounts, which participants frequently deemed the most valuable, 29 participants created passwords containing letters, digits, and symbols. In 27 of these cases (93%), participants first chose the word they would use. In the remaining two cases, the participant first thought of the digit(s) they would use, followed by the word. Seventeen additional passwords contained letters and digits, but not symbols. For fifteen of these passwords (88%), the participant first chose the word; in only two cases did the participant first choose the number.

The consistency with which participants first chose a word to use, followed by digits and symbols, is particularly notable because for 82% of banking passwords, the order in which participants decided on elements is the order those elements appeared in their final passwords. In essence, the password is built from left to right as participants think of elements. Because passwords that begin with a word and end with digits and symbols are most common, one way to induce users to create stronger passwords might be to encourage them to scramble the elements of their password or, even better, to nestle digits and symbols into the middle of the words.

4.4.5 Password Management

Users' password-management strategies [18, 55] are central to their ability to use distinct, complex passwords for each account. We found that 17 of our 49 participants (35%) simply memorize their passwords without writing them down or storing them anywhere. For these participants, the memorability of the password is of paramount concern. In contrast, only two participants (4%) used a third-party password manager (KeyPass and LastPass, respectively), and only 6 participants (12%) used their browser to store passwords. Consistent with what prior research has shown [55], the remaining participants mixed memorization, writing passwords down, and storing passwords in ad-hoc ways on their computer.

A few participants had other considerations that impacted password creation. Two participants regularly reset their passwords, leading them to care greatly about the security of their email account. P29 uses a "family password" for her bank account to enable her parents to access it. As a result, she wants that password to be memorable to her parents, too.

4.5 Strategies and Misconceptions

Finally, we delve into micro-decisions participants made in the course of making a password. We documented micro-decisions through both the think-aloud protocol and the targeted questions in response to participants' behaviors and interview responses. We pay particular attention to participants' misconceptions.

As discussed in Section 3, our qualitative clustering of password-creation strategies enabled us to identify 122 distinct behaviors within the context of 25 broader themes. Table 3 presents the 25 broad themes we identified and how many of our 49 participants exhibited behaviors in each of those themes. As our analysis was purely qualitative, these counts should not be interpreted as generalizable to larger populations or comparable statistically. Instead, we provide these to give the reader a better sense of our data.

As shown in Table 3, we evaluated the security impact of each password-creation strategy based on the overall guessability of the passwords that employed that technique, as well as the frequency with which those techniques appear in passwords leaked in major breaches [21,48,60,62]. In some cases, different applications of the same strategy had very different impacts on security—sometimes beneficial and sometimes detrimental. For many passwords, however, the application of a particular strategy itself did not cause either a substantial increase of decrease in security. We do not explicitly call out these neutral impacts in the table.

Table 3: The categorization of participant strategies that resulted from our qualitative data analysis, along with how many participants (#) exhibited that behavior. Check marks denote that we observed instances of that behavior that made a password substantially *more* or *less* secure.

Categorization	#	More Secure	Less Secure
Choosing words/phrases (Section 4.5.1 and Section 4.5.2)			
Use a phrase	24	✓	✓
Use keyboard pattern for security	5		✓
Use non-English words for security	4	✓	✓
Use address or geographic location	10	✓	✓
Use names of family, friends, or pets	23		✓
Use information not on social media	5	✓	✓
Use uncommon dictionary word	11	✓	✓
Use word(s) inspired by the website	19	✓	✓
Base element is something participant likes	13	✓	✓
Base password on own workplace	5	✓	✓
Base password on pop-culture reference	11		✓
Password structure (Section 4.5.3)			
Create mnemonic	6	✓	✓
Intentionally non-standard capitalization	7	✓	
Capitalize first letters, following grammar	24		✓
Intersperse different character classes	6	✓	
Add string of "random" characters	9	✓	✓
Digits and symbols (Section 4.5.4)			
Replace letters with digits / symbols	14	✓	✓
Use information from bank card or ID card	3	✓	
Use date / year significant to self / family	17		✓
Other meaningful digits / symbols	24	✓	✓
Add symbol (usually "!") at end	18		✓
Expect symbol (e.g., "&") is hard to guess	16		✓
Meeting composition requirements (Section 4.5.5)			
Aim to make password longer than required	13	✓	
Explicitly include extra character classes	7	✓	
Feel any password meeting policy is secure	6		✓

The 25 broad themes we identified conceptually fit into four even broader approach areas. The first area, detailed further in Sections 4.5.1 and 4.5.2, centers on how participants chose the primary content, often semantically significant, that served as the foundation of the password. Usually after choosing this foundational content, participants imbued the password with additional structure (Section 4.5.3) through capitalization, mixing character classes, and added "randomness." Participants had varied strategies for using digits and symbols (Section 4.5.4) and meeting, or intentionally exceeding, the requirements of a password-composition policy (Section 4.5.5).

4.5.1 Choosing Words and Phrases

Choosing words to form the base element of a password was a crucial step in password creation. Our participants most often chose words based on personal topics (e.g., addresses, names, birthdays), associations to the site, their hobbies, nearby items, past events, keyboard patterns, work, and religion. As P23 explained, "I like to come up with words that mean something to me, something that I like, like the name of my favorite author, or a candy that I like to eat." Other participants used nearby objects for inspiration. For instance, P28 built her password around the product number of a camera on a shelf in our lab. Some participants correctly knew to avoid using their employer's name, their own name, their own birthday, or a single dictionary word. However, beyond these correct conceptions of high-level topics to avoid, participants fell victim to a number of misconceptions.

Many of these misconceptions were related to participants not understanding the automated nature of password-guessing attacks. They knew to avoid personal items about themselves, yet thought names and dates related to family members were fine. P6 structured her password around her name, yet placed a birthday (MMDD) in between her first and last name. She expected "a malicious person will try my birthday and my name, so I will not use my birthday...I will use...my pet dog's birthday." Unfortunately, automated attackers often try all possible birthdays rather than targeting a particular user's birthday [34,58,65]. Similarly, P7 built her password around the name of her dog, "Goldie." She expected "hackers cannot guess [it] because I have no pictures of him on my Facebook account." Although her dog's name is not on Facebook, it is a common pet name, making it a very likely target for attackers [60,64].

Misunderstandings about attackers also impacted the characteristics participants expected to be secure. For instance, P2 based his password around the Mahavishnu Orchestra, noting that he expected attackers would not be able to guess his password because "this band name is hard to spell." Because automated attackers use wordlists, words being hard to spell makes no difference for the attacker, only for the user. In crafting the password *purplep@nts*, P11 wanted to pick words that were secure, and she picked purple as "a color that is not often used, unlike red or white." Unfortunately, purple is just as common on word lists as are red and white.

Many participants had heard to avoid dictionary words in making passwords. As a result, P26 used the keyboard pattern *1Qazxsw2* as a password she intended to be secure. She said, "For online banking, I try to make my passwords a little more secure, so I like to follow a pattern on a keyboard." Similarly, P23 based the banking password she hoped would be secure around a keyboard pattern, which she mistakenly considered to be "random letters." Keyboard patterns are an easy target for automated attacks [54,60,64].

In contrast, other participants had developed approaches to word selection that resulted in much more secure passwords. For the banking website, P39 wanted to base his password around a song that he associates with money, and "the first song that comes to mind is Gold Digger. The phrase would be, 'I ain't saying she's a gold digger.'" He transformed this phrase into a mnemonic and added three random characters. P4 similarly based her password, *$0.02CentShow*, on a music album that she associates with money. She chose to spell "2 cent" as "$0.02Cent" to be harder to guess. P28 created the complex password *LCiinf3-n*, explaining, "I usually create a sentence and take all the first letters," which is often considered good advice as long as the sentence is unique [36,49].

Other participants sadly undervalued the importance of choosing unique phrases when constructing a password. P17 chose the common aphorism "be the change" as the basis for one password, yet believed it to be secure because "someone wouldn't think [the phrase] necessarily applies to me." Similarly, P46 used a mnemonic of the famous opening line from *A Tale of Two Cities* ("It was the best of times, it was the worst of times") as the basis for her password. In contrast, P28 securely created a completely unique phrase to describe what was happening while she was creating a password.

A few participants crafted passwords that combined words from multiple languages, which may or may not be a secure strategy based on the languages and words chosen. For example, P6 used a Hawaiian word in her password, expecting its juxtaposition with English to be unpredictable. Other participants created strong passwords by combining unrelated words. P40 unintentionally created a reasonably secure password when he wanted to make a "short and simple" password for his low-value news account. He combined two unrelated words, "tossed in a few symbols," and appended "whooo" in crafting *Squ@shC2ndywhooo*.

4.5.2 Deriving Passwords from a Website

Using the website or service for which they were creating an account as inspiration was participants' second most common strategy after using personal names and dates. We found a sharp dichotomy of secure and very insecure instances of this strategy.

Many participants simply used the name of the site as a core component of their password, making such passwords easy targets for site-specific attacks. For instance, P36 had "heard that, instead of using words and numbers, using a phrase is more secure," so she created the password *Ilove1sttrust!* for the First Trust National Bank. Unfortunately, the name of the site is extremely predictable. Other participants used predictable word associations. For instance, P33 created the password *+Money369*. She used the word money "because it is a bank," and she used an increasing pattern on the number pad to represent an account balance she hoped would also be increasing.

In contrast, other participants used much more distant word associations in crafting far more secure passwords. These participants, like P38, avoided "the name of the service or the type of service because that would be too easy to guess." Many of these participants were inspired not just by the purpose of the site, but also the site's visual design (see Figure 1 in Section 3). For example, P13 said, "I saw the website logo picture and found a brown building on the left. So I used 'left' 'brown' as keywords followed by my favorite number and a symbol that looks like a building" in crafting *LEFTbrown8!* as his news password. Similarly, the news site logo reminded P32 of New York, which itself reminded her of "108." As a result, she created the password *newyorkONE008*, in which she added a 0 and mixed digits and capital letters for security. P39 used creative capitalization and uncommon punctuation to turn the lyrics from the Queen song "The Invisible Man" into a secure password. He chose this song as the basis for a secure email password "because I want the password to be invisible," using a distant association with the goal of the password as inspiration.

4.5.3 Capitalization, Punctuation, and Structure

As they decided how to integrate capital letters and other structures, many participants predictably capitalized the first letter of the password and added a single punctuation mark at the end. While in some cases participants said they did so out of laziness or simply because it is easy to remember, others said that years of schooling had inculcated the idea that capital letters come at the beginning and punctuation comes at the end. For example, P36 ended her password with an exclamation point "because that is how a sentence ends." These participants did not recognize that following the rules of grammar is detrimental to password security [47, 54, 60].

In contrast to the many participants who said they usually capitalize the first letter of passwords, some participants used far more creative approaches. For example, P13 crafted the password *8AX-Ewater<* based on the two items he associated with the job he had 8 years ago. He explained, "The security of an email account is important to me, so I capitalize some words and include a symbol that looks like an axe." The less predictable capitalization and less common symbol made his password stronger. Similarly, P31 knew to structure a password unpredictably, so he capitalized the "E" in *baldErdash49* "to randomize the password for security."

4.5.4 Use of Digits and Symbols

The most common, and most troubling, misconception we observed around the use of digits and symbols is that their inclusion automatically makes a password secure. We hypothesize this misunderstanding stems from advice that strong paswords contain digits and symbols (to increase the space of potential passwords) being misinterpreted as something akin to the assertion that including digits and symbols makes a password secure. Participants with this misunderstanding were well intentioned. For instance, P6 said, "I want to prevent others from predicting passwords, so I want to use all four types of [characters] for my password." Unfortunately, she did so in very predictable ways, with a capital letter at the beginning and a digit and a symbol at the end.

Many participants thought simply adding a symbol at the end of the password made it secure. As P45 said, "I added '!' at the end to make it secure." P34 felt that "usually numbers and a symbol will make the password strong." Therefore, he appended "24!" to each password, which otherwise were just the names of his three siblings. Users must be disabused of the notion that digits and symbols are a silver bullet for password security.

Years were common among participants' passwords. Unfortunately, years and dates are also well represented among top guesses by cracking tools [65]. Most of our participants did not seem to realize how predictable years are. For example, P49 made the password *Its1987* "because [he] was born in [that year]." Similarly, P25 appended "68" onto her full name to create her password because she was born in 1968. Perhaps even more predictably, P36 created the password *IloveNDT2014!* "since it's 2014."

Participants who used dates or years in their passwords commonly seemed to think only about threats from targeted attacks. For instance, P6 explained, "I think a malicious person will try my birthday and my name, so I will not use my birthday itself...I will use family information, such as my sister's birthday or my parents' birthday, or my pet dog's birthday." Compared to using her own birthday, P6's choices would give less of an advantage to an attacker in a user-specific attack, yet would do little to thwart a general attack.

Other choices of digits and symbols were more novel. Some participants used long sequences of digits and symbols that would be hard to predict. For instance, P9 used her student ID number from when she was in high school as the beginning of a password that was not guessed. In many other cases in which the digits or symbols were predictable, other parts of the password contributed enough unpredictability to make the password secure overall. P21 crafted *bAMBANG5$555*, combining his father's hometown with "5" because it "is a lucky number" and "$" "because it is a bank account." P42 also used the "dollar sign for money because this is a banking account." Her password, *ilovebillyC$1*, used the 1 at the end because Billy is her boyfriend and "he's number 1" in her book. A handful of other participants used digits to mirror meta-aspects of thir password. For instance, P46 included "2" in *Lethe+Styx27* because "there are two rivers (Lethe and Styx)" in the password, yet the "7" was random.

4.5.5 Meeting Requirements

One aspect where many participants came up short was deciding what to do when they did not meet a length requirement. The aforementioned transformation of "journal" into "journalistic" was a creative and accidentally effective approach of meeting a length requirement. In contrast, many participants simply tacked on a predictable number and symbol. Having decided to associate a piggy bank with the banking website, P8 came up with "pink piggy," yet found it was not long enough. Therefore, she added "1!," which is the most predictable ending for a password [60]. However, such predictable additions to a password have the added benefit of meeting digit and symbol requirements, which is why P2 said he always appends "1@." P49 found that "elephant" was not long enough and she said, "I don't want to choose another word," so she added numbers. This moment in password creation is ripe for a more thoughtful intervention to help users make more creative, and hence less predictable, micro-decisions.

Other participants did not take full advantage of the tools at their disposal. For example, P1 is one of only two participants to use a password manager browser plugin, yet he reused the same weak password across all three accounts, which he said was typical of his behavior with his real accounts. Because his reused password ends with a three-digit "random number," he uses KeyPass to remember it. Instead, he would be far better served by using KeyPass to generate a unique, much harder to guess password for each account. Similarly, although P14 said that he writes his passwords down, he combined his sister's name and his initials into the password *jennlp1*, which was guessed in the general attack. Because he does not believe he can recall his passwords strictly from memory and thus plans to write the password down, he might be better served by making a less memorable, yet more secure, password.

5. DISCUSSION

Our participants' many misconceptions about passwords reveal that we, the community of password researchers and system administrators, are falling short in helping users understand how to make a secure password. We have become accustomed to shaming users for egregiously bad passwords [9, 35, 62]. However, we seem to have overlooked how to help motivated, well-intentioned users understand what precisely distinguishes a good password from a bad password, as well as what exactly they should be concerned about in the greater ecosystem of password security.

As a result, participants' folk models [68] about secure password behaviors often diverged from reality. In this section, we discuss directions for both improving the static advice given to users about password creation and designing interactive, data-driven tools to help users more intuitively understand why certain behaviors are predictable. These approaches aim to correct major misconceptions we identified in our qualitative data and thereby help users create secure passwords when they intend to do so.

5.1 Improving Advice About Passwords

We found that most of our participants have a human algorithm for generating their password. Some of these algorithms were generally secure, such as generating unique phrases that the participant associated with a site. In contrast, less secure algorithms often centered on reusing a base string with minor additions across accounts.

Summarizing good and bad behaviors for users can be difficult. We observed both secure and insecure variants of the same conceptal behaviors, and it is hard to capture nuances that distinguish them succintly. For example, participants who developed their password from an obscure song they associate with the website's logo were behaving securely. Those who made phrases like "ilove*SiteName*!" were not behaving securely despite also crafting a phrase related to the website.

5.1.1 *Promoting Secure Human Algorithms*

Security advice and requirements could focus on helping users develop and accurately judge human algorithms for developing passwords [49] rather than assuming the enforcement of a password-composition policy is sufficient. Password-composition policies often focus on character-class structures [29,33,56], rather than approaches and algorithms. A better approach might be to help users develop abstract approaches for generating passwords and accurately judge whether decisions they make are predictable.

For example, many participants expected that adding a digit or symbol to a password they considered weak would transform it into a secure password. These users seem to have misinterpreted canonical advice about including digits and symbols in passwords (on the assumption of increasing the password space) to be sufficient on its own for making a password secure. Abstract advice to "include digits and symbols" should be reworked to specify that these should be included randomly throughout the password. Tacking a digit or symbol onto the end of the password is not enough.

5.1.2 *Assigning Value to Accounts*

Participants' understanding and opinion of the relative value of accounts seemed surprisingly out of sync with what the security community might recommend. A rational user would make a simple password for all low-value accounts [48], such as accounts on news sites. Users should be reassured that "newspaper123" is actually a perfectly reasonble password for a low-value account; they should save their limited mental capacity for passwords for more important accounts [18,42]. We had expected email accounts, which can often be reused to reset passwords for other accounts, to be considered most valuable, followed by banking accounts. We expected accounts on news sites to be considered low-value. Only two of the 49 participants shared this appraisal, however, whereas 21 participants considered all accounts to be about the same value. Users need more guidance on how to make such value decisions so that they can reserve their effort for high-value accounts.

5.1.3 *Understanding Threats*

When explaining what makes a password insecure, participants mentioned targeted attacks using their own birthdays, names, addresses, and family members more frequently than attacks on abstract, yet predictable, behaviors. In essence, they were thinking of targeted attacks, rather than large-scale guessing. The community could better explain that both threats should be considered.

Furthermore, participants did not seem to understand automated guessing attacks. For example, the participant who expected that words that are harder to spell are harder for an attacker to guess is likely unaware of the mostly automated attacks [22] that take place when a password database is compromised [5,9,43,57]. Attackers are not typing candidate guesses; they are using word lists. Education efforts might build on Zhang-Kennedy et al.'s infographics helping users understand how password-guessing attacks work [73].

Participants also seemed to misunderstand the impact of password reuse on security. Reuse becomes problematic when an account is compromised and the attacker can then attack a high-value account with the same credentials. Thus, a user should have a set of distinct passwords for each of her handful of high-value accounts, whereas reuse is rational for low-value accounts [18].

5.2 Better Data-Driven Feedback

In concert with improved advice, interactive, data-driven tools could help correct users' misconceptions. In essence, data-driven approaches could help a user understand what everyone else does, pointing out the predictability of their own choices in crafting a password through examples.

We consistently observed that participants wanted different security levels for different sites, yet nonetheless crafted passwords of similar objective security or, in a few cases, more secure passwords for their intended lowest-value accounts. Currently, typical password-strength meters tell a user simply that their password is "very weak" or "very strong" based mainly on the password's length and number of character classes [15]. Large jumps in the password-security estimate by typical meters whenever a user adds an additional character class may have even contributed to the misconception that simply adding a digit or symbol to the end of a password greatly increases security. Instead, a data-driven feedback tool could leverage simulations of adversarial password cracking [31] to tell the user how long an attacker would take to guess that password. Similarly, tools could use large data sets of leaked passwords to help the user learn what patterns are predictable [32].

Participants seemed somewhat oblivious to the predictability and ubiquity of certain approaches, such as using a very common phrase as the basis for a password. Contrary to some participants' misconceptions, adding a "1" or "!" to the end of a password does not make it secure, and a famous quote is not a very good starting point for a password. Targeted, data-driven feedback during password creation could point out insecure behaviors that users seemed to think add security, yet might actually make a password weaker. The security misconceptions we note in this paper are a potential first set of detailed insecure behaviors to target. Such a tool could also point out the percentage of other users who employ such behaviors; leaked password sets could be used to bootstrap this tool.

Finally, we belive there is an opportunity for researchers to help users think of their password elements in unpredictable order and thereby craft passwords with less predictable structure. Participants often thought first of a word and then digits and symbols, thereby constructing their password in that (very common) order. One can imagine a tool that builds on prior work on persuasive tools [19] and automatically prompts users to think of password elements in a different order, or even automatically moves digits and symbols into unpredictable areas in the middle of the password.

6. CONCLUSION

We have reported on the first qualitative lab study of precisely how users construct passwords step-by-step. We found that many users have algorithms for developing passwords. Some of these algorithms are by and large secure, while others continually lead to passwords that are easy to guess. Many participants aimed to construct passwords of different security levels, yet the passwords they intended to be weak were often comparable to those they desired to be very strong. We also delved into participants' micro-decisions in constructing a password, finding numerous secure and insecure sources of inspiration for the words, phrases, digits, and structures they employ to craft a password. Building on participants' decisions and misconceptions, we have outlined directions for improving password guidance and designing data-driven tools to help users craft secure passwords for accounts they care about.

7. ACKNOWLEDGMENTS

This research was supported in part by NSF grants DGE-0903659 and CNS-1116776, and by a gift from Microsoft Research. This research was also conducted with government support under and awarded by DoD, Air Force Office of Scientific Research, via the National Defense Science and Engineering Graduate (NDSEG) Fellowship, 32 CFR 168a.

8. REFERENCES

[1] S. V. Acker, D. Hausknecht, W. Joosen, and A. Sabelfeld. Password meters and generators on the web: From large-scale empirical study to getting it right. In *Proc. CODASPY*, 2015.

[2] A. Adams and M. A. Sasse. Users are not the enemy. *Communications of the ACM*, 42(12):40–46, 1999.

[3] H. Beyer and K. Holtzblatt. *Contextual Design: Defining Customer-Centered Systems*. Morgan Kaufmann, 1998.

[4] C. Bhagavatula, B. Ur, K. Iacovino, S. M. Kywe, L. F. Cranor, and M. Savvides. Biometric authentication on iphone and android: Usability, perceptions, and influences on adoption. In *Proc. USEC*, 2015.

[5] J. Bonneau. The Gawker hack: How a million passwords were lost. *Light Blue Touchpaper* Blog, December 2010. http://www.lightbluetouchpaper.org/2010/12/15/the-gawker-hack-how-a-million-passwords-were-lost/.

[6] J. Bonneau. The science of guessing: Analyzing an anonymized corpus of 70 million passwords. In *Proc. IEEE Symposium on Security and Privacy*, 2012.

[7] J. Bonneau, C. Herley, P. C. van Oorschot, and F. Stajano. The quest to replace passwords: A framework for comparative evaluation of Web authentication schemes. In *Proc. IEEE Symposium on Security and Privacy*, 2012.

[8] J. Bonneau and E. Shutova. Linguistic properties of multi-word passphrases. In *Proc. USEC*, 2012.

[9] J. Brodkin. 10 (or so) of the worst passwords exposed by the LinkedIn hack. *Ars Technica*, June 2012.

[10] W. E. Burr, D. F. Dodson, and W. T. Polk. Electronic authentication guideline. Technical report, NIST, 2006.

[11] D. Cameron. Apple knew of iCloud security hole 6 months before Celebgate. *The Daily Dot*, September 24 2014. http://www.dailydot.com/technology/apple-icloud-brute-force-attack-march/.

[12] C. Castelluccia, M. Dürmuth, and D. Perito. Adaptive password-strength meters from Markov models. In *Proc. NDSS*, 2012.

[13] S. Chiasson, A. Forget, E. Stobert, P. C. van Oorschot, and R. Biddle. Multiple password interference in text passwords and click-based graphical passwords. In *Proc. CCS*, 2009.

[14] A. Das, J. Bonneau, M. Caesar, N. Borisov, and X. Wang. The tangled web of password reuse. In *Proc. NDSS*, 2014.

[15] X. de Carné de Carnavalet and M. Mannan. From very weak to very strong: Analyzing password-strength meters. In *Proc. NDSS*, 2014.

[16] S. Fahl, M. Harbach, Y. Acar, and M. Smith. On the ecological validity of a password study. In *Proc. SOUPS*, 2013.

[17] D. Florencio and C. Herley. A large-scale study of web password habits. In *Proc. WWW*, 2007.

[18] D. Florencio, C. Herley, and P. C. van Oorschot. Password portfolios and the finite-effort user: Sustainably managing large numbers of accounts. In *Proc. USENIX Security*, 2014.

[19] A. Forget, S. Chiasson, P. C. van Oorschot, and R. Biddle. Improving text passwords through persuasion. In *Proc. SOUPS*, 2008.

[20] S. Gaw and E. W. Felten. Password management strategies for online accounts. In *Proc. SOUPS*, 2006.

[21] D. Goodin. Hackers expose 453,000 credentials allegedly taken from Yahoo service. *Ars Technica*, July 2012. http://arstechnica.com/security/2012/07/yahoo-service-hacked/.

[22] D. Goodin. Why passwords have never been weaker—and crackers have never been stronger. *Ars Technica*, August 2012. http://arstechnica.com/security/2012/08/passwords-under-assault/.

[23] D. Goodin. Anatomy of a hack: How crackers ransack passwords like "qeadzcwrsfxv1331". *Ars Technica*, 2013. http://arstechnica.com/security/2013/05/how-crackers-make-minced-meat-out-of-your-passwords/.

[24] D. Goodin. “thereisnofatebutwhatwemake”-turbo-charged cracking comes to long passwords. *Ars Technica*, August 2013. `http://arstechnica.com/security/2013/08/thereisnofatebutwhatwemake-turbo-charged-cracking-comes-to-long-passwords/`.

[25] Google. Web 1T 5-gram version 1, 2006. `http://www.ldc.upenn.edu/Catalog/CatalogEntry.jsp?catalogId=LDC2006T13`.

[26] Google. 2-step verification. `https://www.google.com/landing/2step/`, 2015.

[27] B. Grawemeyer and H. Johnson. Using and managing multiple passwords: A week to a view. *Interacting with Computers*, 23(3), June 2011.

[28] C. Herley. So long, and no thanks for the externalities: the rational rejection of security advice by users. In *Proc. NSPW*, pages 133–144, 2009.

[29] P. Inglesant and M. A. Sasse. The true cost of unusable password policies: password use in the wild. In *Proc. CHI*, 2010.

[30] M. Jakobsson and M. Dhiman. The benefits of understanding passwords. In *Proc. HotSec*, 2012.

[31] P. G. Kelley, S. Komanduri, M. L. Mazurek, R. Shay, T. Vidas, L. Bauer, N. Christin, L. F. Cranor, and J. Lopez. Guess again (and again and again): Measuring password strength by simulating password-cracking algorithms. In *Proc. IEEE Symposium on Security and Privacy*, May 2012.

[32] S. Komanduri, R. Shay, L. F. Cranor, C. Herley, and S. Schechter. Telepathwords: Preventing weak passwords by reading users’ minds. In *Proc. USENIX Security*, 2014.

[33] S. Komanduri, R. Shay, P. G. Kelley, M. L. Mazurek, L. Bauer, N. Christin, L. F. Cranor, and S. Egelman. Of passwords and people: Measuring the effect of password-composition policies. In *Proc. CHI*, 2011.

[34] KoreLogic. “Crack Me If You Can” - DEF CON 2010. `http://contest-2010.korelogic.com/rules.html`, 2010.

[35] L. Kornblatt. When “most popular” isn’t a good thing: Worst passwords of the year – and how to fix them. `http://www.splashdata.com/press/PR111121.htm`, November 2011.

[36] C. Kuo, S. Romanosky, and L. F. Cranor. Human selection of mnemonic phrase-based passwords. In *Proc. SOUPS*, 2006.

[37] D. Love. Apple on iCloud breach: It’s not our fault hackers guessed celebrity passwords. *International Business Times*, September 2 2014. `http://www.ibtimes.com/apple-icloud-breach-its-not-our-fault-hackers-guessed-celebrity-passwords-1676268`.

[38] M. L. Mazurek, S. Komanduri, T. Vidas, L. Bauer, N. Christin, L. F. Cranor, P. G. Kelley, R. Shay, and B. Ur. Measuring password guessability for an entire university. In *Proc. CCS*, 2013.

[39] B. D. Medlin and J. A. Cazier. An empirical investigation: Health care employee passwords and their crack times in relationship to hipaa security standards. *IJHISI*, 2(3), 2007.

[40] Microsoft. About two-step verification. `http://windows.microsoft.com/en-us/windows/two-step-verification-faq`, Accessed 2015.

[41] D. A. Milman. Death to passwords. ComputerWorld. `http://blogs.computerworld.com/17543/death_to_passwords`, 2010.

[42] R. Nithyanand and R. Johnson. The password allocation problem: Strategies for reusing passwords effectively. In *Proc. WPES*, 2013.

[43] N. Perlroth. Adobe hacking attack was bigger than previously thought. *The New York Times Bits Blog*, Oct. 2013. `http://bits.blogs.nytimes.com/2013/10/29/adobe-online-attack-was-bigger-than-previously-thought/`.

[44] PHDays. “Hash Runner” - Positive Hack Days. `http://2013.phdays.com/program/contests/`, 2013.

[45] R. W. Proctor, M.-C. Lien, K.-P. L. Vu, E. E. Schultz, and G. Salvendy. Improving computer security for authentication of users: Influence of proactive password restrictions. *Behavior Research Methods, Instruments, & Computers*, 34(2):163–169, 2002.

[46] N. Provos and D. Mazieres. A future-adaptable password scheme. In *Proc. USENIX ATC*, 1999.

[47] A. Rao, B. Jha, and G. Kini. Effect of grammar on security of long passwords. In *CODASPY*, 2013.

[48] B. Schneier. MySpace passwords aren’t so dumb. `http://www.wired.com/politics/security/commentary/securitymatters/2006/12/72300`, 2006.

[49] B. Schneier. Password advice. `www.schneier.com/blog/archives/2009/08/password_advice.html`, August 2009, retrieved September 2012.

[50] SCOWL. Spell checker oriented word lists. `http://wordlist.sourceforge.net`, 2015.

[51] R. Shay, S. Komanduri, A. L. Durity, P. S. Huh, M. L. Mazurek, S. M. Segreti, B. Ur, L. Bauer, N. Christin, and L. F. Cranor. Can long passwords be secure and usable? In *Proc. CHI*, 2014.

[52] S. Singh, A. Cabraal, C. Demosthenous, G. Astbrink, and M. Furlong. Password sharing: Implications for security design based on social practice. In *Proc. CHI*, 2007.

[53] F. Stajano. Pico: No more passwords! In *Proc. SPW*, 2011.

[54] J. Steubbe. Hashcat. `http://hashcat.net/oclhashcat-plus/`, 2009.

[55] E. Stobert and R. Biddle. The password life cycle: User behaviour in managing passwords. In *Proc. SOUPS*, 2014.

[56] W. C. Summers and E. Bosworth. Password policy: The good, the bad, and the ugly. In *Proc. WISICT*, 2004.

[57] Trustwave. eHarmony password dump analysis, June 2012. `http://blog.spiderlabs.com/2012/06/eharmony-password-dump-analysis.html`.

[58] Trustwave Spiderlabs. SpiderLabs/KoreLogic-Rules. `https://github.com/SpiderLabs/KoreLogic-Rules`, 2012.

[59] B. Ur, P. G. Kelly, S. Komanduri, J. Lee, M. Maass, M. Mazurek, T. Passaro, R. Shay, T. Vidas, L. Bauer, N. Christin, and L. F. Cranor. How does your password measure up? The effect of strength meters on password creation. In *Proc. USENIX Security*, August 2012.

[60] B. Ur, S. Komanduri, R. Shay, S. Matsumoto, L. Bauer, N. Christin, L. F. Cranor, P. G. Kelley, M. L. Mazurek, and T. Vidas. Poster: The art of password creation. In *Proc. IEEE Symposium on Security and Privacy*, 2013.

[61] B. Ur, S. M. Segreti, L. Bauer, N. Christin, L. F. Cranor, S. Komanduri, D. Kurilova, M. L. Mazurek, W. Melicher, and R. Shay. Measuring real-world accuracies and biases in modeling password guessability. In *Proc. USENIX Security*, 2015.

[62] A. Vance. If your password is 123456, just make it HackMe. New York Times, **http://www.nytimes.com/2010/01/21/technology/21password.html**, 2010.

[63] A. Vance, D. Eargle, K. Ouimet, and D. Straub. Enhancing password security through interactive fear appeals: A web-based field experiment. In *Proc. HICSS*, 2013.

[64] R. Veras, C. Collins, and J. Thorpe. On the semantic patterns of passwords and their security impact. In *Proc. NDSS*, 2014.

[65] R. Veras, J. Thorpe, and C. Collins. Visualizing semantics in passwords: The role of dates. In *Proc. VizSec*, 2012.

[66] K.-P. L. Vu, R. W. Proctor, A. Bhargav-Spantzel, B.-L. B. Tai, and J. Cook. Improving password security and memorability to protect personal and organizational information. *IJHCS*, 65(8):744–757, 2007.

[67] C. Warner. Passwords with simple character substitutions are weak. **http://optimwise.com/passwords-with-simple-character-substitution-are-weak/**, 2010.

[68] R. Wash. Folk models of home computer security. In *Proc. SOUPS*, 2010.

[69] M. Weir. The RockYou 32 million password list top 100. **reusablesec.blogspot.com/2009/12/rockyou-32-million-password-list-top.html**, December 2009, retrieved September 2012.

[70] M. Weir, S. Aggarwal, M. Collins, and H. Stern. Testing metrics for password creation policies by attacking large sets of revealed passwords. In *Proc. CCS*, 2010.

[71] M. Weir, S. Aggarwal, B. d. Medeiros, and B. Glodek. Password cracking using probabilistic context-free grammars. In *Proc. IEEE Symposium on Security and Privacy*, 2009.

[72] Y. Zhang, F. Monrose, and M. K. Reiter. The security of modern password expiration: An algorithmic framework and empirical analysis. In *Proc. CCS*, 2010.

[73] L. Zhang-Kennedy, S. Chiasson, and R. Biddle. Password advice shouldn't be boring: Visualizing password guessing attacks. In *Proc. eCRS*, 2013.

APPENDIX

A. INTERVIEW SCRIPT

Good {morning, afternoon}. My name is ______ and I will be moderating your interview today. First, I would like you to review this consent form. It contains important information about today's interview. If you consent to the terms and would like to participate in the study, please sign the form and hand it back to me.

Today, we will be asking you questions that relate to how you create passwords. You are free to choose not to answer any questions, and to stop the interview at any point if you feel uncomfortable. We will ask you to create new passwords as if you were creating them for certain services. Please note that we will not ask you to tell us your passwords you use for the services you are currently signed up for in your real life. We greatly value your honest and candid responses.

Later in this session, we would like to make an audio recording of this session. This recording will only be used for the purposes of this study and will only be accessible to the researchers. Do you consent to having this session audio recorded?

First, I would like to ask you about yourself.

Demographics

1. How old are you?
2. What is your nationality?
3. What is your gender?
4. What is your occupation?
5. (If relevant from the answer to the previous question) What is your {role, expertise, major} in your occupation?
6. What is the highest level of education that you have completed?
7. Please rate your computer skills and knowledge from 1 to 5, with 1 being the lowest and 5 being the highest?
8. For how many hours a day on average do you use the Internet?
9. Do you use your own computer, or a shared computer, when using the Internet?
10. What kind of computer devices do you use? (PC, smart phone, tablet, etc.)
11. (If answer to the question 10 includes a smart phone) What kind of smart phone do you use?

Now, let me start recording audio.

12. What is the purpose of using the Internet on your device(s)? What kinds of websites do you visit?
13. (If the participant has multiple devices) When using the Internet, do you use all the devices for the same purposes at the similar frequency, or do you use them differently?
14. (If the answer to the question 13 is "differently") Why do you use those devices differently when using the Internet? Do you think one device is more secure than others? Other reasons?
15. Do you need to type in a password for any of the Internet services on your device(s)?
16. What kind of Internet services have you signed up for?

Password Creation

Thank you. Now, I am going to ask you to create new passwords for 3 different types of services.

You will be asked to type the passwords in the password field on the computer. While creating the passwords, we will ask you to "think aloud" to help us understand what your password-creation strategy is.

Please listen carefully to the example of "thinking aloud" I am going to give you, and please follow the same method when you create new passwords.

> I am thinking aloud to design an anti-drunk driving bumper sticker as an example. Now let's see.., I'll start with the colors I will be using. One of the colors I will use is red because it symbolizes both warning and blood. I thought of blood because drunk driving may cause accidents that involve injuries or deaths. And for the design... hmm, well, I want to use something like the no-smoking sign, that is, a red circle with backslash, and... let's place something that represents alcohol behind it.... say, a beer bottle, because it is very straightforward at a glance. And having a beer bottle placed behind a red circle with the backslash would be, uh, pretty easy to understand that it means alcohol beverages are prohibited. Hmm, actually, should I use a can or bottle? Um, I think I will stick with a beer bottle, because a beer can may look like a soda can. Hmm, and the background color should be definitely white because it would make the symbol stand out more. I will also put "Don't drink and drive" on the right hand side of the symbol using a big font, uhh, probably in green because it is an opposite color to red, so along with the red symbol, I think it will make the sticker look really striking. Oh, and I will put an exclamation mark at the end of the sentence, because it would give us impression of urgency and importance.

Now, we are going to ask you to use the same thinking-aloud process when you create passwords. Again, please do not use the same passwords you are currently using for any of your real accounts. Please pretend that you are actually creating new passwords to sign up for new services, and remember to say aloud what exactly you are thinking when doing so. Please create a password in the same way you would if this were your real password and you need to use those passwords again to log in to the account you sign up for. In addition, please take the steps you would normally take to remember your password and protect this password as you normally would protect the password for this type of account. For example, if you normally write down this type of password, you should go ahead and do that. Please behave as you would if this were your real password! After you finish creating your passwords for this study, we're going to have you do some tasks to clear your mind, and then we're going to have you try to log in again using the password you created. So keep that in mind as you create your password. Please also explain the reason for choosing a certain words or characters, for example, please say "I will choose 1 as a numerical character because it is easy to remember," instead of just "I will choose 1 as a numerical character." Also please note that I may ask you for a clarification for the reason you chose certain words or characters.

You are asked to create 3 passwords for different type of services and those are free news subscription, email, and online banking services. All of your passwords must follow a certain composition rule.

You will be asked to create your new passwords starting at the next step, but please do not start typing a password until I ask you to do so.

Please click continue.

Password for of News/Email/Bank
Now, please pretend that you are creating a new password to {log into a free newspaper website such as NY Times, CNN, etc. | sign up for an email account such as Gmail, Hotmail, etc. | sign up for an online banking service}. Again, while creating password, say aloud exactly what you are thinking, as in the example I gave you. Please do not click continue until I ask you to do so. Please use the computer and start typing in the password you create as you explain your thinking process.

After confirming the thinking aloud process
Thank you. Please click continue and do not start typing until I ask you to do so.

Password for News/Email/Bank
Now, please pretend that you are creating a new password to {log into a free newspaper website such as NY Times, CNN, etc. | sign up for an email account such as Gmail, Hotmail, etc. | sign up for an online banking service}. Again, while creating password, say aloud exactly what you are thinking, as in the example I gave you. Please do not click continue until I ask you to do so. Please use the computer and start typing in the password you create as you explain your thinking process.

After confirming the thinking aloud process
Thank you. Please click continue and do not start typing until I ask you to do so.

Password for News/Email/Bank
Now, please pretend that you are creating a new password to {log into a free newspaper website such as NY Times, CNN, etc. | sign up for an email account such as Gmail, Hotmail, etc. | sign up for an online banking service}. Again, while creating password, say aloud exactly what you are thinking, as in the example I gave you. Please do not click continue until I ask you to do so. Please use the computer and start typing in the password you create as you explain your thinking process.

After confirming the thinking aloud process
Thank you. Please click continue and please do not click anything until I ask you to do so.

Now, please answer a few additional questions related to your password-creation strategy.

17. What is the first thing on your mind that comes up when creating a password? (If clarification requested: Is it a digit, keyword, something that associates the type of services, or something else?)
18. How do you pick your keywords? Are they related to your hobbies, what you find in your room, or something else?
19. Could you tell me what process you go through to select what you use for your passwords? (If clarification requested: For example, you can say "First, I pick up a favorite song, and then a songwriter. Then I choose a number associated with the song like a released year," etc.)
20. What aspect of your password do you think make the password harder to crack?

Password Challenges

21. As far as you know, have any of your Internet service account passwords been stolen or leaked?
22. (If answer to the question 21 is yes) Did it cause you to change anything about the way you create passwords? How?
23. Was your password creation strategy in our study today different from the way you create passwords normally?
24. (If the answer to the question 23 is yes) How different was it, and why?

25. When was the last time you created a new password before the study? You may answer by rough estimate.
26. What strategy did you use to create the password you created last time? Was it different from the one you used today?
27. How did you come up with the strategy you used today (and you use normally)?
28. (If the passwords contain a number or symbol) How did you come up with the number/symbol you used today?
29. Have you ever reused any of your passwords exactly as they are before?
30. Why? / Why not?
31. Have you ever reused a part of any of your passwords before?
32. Why? / Why not?
33. (If the answer to question 29 or 31 is yes) Do you reuse your passwords rarely, always, often or only occasionally?
34. (If the answer to question 29 or 31 is yes) When are you likely to reuse your password and why?
35. (If the answer to question 29 or 31 is yes) When are you not likely to reuse your password and why?
36. (If the answer to the question 31 is yes) Would you explain how you modify existing password for reuse?
37. Are you concerned about the security of reusing passwords?
38. (If the answers to both question 29 or 31 and 37 are yes) Then why did you reuse your passwords?
39. Some passwords are required to be changed periodically, like every 90 days. Have you changed a password for an existing account because you are required to change it periodically? If yes, what was the strategy you used to change the password? Did you create a whole new password or modify the old password?

Thank you.

Distraction Task
Now, I am going to ask you to do some simple task. This task is not related to passwords, so you can relax and take your time to finish it. Please count backwards from 100 in sevens, that is, starting from 100, subtract 7 from each number you say, like 100, 93, and so on.

Now, recall you created 3 new passwords during the first session. You will be asked to enter the passwords you created to see if you remember them correctly. For this session, please think aloud again to help me understand how you are remembering your password. Please note that the type of service shown on the screen may be in a different order from the one you saw when you created your passwords.

You can click the continue button on the screen to start entering the password. You are allowed to make up to 5 attempts per password.

Remembering {News/Email/Bank} Password
Please enter the password you created for the {News/Email/Bank} account, and while doing so, please think aloud again how you are remembering your password. Please click continue to see if you remembered it correctly or not.

40. If they did not remember the password in 5 times: Why do you think you had a problem remembering the password?

Remembering {News/Email/Bank} Password
Please enter the password you created for the {News/Email/Bank} account, and while doing so, please think aloud again how you are remembering your password. Please click continue to see if you remembered it correctly or not.

41. If they did not remember the password in 5 times: Why do you think you had a problem remembering the password?

Remembering {News/Email/Bank} Password
Please enter the password you created for the {News/Email/Bank} account, and while doing so, please think aloud again how you are remembering your password. Please click continue to see if you remembered it correctly or not.

42. If they did not remember the password in 5 times: Why do you think you had a problem remembering the password?
43. How do you usually remember your passwords? Do you store it electronically, such as using a password manager, or write it down on paper?
44. (If the answer to 43 is electronically) Which tool do you use to store passwords?
45. (If the answer to 43 is paper) Where do you write it down? Do you look at it every time you have to enter it or do you try to to remember it first?
46. How hard do you think it would be for your friends or family to guess your password? If a cybercriminal were to compile a list of the most common passwords, do you think your password would be on it? What if that list was compiled by the government?
47. Is there anything else about password creation strategy you used you think is useful for us to know about how you create passwords?

Interview Conclusion
Thank you very much for participating! Your feedback has been valuable to our research. We will eventually write a research paper about conversations we are having with you and a number of other research participants. In the research paper, we would like to include quotes from some of our participants, attributing these quotes to "Participant #."

47. Do you give us your permission to use quotes from you in this research paper?
48. Are there any things we discussed today that you would like us to not quote?

Thanks again!
Compensates participant.

Social Media as a Resource for Understanding Security Experiences: A Qualitative Analysis of #Password Tweets

Paul Dunphy[1], Vasilis Vlachokyriakos[1], Anja Thieme[1], James Nicholson[2],
John McCarthy[3], Patrick Olivier[1]
[1] Culture Lab, School of Computing Science, Newcastle University
[2] PaCT Lab, Northumbria University
[3] School of Applied Psychology, University College Cork
1{forename.surname}@ncl.ac.uk, 2james.nicholson@northumbria.ac.uk,
3john.mccarthy@ucc.ie

ABSTRACT

As security technologies become more embedded into people's everyday lives, it becomes more challenging for researchers to understand the contexts in which those technologies are situated. The need to develop research methods that provide a lens on personal experiences has driven much recent work in human-computer interaction, but has so far received little focus in usable security. In this paper we explore the potential of the micro blogging site *Twitter* to provide experience-centered insights into security practices. Taking the topic of passwords as an example, we collected tweets with the goal to capture personal narratives of password use situated in its context. We performed a qualitative content analysis on the tweets and uncovered: how tweets contained critique and frustration about existing password practices and workarounds; how people socially shared and revoked their passwords as a deliberate act in exploring and defining their relationships with others; practices of playfully bypassing passwords mechanisms and how passwords are appropriated in portrayals of self. These findings begin to evidence the extent to which passwords increasingly serve social functions that are more complex than have been documented in previous research.

1. INTRODUCTION

Nearly twenty years have passed since early calls for usable security [40] where security researchers were encouraged to lend contemporary research methods from the field of human-computer interaction (HCI). The HCI methods of the day were focused upon improving the efficiency of users in their interactions with digital technology in the workplace; however, even at that point in time, digital technologies were already accelerating on a trajectory of becoming tightly interwoven into people's everyday lives. Designing and evaluating digital technologies that will play this role is challenging, because the success criteria for that technology is not only related to the efficiency of the interface interactions but must include a broader agenda of understanding how technologies involve people emotionally, intellectually and sensually [25].

It is challenging to understand and design for these facets of technology usage as people appropriate technologies differently according to their own personal circumstances, past experiences and anticipations for the future. However, engaging with this challenge and taking an experience-centered [39] approach to design requires designers to continuously seek new perspectives on how people live with the technologies they design [27] and accept that interactions and experiences are unique to the person through their own interpretations, feelings and value judgments [24]. By trying to see the world through the eyes of another person and respecting the different values held by others, the designer is better placed to generate a rich, contextual understanding of the problem at hand as well as new design ideas. Recent work has argued the value of a focus of an experience-centered approach to usable privacy and security [9], but while some research is beginning to address this challenge [29, 36] there is still a dearth of methods for eliciting, evoking and developing descriptions of everyday security experiences.

Social media platforms are used by people to share their thoughts and opinions with friends, family, colleagues, or others with similar interests. Communicating online, people exchange information via 'posts'; an important attribute of these posts is that they are made in a naturalistic setting and in the course of daily activities. The content of these posts is increasingly the focus of analysis for those seeking to better understand people's behaviors and opinions, ranging from understanding political sentiment [2], to tracking the spread of the flu virus [22]. However, while platforms such as Twitter —used for everyday conversations, the sharing of news, and to document daily occurrences [17] —had a user base of 232 million at the end of 2013 [28], no work has yet investigated its potential as a resource of everyday security experiences, nor considered how data of such scope and scale might enhance our understanding of how people appropriate security mechanisms into their lives.

To these ends, this paper makes two key contributions to the study of security experiences: firstly we identify social media, specifically Twitter, as a resource of naturally generated reflections on security practices and workarounds in social settings, and qualitative content analysis (QCA) as an appropriate method of analysis. Secondly, analyzing

Symposium on Usable Privacy and Security (SOUPS) 2015, July 22–24, 2015, Ottawa, Canada.

tweets on the topic of the password we identify key qualitative themes and present data that provide new perspectives on everyday password usage situated within the context and complexity of everyday life, social relations, and practices.

2. RELATED WORK

2.1 Understanding Password Practices

Conventional wisdom on managing passwords securely has not changed much since early guidelines were published [3]. Such guidelines are typically centred upon changing passwords regularly, not sharing passwords with others, not writing them down, and making them difficult to guess for others. Much work has taken place in the usable security community to understand how desired password behaviors (set out in guidelines) manifest in practice. Klein [19] showed that people were likely to choose very short passwords that were vulnerable to guessing attacks structured by a standard word dictionary. Sasse et al. [30] conducted a survey in a workplace and uncovered the challenges people faced to remember passwords; particularly those that were infrequently used. This work also highlighted that the requirements of securely managing passwords had considerable potential to conflict with everyday work practices.

Password sharing is a topic that recurs in studies both in and out of the workplace. Inglesant and Sasse [15] report that shared passwords were a de facto means to access shared resources in the workplace. Outside of the workplace, the literature is more sparse. Kaye [18] reports the results of a survey that describe how passwords were routinely, yet thoughtfully, shared amongst partners and spouses, friends and family and even work colleagues. Indeed, one third of respondents claimed to share their personal email password, compared to one quarter sharing their Facebook password; in both cases sharing was predominantly with partners and close friends. Examples that could be considered more extreme shed light on the how people share passwords to overcome physical or geographical difficulties. Dunphy et al. [8] report how older adults would delegate personal identification numbers (PINs) to helpers who would visit the ATM on their behalf. Singh et al. [32] describe occurrences of password sharing amongst couples but also report how an entire village would delegate bank credentials to a single person who would travel a considerable distance to the nearest bank to conduct business on everybody's behalf.

Each of the presented research studies represent considerable investments in time to work together with user groups 'in the wild' to surface relevant insights. However, while it may be attractive to dismiss some of the examples as particularly extreme, the reality is that for the people living in those contexts, the practices are completely normal and represent a very personal and experience-centered trade-off between usability and security.

2.2 Understanding People via Social Media

Piskorski [28] explains that people are attracted to social media platforms as they offer different ways to interact with others and help fulfil social needs. Twitter is a micro blogging-based social network site that is arguably the most popular example of its kind; it grew in popularity around 2010. The platform offers functionality to connect to strangers, yet it provides some opportunities for interactivity or more private communication with friends. Morris et al. [26] describe how Twitter is used to distribute content such as breaking news, describing the platform not only as a social network, but also as a news source that people access to specifically search for a topic of personal or community interest. Duggan and Smith [7] published a survey report about social media use by Americans, which showed that about 18% of all online adults are Twitter users, compared to the 71% being users of Facebook the largest online social platform. While Facebook is popular across a range of demographic groups, 31% of Twitter users are drawn from the age range 18-29 (19% of Twitter users are in the age range 30-49 years; 9% for 50-64 years; 5% for 65+ years) with a particular presence of urban dwellers, African-Americans and Hispanics. Moreover, Twitter users are split equal in gender and 46% of all Twitter users tend to visit the site daily (29% multiple times per day). Smith and Brenner [33] further reported on a high correlation between the use of Twitter and of mobile technology, especially smartphones, which is generally high among African Americans and Hispanics, and is an increasing trend amongst younger adults.

The challenge to understand peoples' behaviors and opinions outside of the laboratory context in a resource respectful manner is faced by a multitude of researchers. One approach that is increasingly popular concerns the analysis of social media posts as empirical data. For example, Kramer [20] analyzed posts from Facebook and applied techniques of quantitative sentiment analysis to explore the measurement of levels of happiness in the user group. Lampos and Cristianini [22] used a content analysis of tweets to estimate the spread of the H1N1 virus in the UK; their estimates showed a fair agreement with the estimates of central government at the time. Golder and Macy [11] found that social media posts can mirror seasonal behavior fluctuations across different communities. Moreover, social media (specifically Twitter) has been used in the last few years to monitor political sentiment and predict election results [2, 33]. While communication and politics researchers have contrasting views on the validity of Twitter for these purposes —mainly due to concerns around the representativeness of the sample —Twitter has shown itself to be an reasonable predictor of election results (e.g. as an exit poll) when both volume-based measures and sentiment analysis are applied [35].

The analysis of posts on social media platforms has so far served a variety of purposes in the field of human-computer interaction. For example, the act of tweeting is increasingly integrated into large events or even television shows as a way to encourage audience participation, gather feedback, or provoke debate. Doughty et al. [5] analyzed tweets published during a UK-based television show that focused upon the Irish traveler community. They discuss how the tweet contents demonstrated the potential for Twitter to reveal prejudices held by an audience, but also how the micro blogging might have served to reinforce those prejudices.

3. SOCIAL MEDIA AND SECURITY EXPERIENCES

Social media analysis has not yet been considered as a means to understand everyday experiences of security technologies. When considering why we might need new methods to specifically capture naturalistic everyday experiences, it is important to note the two types of knowledge that researchers may aim to elicit from users through research: *explicit knowledge*, and *tacit knowledge* [39]. Explicit knowledge comprises that which is easy to transmit to another person e.g. the number of siblings a person may have, or the number of passwords that a person uses. Much of what we know about the world however is tacit knowledge; knowledge that is difficult to transfer to another person e.g. how to use complex equipment (where experience has guided learning), or reasons for why a system feels secure.

Accessing tacit knowledge typically requires researchers to apply a mix of methods in their interactions with people (e.g. observations, surveys, diaries etc.) to identify and capture instances of interesting behaviors, and may even include working together with participants to highlight that interesting behaviors exist in the first place, but also to understand why those behaviors come about [38]. Relatively few methods in usable security go beyond traditional interviews and surveys to elicit this tacit knowledge; these classic methods best serve as tools to capture explicit knowledge (except where *dialog* [39] is explicitly supported), as these do allow respondents to reflect on their own behaviors, however, within an experimenter defined framework.

The use of social media posts as a lens on everyday security experiences would immediately present a number of methodological benefits:

- People are free to use their own vocabulary to describe their feelings and practices.
- Posts are created in naturalistic contexts and in the course of everyday activities.
- Security would be positioned as social and collective [6], rather than a personal and secretive practice.
- There is a large volume of social media posts to study, and they are typically short in length.

3.1 Everyday Vocabulary

Despite the common mantra that users are not interested in security, recent research suggests that, at times, users give it the utmost care and attention [18]. However, people might not describe their concerns or understanding in ways that experts would immediately recognize. As security features permeate the everyday technologies that people encounter, security becomes more of a subject of public discourse and people are able to collectively develop their own ways to articulate concerns as security becomes a less guarded topic. This vocabulary can spread across communities in the form of colloquial language or even form folk stories [36]. Indeed, the traditional technical lexicon of the security field is one of the reasons why it can be challenging for people to contribute effectively to conversations around information security.

3.2 Naturalistic Insights

Social media posts can be created in the course of everyday living where participants are taking part in neither a lab nor a field study, which frees the data from biases introduced by a researcher or by a specific topic under investigation. There can be many reasons for people to create personal social media profiles to document their everyday lives, however, one of those reasons is unlikely to be to purely discuss security or privacy (although this is possible in the expert community). Consequently, we can think of social media as a side channel into everyday life where personal experiences are foregrounded. This approach can mitigate biases introduced by the researcher that occur naturally in a face-to-face scenario such as pursuing personal interests or directing conversations towards active hypotheses. Of course, alleviating the biases introduced by a researcher in a face-to-face context does not imply that the data are not skewed in any other way (see Limitations and Related Work) or that eliminating all bias is desirable. Indeed, by focusing upon experiences we are placing personal biases of users as the core item of enquiry; micro-blogging sites have been designed to allow fast and concise user interaction and are designed to support users' self-portrayal. As a result, the affordances of the social media platform have an effect on people's interactions and the types of posts they may make.

3.3 Security as a Collective Practice

Information security behaviors are typically considered at the level of the individual. Social media platforms encourage social interactions; so far researchers have studied social media posts and have succeeded mainly in generating insights relating to group behavior [2, 33]. While research documenting group security behaviors does exist e.g. [32, 8, 18], this strand of work is still underexplored and chiefly serves to provide contrast to the mainstream research agenda focused upon the individual; evidenced by the observation that insights from such studies are slow to find their way into proposed system designs. The existence of security-related experiences on social media could help redress the balance and provide opportunities to study diverse online communities and better understand how security is appropriated in such groups. A group perspective on a security technologies could provide opportunities to question how a technology 'should' be used, and shed new light on the ways in which security is actually socially practiced and negotiated [6].

3.4 Big Data of Security Experiences

It is estimated that 500 million tweets are posted per day on Twitter [1]. The sheer volume of social media posts online lends itself to a number of methods of data analysis both qualitative and quantitative. In this paper we take a qualitative approach which reflects our contention that personal experiences can be complicated and personal. However, other methods exist which can support the data analysis process or the filtering down of a large dataset such as *sentiment analysis* [34]; this is one quantitative method widely used in social computing and provides one perspective on the overall mood in the data according to the identification of positive and negative word combinations. While the social media data is owned by a private company, the data is openly searchable which can provide a certain element of transparency in the sourcing of data and comparisons across online communities, even at different points in time according to events that happen in society. Research across a number of disciplines is pursuing methods to aid exploration of large quantities of data, but open questions remain around methods of feeding insights back into the design of digital technology.

4. STUDY METHOD

Over a period of 26 days in February 2014 we collected password-related tweets using the Twitter Search API (this preceded the streaming API which is now widely used). Our search criteria returned all available tweets containing the hashtag '#password' or the keyword 'password', in combination with searches that additionally included specific pronouns e.g. 'I', 'me', 'you', and possessive pronouns e.g. 'my', 'your'. This was to ensure that tweets were as closely related as possible to personal experiences. The search criteria would also return replies to tweets containing that criteria and retweets. During the time period of the study, we downloaded the most recent 15 'pages' of tweets each hour, using individual timestamps as a check to determine whether we had downloaded a particular tweet before. Our focus was on tweets with unique content, so after one pass to filter out retweets, our dataset comprised just over 500,000 tweets ($\mu = 19222$ per day, $\sigma = 3623$). This relatively large dataset indicates a common occurrence of password-related discussion within daily communications on Twitter.

4.1 Data Analysis

The approach of our research was qualitative. We conducted a Qualitative Content Analysis (QCA) [21] on a sample of 1000 tweets that were randomly selected from the dataset, and were roughly balanced across each day of data collection. We chose QCA over alternative approaches such as Grounded Theory [10], since our research is exploratory in nature, with a focus on latent expressions of personal experiences (inductive) and led by existing theories and previous research on peoples' password practices (deductive).

For our QCA, three members of the research team independently familiarized themselves with the data to identify and systematically search for (recurring) themes in the tweets. Identified themes were then coded, first individually and then recoded following conversations between the researchers, and then synthesized to higher-level categories. The comparing of content codes and discussing them with each other formed an essential part in this process, since, for a large number of the tweets, and influenced by our individual perspectives, multiple interpretations and thematic groupings seemed possible. Partly, this was due to a lack of context information that was available in relation to each tweet, which in itself provided little insights about the situation that may have motivated the post, or the intentions of the person sending the message. For some messages it was also not always clear if they were intended to reflect a serious statement or an expression of sarcasm or a joke, leading us to be particularly cautious in their interpretation. After one pass of our QCA we disregarded 302 tweets from further inquiry, as they were either not written in English (124 tweets); presented advertisement or spam messages (52 tweets); were unreadable (e.g. cryptic composition of numbers or characters; 29 tweets); or presented posts that were ambiguous (97 tweets) to the extent that the researchers could not reach agreement about their content. The remaining 698 tweets that were conclusive and agreed on by all three coders were analyzed with regard to the insights they provide for understanding peoples' password practices and concerns. Although we could have included more tweets, we noticed towards the end of our analysis that the themes that we had identified reached saturation, whereby any additional tweets only provided more examples of a similar themes. Our findings present the themes that evolved through this mode of analysis.

4.2 Limitations

Before the presentation of our findings, it is worth clarifying some limitations of the data that will be presented. From an analysis perspective, we disregarded tweets that were not written in English. At the time of our study, The Twitter Search API allowed the retrieval of at most a 1% of all the data matching some specified criteria. After the threshold of 1% has been reached only sampled data is available to the API calls of the user, with the methods of sampling unknown (although premium search options did exist). Such restrictions may be removed in the newer Streaming API. As with any online community there are discrepancies in which members make the biggest contributions; on Twitter, some estimates suggest that 40% of users do not tweet at all, and 90% have less than 10 tweets [14]. Secondly, as with any qualitative data, the data we present should not be scrutinized for its generalizability. The data provides insights confined to the group of people it was collected from and the reader should consider the transferability of the insights with caution. As such, the reader should scrutinize the data for its trustworthiness [12], which is composed of credibility, transferability, dependability, and confirmability.

5. FINDINGS

Our QCA revealed three high level themes. Firstly, people commonly tweeted about specific practical difficulties that they encountered in the set-up or retrieval of passwords; how they experienced such difficulties, as well as a range of individual workarounds they developed to manage these; all of which highlight classic usability and security issues. Secondly, tweets contained insights into some of the complexities that surround peoples' everyday uses of passwords within their social life. In this regard, users expressed how the request, receipt and sharing of passwords with others assisted in exploring and defining a person's relationship with others. Thirdly, some tweets served to portray the personality of a person in a certain light, where the password was used to reinforce certain desired aspects of a person's character. Thus, the final theme presents how people made use of their understanding of, and practices around, their use of passwords to support a specific display of identity.

The particular passwords that people appeared to discuss were mostly online accounts such as social media sites, video-streaming services, and so forth. This might indicate that Twitter is an adequate platform for understanding the experiences of these types of (more social) passwords, or it might be our search criteria were inadequate for collecting insights on other types of password. In the following sections we provide example tweets that illustrate our three themes. How these initial observations can be interpreted is then developed further in the discussion section of the paper.

5.1 Password Practices and Workarounds

Approximately one third of the tweets that we analyzed ($n = 237$, 34%) presented descriptive accounts of the difficulties that people encountered in generating, remembering and recovering their passwords. Even though this could have been expected considering the search criteria that we used to assemble the dataset, it is particularly interesting that these pragmatic accounts were also followed by feelings of anxiety and frustration about a potential loss of access to one's account, as well as some of their personal strategies for approaching password recovery.

5.1.1 Frustration about System Demands for Secure Practices

In a number of tweets ($n = 87$, 12%) users expressed, often in a sarcastic or joking manner, their feelings about current requirements posed upon them in the creation of secure passwords for their accounts. They expressed critique of demands to generate passwords that are very long, include upper or lower case characters; numbers; and to also have to change and to re-enter these frequently. Users tweeted for example: *"Sorry your password must contain a capital letter, two numbers, a symbol, and a inspiring message #BumpDat"*; *"I thought it'll be a password not a pass-speech"*; *"It's always a test of my intelligence when I change my password"*; or *"Why must I type in my Apple ID password EVERY TIME I download an app?!? #Annoying"*.

It is known that such system-enforced security demands are often experienced as a chore and can challenge users' ability to remember their passwords. Thus, we frequently identified posts in which people expressed their frustration about forgotten passwords. Expressions of anger or melancholy in this regard included: *"Omg I forgot my password to my fone!!!!!RS"*; or *"I wish I could change my Wi-Fi name, but apparently I don't have the correct password to do so. #sadface"*. Users also, albeit less often, tweeted about being relieved when they had regained access to an account: *"OMG I forgot you had my password. Wheeeww I thought I was going crazy [username]"*.

In some cases people also commented that losing or forgetting their passwords resulted in long-term loss of access to their online accounts as illustrated by these tweets: *"[username] I forgot my password 4 years ago, soooooo, I don't use that account anymore, Sherlock"*; or *"[username] I got a new twitter because I couldn't retrieve the password from my other one"*. While most online systems offer functionality for password retrieval, for some users this did not appear to be a possible solution or straight-forward process. Expressing frustration about difficulties to recover forgotten credentials, people tweeted for example: *"why is there no way for me to recover my skype name and password? gawd-dddd"*. Moreover, at times, available recovery options did not reflect the authentication problem that the user was facing: *"#ThatawkwardMoment when you can't remember your username, and the only option is "Forgot Password". #Why"*.

As a result, we identified a number of tweets in which users reacted by creating a new account. However, specifically with regard to Twitter this comes at the cost of the person needing to reconstruct the list of people they were following and to rebuild their list of followers. Where the creation of a new account appeared to be the only possible course of action, this did not only feel frustrating to the account owner, but was also greeted by perplexion by others: *"Why do people make a new Twitter account when they forget their password instead of clicking the "Forgot Password button? #comeonpeople"*.

5.1.2 Password Management: Personal Practices & Workarounds

Difficulties related to the remembering of passwords described above led to descriptions of people abandoning existing accounts. In order to prevent this scenario, many users further admitted to have chosen the same password for many (if not all) of their accounts, undermining some of the security demands posed upon them: *"I Got The Same Password For Everythang!"*. Others openly described their approaches to recovering their passwords, explaining how they keep for example a record of these in their email account: *"[username] I can never remember the password LOL [laugh out loud]. I do have it saved in my mail though"*.

In addition to descriptions of such workarounds, people tweeted about how their remembering of certain passwords, especially online passwords, was complicated and addressed through the techniques they had developed through regular experience of password use (n = 150, 15%). Some described difficulties remembering their passwords due to routine, embodied typing mechanics: *"I barely remember my password, my fingers are just used to typing it so much"*. Again, others explained how the practice of apps caching login credentials on mobile devices can cause authentication problems when the user deletes certain mobile apps that hold their saved account details, or if they lose or change their devices. Describing these memorability problems users tweeted for example: *"I dont even know my password on twitter so if i ever lose my phone im f[*]d"*; *"[username] lol I know I forgot my twitter password so I can only tweet from one device where the password is saved"*; or *"[username] got anotha phone n dnt got my password to the account"*.

To gain general advice and support by others about how to recover one's passwords or manage hacked accounts, we found that users tweeted for example IT queries regarding Twitter to the official Twitter account: *"twitter won't let me login on my iPod. It used to but now it tells me "incorrect password or username" I've checked everything"*; or contacted the official support account: *"Support please fix the recaptcha thing it won't let me log in. im only able to log in if i reset my password"*. Moreover, users also frequently posted advice to each other about password management and privacy settings, and they warned one another about security breaches in relation to their accounts being hacked, providing advice how they could regain control. The following tweets illustrate this: *"[username] just go to your settings, go to password, change it, and then review who's on your account and you can revoke their access"*; or *"hi Mar, just to let u know that you may need to change your twitter password as i keep getting spam msgs from your account. Mark xx"*.

5.2 Social Practices around Passwords

A large proportion of the tweets that we analyzed ($n = 306$, 44%), described a wide range of social practices that evolved around the giving, receiving and resetting of a person's password. In the following we present examples of tweets that outline how the sharing of passwords has been a deliberate act in enabling access to certain resource for a specific individual or group of people; and in defining a relationship as close, intimate and trusting, or as doubtful and disappointing. Furthermore, the findings show how trying to guess the (online) password of a person in one's social network can become a playful social challenge rather than being perceived as a security threat; and how people who broadcast their passwords online do so under consideration as to how the target recipients would gain access to their devices or (online) accounts without compromising their overall security. All of these examples contribute to an experiential account of how passwords are embedded and have become of immense importance in peoples' daily and social lives, demonstrating how their role undoubtedly exceeds securing access to personal data.

5.2.1 Sharing Access to Resources with Social Network

We identified a number of tweets in our data set (n = 56, 8%) of users describing how they were sharing passwords as a way to manage and distribute access to resources such as WiFi or a person's Netflix account. For example, users tweeted to request or offer the password to certain online platforms for sharing the benefits and/or costs of a particular service with members of their social network: *"Getting my Netflix account back tomorrow..who wanna go half n get this password??"*; or *"If anyone wants to share with me their Netflix username and password I'll give you a prize (will be food)"*. Passwords were also shared with members of a specific social group to organize access to materials by these people. For example: *"just dm [direct message] me if you want the password to this account, ANYONE from the phandom can use it:)"*; or *"[username] can we recreate 6th grade and make a mom jean patrol club, only those with the password are allowed in"*.

5.2.2 Defining Relationships as Close and Intimate

Further challenging conventional assumptions that passwords are something to be kept private, several tweets (n = 52, 7%) described how people shared their passwords as a way of defining their relationship with friends, family members, or romantic partners. These tweets portrayed, at times playfully, how the giving, receiving, exchanging and possessing of access to someone's passwords reflected a certain closeness and intimacy between, and knowledge of, each other. For example, the following tweets indicate how, quite literally, the sharing and knowledge of a person's password with another person defined their relationship for example as 'romantic' or as 'friends': *"I cant call you my girl till you know my email password?"*; or *"[username] We're not friends until I have your wi-fi password"*.

Similarly, if a person was considered to be a romantic partner or friend, the sharing of their passwords appeared to be a social practice that was expected of them: *"If you don't give me your wifi password then we can't be friends anymore"*. However, this social obligation did not find acceptance by all, as is indicated in this tweet: *"idc [I don't care] if we date what you need the password to all my stuff for"*. Moreover, if the relationship had taken a hit or people quarelled with each other, passwords were revoked: *"[username] I forgot the password on one and the other got blocked by [username] who I had an argument with :D xxx"*. In the context of exchanging passwords between people, we also found posts that reflected an adherence to the social norm of reciprocity, which was especially apparent in the following tweet: *"[username]... she has my password too... its equal"*.

Most striking however were expressions of closeness whereby the knowing of a person's password was considered a prerequisite of knowing the person: *"If you don't know my password than you don't know me"*. Equally, if a person had intimate knowledge of the other, they were better positioned to guess their passwords, as became apparent in the following two tweets: *"I'll give ya'll a hint. It's an 8 character password and the second letter is A"*; or *"my phone password is my crushes last name lol"*. This last tweet further indicates that choosing the names of people one is close to as passwords presents a frequent practice and one that can also cause difficulties, when the social bond is broken off: *"And your name is still my password so I'm reminded of that shit everyday"*.

5.2.3 Giving and Revoking Trust to Others

In this context of password sharing, a number of tweets (n = 45, 6%) linked giving someone access to personal accounts or WiFi networks to an awarding of trust and the expectation that the recipient would treat the password appropriately (however socially defined in the different instances). Reminding the recipient of the responsibility that they had been given through the password, one user tweeted for example: *"Kaitlin's the only one that knows my password now. Sooo, I know it's you that just tweeted that Kaitlin"*.

The relationship between password sharing and trust, however, became most apparent in tweets where users openly expressed their disappointment and frustration about others 'abusing' their online accounts by deleting, changing or adding information: *"Never give your password to sum1 else i learned from that cause they will delete your things"*; or *"i'm so paranoid i'm never giving anyone my password ever from now on oh my god"*. In response to the violation of trust that the password recipient had been given, Twitter users described to have revoked other peoples' access (e.g. by resetting their passwords): *"I'm not giving you the new wifi password next time you come home."*; or *"Gonna have to change my password so Casandra can't see these messages"*.

5.2.4 Playful Social Hacking

We also gathered tweets describing people's frustration and concerns about their accounts being struck by system hacking or phishing attacks (n = 36, 5%) as part of the password difficulties theme above). For example, a common observation was the attempt to deceive people to disclose confidential information, such as: *"#retweet If you type your password on twitter, it shows up as stars! :D ******"*. However, we also identified various tweets (n = 87, 12%) where protecting one's password from known others (usually social connections/friends of the person) became an invitation and playful challenge for some to try 'guess', 'hack' or 'rape' their account. The following tweets exemplify such intents: *"What Breanna thinks my twitter password is?"*; *"If I had Josh's password I'd totally be raping his account right now. #StrangeUrges"*; or *"[username] Jon has my twitter password and conversates with himself"*.

To those who succeeded in guessing other peoples' passwords and hacking into their online accounts, this felt like an achievement, but we also found complaint tweets of those people who fell victim to such an attack: *"Ayeee, Boy Who hacked My Twitter I Will give Youu 24 Hours to Change My Damn name Back or My password Will Be changed"*.

5.2.5 Broadcasting Passwords: Physically and Socially Secured

Occasionally (n = 66, 9%), users also publicly posted their own or other peoples' passwords and pin codes on Twitter. However, those tweets may not have presented an immediate security threat, as they often required additional access to the device to which they would provide access. For example: *"Joes iPod password is 1337"*; or *"[username] Alright, I'm gonna leave my phone in here w/ you guys. My password's 0209 if you need it"*. Moreover, in tweets where users requested a certain password of others, they often suggested a different channel for receiving it (e.g. send as a text message rather than a tweet). The following tweets illustrate this: *"[username] text me the password for raleys wifi"*; and *"[username] whats the WiFi password. Dm [direct message] it lol"*. In other words, only certain parts of an authentication process were publicly revealed that either required the person to be in reach of a particular password-protected device to complete access; to have sufficient 'knowledge' about that person (e.g. their home address to be able to make use of their WiFi credentials); or to qualify (e.g. as a 'close' friend) to be send a text message with the remaining details.

5.3 Use of Passwords and their Practices in Self-Portrayals

As a platform that invites the open sharing of personal opinions and thoughts, Twitter enables people not only to share information, but to communicate something about their self. Thus, we frequently observed how people used their understanding and (mis)uses of passwords to portray a particular image of themself (n = 155, 22%).

5.3.1 Raising Authenticity of One's Online Profile

A large proportion of those posts (n = 80, 11%) included messages of users admitting that they had forgotten their password or had trouble entering it correctly. About trying to remember their passwords and describing their difficulties to log into their accounts, users tweeted for example: *"Ok 12 year old self....what would you have made your password"*; *"I was typing the wrong password into Facebook #FacePalm"*; or *"Me: 'types in password, Password Doesn't Work' OMG I'M HACKED.... oh wait... never mind, CAPS LOCK WAS ON..' "*; and yet another user tweeted: *"[username] yeah but still trying to figure my password to my heathmax12 account. I swear I should be a natural blonde"*. While these tweets say very little about practical problems in relation to how people manage their passwords, they present expressions around the mishandling of situations and the difficulties that arise.

5.3.2 Insights into One's Security Outlook

In contrast to tweets in which people present inadequacies in how they manage their passwords, we also found a small proportion of tweets (n = 11, 2%) in which users presented themselves as concerned about protecting their passwords and keen to keep their accounts and devices secure. Most tweets in this regard included complaints about shoulder surfing attacks, and peoples' personal attitudes towards adding or regularly changing passwords. Examples include: *"I hate when people stare at the keyboard while I'm typing my password"*; *"I should really put a password on my phone."*; and *"Treat your password like your toothbrush. Don't let anybody else use it, and get a new one every six months"*.

5.3.3 Presenting Oneself as Faithful, Cool or Funny

Some of the tweets (n = 27, 4%) with regard to peoples' portrayals of their identity further link to the previously described theme on 'trust', in that users related to their practices of password sharing as a means to describe themselves as faithful and to have nothing to hide. Users tweeted for example: *"I have nothing to hide he can have my password to everything including my phone. #faithful!"*; or *"?If I had a boyfriend, he'd know my password to my phone. I have nothing at all to hide"*.

In general, peoples' tweets would present a snapshot into their personality, presenting them for example as stubborn, unlucky, cool, and most commonly, as funny (n = 37, 5%). For example: *"I do have a life outside of Twitter, but I can't remember the password for it"*. Furthermore, some Twitter users specifically exploited the particular affordances and qualities of (online) passwords, their character composition and length, or the assumptions that people commonly have of them as something that should be kept private and secured from access by others in the telling of jokes. One user for instance tweeted the following: *"I was choosing a password for my new computer last night, I tried LiverpoolFC but apparently it was too weak"*; and another one posted: *"What's Forest Gump's password on Facebook? 1forest1. aha get it."*.

6. DISCUSSION

Over the past twenty years, digital technology and its associated security mechanisms have diffused into almost all aspects of people's lives which has considerable implications for how security technologies must be designed and studied [9]. After identifying social media as a potential resource of security-related experiences, we analyzed posts on the topic of the password due to its status as a security technology that has resisted a concerted research effort to find usable and secure alternatives [13]. The research we have conducted raises a number of discussion points around how studying personal experiences on social media sites can contribute to a better understanding of the design challenges facing security technologies.

6.1 Passwords as a Social Currency

Our main findings were around passwords being appropriated as a social currency, where people were often thinking carefully about passwords and how the maximum value (both pragmatically and socially) could be derived from that password. We saw how the inherent value of passwords made them an item fit to be protected, shared or sought after in a social circle. The need to protect passwords was demonstrated in no small part through users' expressions of frustration at password loss (e.g. *"#sadface"*, *"#Annoying"*), and relief at rediscovery (e.g. *"Wheeeww I thought I was going crazy"*). The sharing of a password with another person can be a powerful act in defining a relationship (e.g. as trusting, as being close). Social expressions of trust were defined by the sharing of the password that regulated access to certain accounts or services, and were ended through the revoking and resetting of that password if a violation of this trust relationship occurred. This form of 'access control' is enabled through the affordances of passwords as they currently are. Twitter was often used to broadcast or discuss any violation of a trust relationship, and various memes that circulate in tweets shed light on how a person might like to respond in such circumstances.

Conventional wisdom considers that security is a secondary concern for users [30, 37]. This is typically true, however recent work has shown how passwords can be foregrounded in relationships where the sharing of credentials can serve functional purposes due to e.g. disability [8] or geographic isolation [32]. The trend of social hacking is particularly interesting, and is likely to be facilitated due to the increased user awareness of the limitations of password choice, but also the large number of opportunities that exist today to compromise passwords in some way. These activities are also less likely to be interpreted as being deviant due to the lack of technical sophistication required to achieve such compromise e.g. displays are increasingly portable and high resolution, which permits shoulder surfing; devices can even be surreptitiously accessed shortly after a user has authenticated to a device; or people may forget to logout while checking emails on the device of a friend. Exploiting these situations can simultaneously provide friendly feedback that some security practices leave something to be desired, but creates a learning feedback of how accounts can be compromised in the social domain if that password are not appropriately managed.

Security researchers often design systems that must impede opportunities for users to perform certain actions. Traditional conceptions of passwords involved the assumption that they were strictly personal and not to be shared [23]. Our findings reinforce that technology designers should be mindful of the ways that people appropriate security mechanisms when designing systems that may aim to control behaviors; a goal that may contribute to the reduction rather than the increase of security.

6.2 Using Twitter for Understanding Security Experiences

The tweets we collected provided us with a snapshot into a wide range of very practical difficulties that people encounter in their daily routines around the use, set-up and maintenance of their passwords as well as a glimpse of their emotional responses. While in this paper we have focused upon tweets that concerned the use of passwords, it is likely that other areas of security and privacy are also discussed prominently on Twitter and could be studied in a similar way. The most suitable issues are likely to be those that are relatively common in everyday life, yet are ever-changing in how they manifest; for example: phishing [16], identity theft, email account hijacking [31], and computer viruses. Each of these areas has social engineering and deceit at its core, and studying tweets can likely provide a useful window to understand how users make sense of these attacks while they are underway, and how people might recover from them. Future research can verify whether this is the case.

An alternative approach to social media analysis could focus upon timely events in society that have privacy and security implications. Twitter was a particular hub of discussion during the recent heartbleed [4] controversy. Heartbleed was the name given to a bug discovered in the OpenSSL library that underpins much secure communication on the Internet. One piece of advice that emerged from this controversy was the need for users to change all of their passwords. At that time people turned to Twitter and used the hashtag #heartbleed extensively to complain about the possibility of changing passwords, and to seek more information from the crowd about the problem. This could lead to analyses of how attitudes to security and privacy change after significant events, or simply change naturally over time.

Recent work focused on security stories [29] and folk models [36] reinforce the need for a sustained focus on how people make sense of security technologies; it is possible that social media analysis can complement that research agenda and provide an accessible resource of stories for designers wishing to gather insights for technology design or simply understand existing behaviors better.

6.3 Challenges of using Twitter as a Research Method

Although tweets are restricted to only 140 characters in length and therefore require users to keep their statements brief, our analysis has shown how even short textual accounts provided insights into the practices and manifest experiences that people have with regard to passwords. Tweets however are limited in the extent to which they offer contextual background. Greater attention to capturing 'retweets' and 'replies' could alleviate this problem, but this issue mainly arises from the general absence of rich social cues online that are inherent to face-to-face rather than online mediated communication; this presented a challenge at times for the researchers to make sense of the tweets, some of which therefore had to be excluded from the qualitative analysis.

However, as intended for our present study, a manifest analysis of tweets provided us with a sense of the many pervasive and openly communicated (albeit often ignored in design and policy within the field) relational, social and personal factors that are at play and apparent in peoples' experiences of passwords. We regard this outcome as a first step, on which to build more in-depth and contextually richer future research. In this regard, Twitter itself may even become a useful vehicle to help identify potential research participants that have demonstrated through their public tweets certain interests or difficulties in a specific domain. Moreover, brief interviews could even be conducted via Twitter itself. Examples of this have been seen already from politicians (*#askboris* is a hashtag used by the Mayor of London to elicit questions from Londoners).

7. CONCLUSION

As security and privacy technologies increasingly penetrate most aspects of our lives, there is a need to develop research methods that provide a window into the ways that people appropriate security technologies into their everyday lives. It is particularly important to question the ways that security experts expect their technologies to be used, and contribute to the design of new, more experience-centred security mechanisms. In this paper we explored the use of social media as a window into everyday security practices. To do this we assembled a dataset of tweets related to passwords. Our analysis suggests that Twitter is a platform for active discussion around password practices. Our findings shed light on contemporary password practices, and suggest that today, passwords can be considered a social currency that people seek to protect, share and obtain from others. These results have implications for our collective understanding of how people integrate passwords into their lives, and suggest Twitter as a platform where other learnings around security and privacy practices could be focused. Future work can consider how content from tweets can be source of design inspiration and feed into a process of security or privacy design.

8. ACKNOWLEDGEMENTS

This work was supported by the RCUK Digital Economy Hub on Social Inclusion through the Digital Economy (SiDE) (EP/G066019/1).

9. REFERENCES

[1] About twitter. ”https://about.twitter.com/company”. Accessed: 2015-05-30.

[2] A. Bermingham and A. F. Smeaton. On using twitter to monitor political sentiment and predict election results. In *Workshop on Sentiment Analysis where AI meets Psychology (SAAIP)*, 2011.

[3] W. E. Burr, D. F. Dodson, and W. T. Polk. *Electronic authentication guideline*. Citeseer, 2004.

[4] M. Carvalho, J. DeMott, R. Ford, and D. A. Wheeler. Heartbleed 101. *Security & Privacy, IEEE*, 12(4):63–67, 2014.

[5] M. Doughty, S. Lawson, C. Linehan, D. Rowland, and L. Bennett. Disinhibited abuse of othered communities by second-screening audiences. In *Proceedings of the 2014 ACM international conference on Interactive experiences for TV and online video*, pages 55–62. ACM, 2014.

[6] P. Dourish and K. Anderson. Collective information practice: emploring privacy and security as social and cultural phenomena. *Human-computer interaction*, 21(3):319–342, 2006.

[7] M. Duggan and A. Smith. Social media update 2013. *Pew Internet and American Life Project*, 2013.

[8] P. Dunphy, A. Monk, J. Vines, M. Blythe, and P. Olivier. Designing for spontaneous and secure delegation in digital payments. *Interacting with Computers*, 2013.

[9] P. Dunphy, J. Vines, L. Coles-Kemp, R. Clarke, V. Vlachokyriakos, P. Wright, J. McCarthy, and P. Olivier. Understanding the experience-centeredness of security and privacy technologies. In *Proc. of the New Security Paradigms Workshop (NSPW)*, 2014.

[10] B. G. Glaser and A. L. Strauss. *The discovery of grounded theory: Strategies for qualitative research*. Transaction Publishers, 2009.

[11] S. A. Golder and M. W. Macy. Diurnal and seasonal mood vary with work, sleep, and daylength across diverse cultures. *Science*, 333(6051):1878–1881, 2011.

[12] G. R. Hayes. The relationship of action research to human-computer interaction. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 18(3):15, 2011.

[13] C. Herley and P. Van Oorschot. A research agenda acknowledging the persistence of passwords. *Security & Privacy, IEEE*, 10(1):28–36, 2012.

[14] Y. Hwang. Antecedents of interpersonal communication motives on twitter: Loneliness and life satisfaction. *International Journal of Cyber Society and Education*, 7(1):49–70, 2014.

[15] P. G. Inglesant and M. A. Sasse. The true cost of unusable password policies: password use in the wild. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 383–392. ACM, 2010.

[16] M. Jakobsson and S. Myers. *Phishing and countermeasures: understanding the increasing problem of electronic identity theft*. John Wiley & Sons, 2006.

[17] A. Java, X. Song, T. Finin, and B. Tseng. Why we twitter: understanding microblogging usage and communities. In *Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 workshop on Web mining and social network analysis*, pages 56–65. ACM, 2007.

[18] J. Kaye. Self-reported password sharing strategies. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 2619–2622. ACM, 2011.

[19] D. V. Klein. Foiling the cracker: A survey of, and improvements to, password security. In *Proceedings of the 2nd USENIX Security Workshop*, pages 5–14, 1990.

[20] A. D. Kramer. An unobtrusive behavioral model of gross national happiness. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 287–290. ACM, 2010.

[21] K. Krippendorff. *Content analysis: An introduction to its methodology*. Sage, 2012.

[22] V. Lampos and N. Cristianini. Tracking the flu pandemic by monitoring the social web. In *Cognitive Information Processing (CIP), 2010 2nd International Workshop on*, pages 411–416. IEEE, 2010.

[23] S. Mandujano and R. Soto. Deterring password sharing: User authentication via fuzzy c-means clustering applied to keystroke biometric data. In *Computer Science, 2004. ENC 2004. Proceedings of the Fifth Mexican International Conference in*, pages 181–187. IEEE, 2004.

[24] J. McCarthy and P. Wright. Putting ’felt-life’ at the centre of human-computer interaction (hci). *Cognition, Technology & Work*, 7(4):262–271, 2005.

[25] J. McCarthy and P. Wright. *Technology as Experience*. The MIT Press, 2007.

[26] M. R. Morris, S. Counts, A. Roseway, A. Hoff, and J. Schwarz. Tweeting is believing?: understanding microblog credibility perceptions. In *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work*, pages 441–450. ACM, 2012.

[27] D. A. Norman. *Emotional design: Why we love (or hate) everyday things*. Basic books, 2004.

[28] M. J. Piskorski. *A Social Strategy: How We Profit from Social Media*. Princeton University Press, 2014.

[29] E. Rader, R. Wash, and B. Brooks. Stories as informal lessons about security. In *Proceedings of the Eighth Symposium on Usable Privacy and Security*, page 6. ACM, 2012.

[30] M. A. Sasse, S. Brostoff, and D. Weirich. Transforming the ’weakest link’- a human/computer interaction approach to usable and effective security. *BT technology journal*, 19(3):122–131, 2001.

[31] R. Shay, I. Ion, R. W. Reeder, and S. Consolvo. My religious aunt asked why i was trying to sell her viagra: experiences with account hijacking. In *Proceedings of the 32nd annual ACM conference on Human factors in computing systems*, pages 2657–2666. ACM, 2014.

[32] S. Singh, A. Cabraal, C. Demosthenous, G. Astbrink, and M. Furlong. Password sharing: implications for security design based on social practice. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 895–904. ACM, 2007.

[33] A. Smith and J. Brenner. Twitter use 2012. *Pew Internet & American Life Project*, page 4, 2012.

[34] M. Thelwall, K. Buckley, G. Paltoglou, D. Cai, and

A. Kappas. Sentiment strength detection in short informal text. *Journal of the American Society for Information Science and Technology*, 61(12):2544–2558, 2010.

[35] A. Tumasjan, T. O. Sprenger, P. G. Sandner, and I. M. Welpe. Election forecasts with twitter: How 140 characters reflect the political landscape. *Social Science Computer Review*, page 0894439310386557, 2010.

[36] R. Wash. Folk models of home computer security. In *Proceedings of the Sixth Symposium on Usable Privacy and Security*, page 11. ACM, 2010.

[37] A. Whitten and J. D. Tygar. Why johnny can't encrypt: A usability evaluation of pgp 5.0. In *Usenix Security*, volume 1999, 1999.

[38] P. Wright and J. McCarthy. Empathy and experience in hci. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 637–646. ACM, 2008.

[39] P. Wright and J. McCarthy. Experience-centered design: designers, users, and communities in dialogue. *Synthesis Lectures on Human-Centered Informatics*, 3(1):1–123, 2010.

[40] M. E. Zurko and R. T. Simon. User-centered security. In *Proceedings of the 1996 workshop on New security paradigms*, pages 27–33. ACM, 1996.

“I’m Stuck!”: A Contextual Inquiry of People with Visual Impairments in Authentication

Bryan Dosono
Syracuse University
bdosono@syr.edu

Jordan Hayes
Syracuse University
jhayes05@syr.edu

Yang Wang
Syracuse University
ywang@syr.edu

ABSTRACT

Current authentication mechanisms pose significant challenges for people with visual impairments. This paper presents results from a contextual inquiry study that investigated the experiences this population encounters when logging into their computers, smart phones, and websites that they use. By triangulating results from observation, contextual inquiry interviews and a hierarchical task analysis of participants’ authentication tasks, we found that these users experience various difficulties associated with the limitations of assistive technologies, suffer noticeable delays in authentication and fall prey to confusing login challenges. The hierarchical task analysis uncovered challenging and time-consuming steps in the authentication process that participants performed. Our study raises awareness of these difficulties and reveals the limitations of current authentication experiences to the security community. We discuss implications for designing accessible authentication experiences for people with visual impairments.

1. INTRODUCTION

Logging into a website with usernames and passwords (i.e., authentication) is an essential part of users’ everyday Internet activities. However, this mundane operation can be daunting for users with disabilities. While many users can input a username and a password (their “login credentials”) to verify their online identities with relative ease, users with disabilities contend with challenges that may prevent them from experiencing a straightforward login process. In this paper, we focus on users with visual impairments. We seek to illuminate their challenges to portray the experiences of these users and raise awareness of current technology limitations that may inhibit them from taking full advantage of these technologies.

We conducted a contextual inquiry to understand the difficulties users with visual impairments encounter when using their computers, mobile phones, and the Internet. Our participants reported their experiences and opinions using different authentication mechanisms, such as passwords and biometrics. Participants experienced the most difficulty authenticating due to inaccessible design within the systems they were using. We found that many websites bury their authentication forms beneath cluttered graphics, flash advertisements and a myriad of other web elements. Encountering these unnecessary elements further prolonged their ability to successfully locate the authentication area on a webpage. Assistive technologies like screen readers offered limited options for users to receive appropriate feedback regarding the degrees of accuracy and success when entering in their login credentials.

These system limitations significantly inconvenience users with visual impairments. They lead participants to experience significant lags and frustration when attempting to authenticate to the services they enjoy when using their computers. As a result, users are required to explore several alternative strategies such as using keyboard shortcuts to navigate their way around cluttered website designs to compensate for poor design.

This paper makes three main contributions. First, we discover specific difficulties users with visual impairments experience in a wide range of authentication scenarios as well as how they mitigate these challenges. Second, we reveal limitations of current authentication systems. Some of these limitations were related to web accessibility issues, which, to our knowledge, have not been systematically examined in the context of authentication. Third, we provide concrete recommendations towards making authentication experiences more accessible.

2. RELATED WORK

Authentication ensures that users are who they claim to be. There exist numerous types of authentication mechanisms in use within today’s security systems. Research and development in identity management [1] categorize modern authentication schemes into three main types: knowledge-based, token-based and biometric authentication systems. Each authentication scheme comes with its strengths and weaknesses. At the moment, however, no single authentication method satisfies the needs of all users, especially considering the wide range of conditions that users may have.

Cassidy et al. researched haptic ATM interfaces for assisting visually impaired users and reported that audio-assisted systems reduce users’ awareness of environmental sounds, meaning that users are less likely to hear someone come up behind them, which increases their vulnerability to

Symposium on Usable Privacy and Security (SOUPS) 2015, July 22-24,

potential attackers [2]. Braille labels and keyboards provide limited tactile feedback to blind users due to the small density of information they can encode [3]. While this is true, not all users with visual impairments utilize Braille or know how to use it well [4]. Emerging technology, such as brain computer interfacing systems, are highly dependent on outside factors such as background noise and the health condition of the individual user [5]. Due to the delicate balance between usability, accessibility, and security in designing authentication systems, adding one modality to user interfaces may affect their usability and can increase the resulting complexity of these systems [6].

A number of papers related to accessible authentication research examine the needs of authentication and proposed technologies to support blind users [7]. In Azenkot's study of 13 blind smartphone users, most participants were unaware of or not concerned about potential security threats [8]. Ahmed et al. conducted an exploratory user study with 14 visually impaired participants to understand how new technologies such as Google Glass may be able to help protect their privacy [9]. The findings of this study show that forced dependence on others, especially strangers was a reoccurring privacy risk. Although low-cost wearable and mobile computing are likely to drive even more advances in accessible authentication [11,15], the unique privacy and security needs of blind users remain largely unaddressed.

Visually impaired users run into problems when interacting with the web. Borodin et al. highlights browsing strategies that they observed screen reader users employ when faced with challenges, ranging from unfamiliar web sites and complex web pages to dynamic and automatically-refreshing content [12]. However, they have not attempted to quantitatively evaluate the effectiveness of employing these strategies. User interface design for effective security remains an ongoing problem [13] and current authentication schemes are not usable enough for those with vision impairments. Relating tactics to technical problems and coping situations allows researchers to understand how users with visual impairments manage undergoing problematic situations [14]. For example, audio CAPTCHAs (Completely Automated Public Turing test to tell Computers and Humans Apart) were introduced as an accessible alternative for those unable to use more common visual CAPTCHAs. A study of more than 150 participants demonstrated that existing audio CAPTCHAs were more difficult and time-consuming to complete compared to visual CAPTCHAs for both blind and sighted users [15].

We did not find any empirical studies investigating concrete challenges and difficulties in authentication for users with visual impairments. In response to a dearth of literature that documents computer and web authentication experiences of these users, our work shares in-depth accounts from the perspective of participants through a contextual inquiry approach. These users shed novel insight into the various types of authentication challenges that designers and developers should consider and address in creating accessible authentication experiences.

3. METHODOLOGY

3.1 Contextual inquiry

To understand how people with visual impairments use computer systems and authentication mechanisms in their natural environment, we adopted a contextual inquiry approach by which we visited participants at places where they regularly used computers or mobile devices (e.g., home, workplace, public library). This approach consists of three main components of gathering qualitative data [16]. First, researchers observe and talk with users in the settings where they perform their everyday tasks. Second, researchers establish a mutual understanding with the user to examine the topic at hand. Acknowledging the user as the expert clarifies that the researchers did not come to solve problems and answer technical questions, which saves the researchers from misinterpreting actions [17]. Third, researchers guide the contextual inquiry on a clearly defined set of participants' concerns, allowing room for conversation to extend beyond a list of specific questions.

We began our contextual inquiry with a semi-structured interview by asking participants a series of questions to understand their computer and Internet use as well as their knowledge and perception of authentication systems. We then asked them to perform a set of authentication tasks. We first asked participants to log into their computer, second their primary email account, third their online banking account or an e-commerce site they use, fourth their social media network of choice and fifth their mobile phones. These tasks were chosen because they cover a diverse set of common authentication scenarios. We also told participants that they could skip any of these tasks if they do not feel comfortable. We encouraged participants to think aloud during these tasks. We audio and video recorded how they performed these tasks with their permission. We conducted each study session with at least two researchers: one leading the study and another taking notes and recording the session. To protect their privacy, we turned the camera away from the keyboard and focused the camera on the device's screen any time they logged in with their credentials. We did not ask them to reveal their usernames and passwords to us during the study. The script we used for each session is included in the Appendix (Figure 7). Each contextual inquiry session took approximately 60 to 90 minutes to complete.

We compensated participants with $30 in cash. We also rewarded participants an additional $10 payment for any extra referrals that completed the study. Our Institutional Review Board (IRB) approved the study.

Table 1: Demographic information of participants and their measured time of logging into various domains of authentication.

	Participant Characteristics					Timed Attempt at Authentication in Seconds					
ID	**Age**	**Sex**	**Occupation**	**Self-description**	**Assistive Tech**	**Computer**	**Email**	**Banking**	**Commerce**	**Social Media**	**Mobile Phone**
P1	50-60	M	Librarian	Blind	JAWS	271	351	376	N/A	86	N/A
P2	40-50	F	Sales	Low Vision	None	N/A	65	N/A	62	40	192
P3	40-50	M	Instructor	Low Vision	ZoomText	78	123	N/A	N/A	N/A	N/A
P4	50-60	M	Banker	Blind	JAWS	12	49	N/A	N/A	N/A	N/A
P5	50-60	M	Retired	Blind	JAWS	N/A	215	N/A	N/A	N/A	N/A
P6	50-60	M	Veteran	Low Vision	ZoomText	229	67	N/A	N/A	N/A	N/A
P7	50-60	F	Retired	Blind	JAWS	N/A	127	58	400	N/A	N/A
P8	50-60	M	Sales	Blind	JAWS	396	37	N/A	223	N/A	10
P9	50-60	F	Instructor	Blind	JAWS	154	11	263 (failed)	N/A	N/A	10
P10	50-60	M	Retired	Blind	JAWS	164	39	N/A	N/A	N/A	N/A
P11	50-60	F	Instructor	Blind	JAWS	33	33	308 (failed)	129	N/A	5
P12	50-60	M	Lawyer	Low Vision	JAWS	254	43	N/A	N/A	N/A	8
					Mean	177	97	316	174	63	54
					Median	164	57	308	129	63	10
					Std. Dev.	124.4	98.0	56.9	142.7	32.5	92.2

* Note: N/A (not applicable) indicates that the participant does not own either a relevant device or an account to authenticate.

3.2 Participant recruitment

From May to July 2014, we recruited 12 participants who self-described as having a visual impairment, including eight blind users and four with low vision. We conducted all study sessions face-to-face. Table 1 describes their demographics. In summary, eight males and four females with an estimated age range of 40-60 volunteered to participate in the study. Three participants reported being retired while eight reported they were still employed and one reported being a veteran. Nine participants used the JAWS (Job Access With Speech) screen reader as their preferred assistive technology while two used ZoomText and one participant did not use any assistive technologies at all. We reached a point of saturation where no new themes emerged after our tenth contextual inquiry session. The remaining two participants confirmed the results.

We recruited participants in the Syracuse, NY metropolitan area via online discussion boards, mailing lists, flyers, YouTube videos, online advertisements and newsletters affiliated with local disability organizations. We also volunteered in local events to gain familiarity with local disability communities. Due to the nature of the study, we found recruiting participants a challenging task. In this vein, we recruited ten of our participants via word of mouth and relied on snowball sampling techniques to recruit from among their acquaintances. We directed potential participants to a recruitment survey that asked respondents to self-describe their disability statuses and we then selected respondents accordingly (see Figure 5).

3.3 Content analysis

We analyzed data collected from each participant by reviewing from each session the transcribed interview and video observation components. We segmented each transcript according to the various parts of the contextual inquiry and proceeded to develop an open coding scheme to generalize the key findings each participant contributed to the study using a grounded theory approach [18]. Seeking to highlight difficulties most salient to authentication, we probed into the various opinions, common practices and difficulties participants encountered when logging into their accounts; their reasons for or against password-protecting their computers and other accounts they used; whether or not they automatically save their login credentials; their willingness to give their login credentials to others; the mental models they conceptualize when creating personal usernames and passwords; their general difficulties using computers and navigating the web and difficulties they encountered when performing the tasks of logging into the computers and accounts they use.

We developed approximately 15 qualitative codes to summarize the most relevant findings we learned from each participant, which we clustered into sets of high-level themes. We also timed attempts of all authentication tasks to get a sense of how time-consuming they were for each participant (see Table 1). Not all participants performed every task as some of them did not use social media or own a mobile phone. We also reviewed videos captured of each participant and noted the actions they took, the visual output observed on the device interface and any voice feedback from assistive technology.

3.4 Hierarchical task analysis

We recorded participants' responses in both audio and video formats with their permission, cleaned up and organized notes from observations and interviews, transcribed audio recordings, coded qualitative data for inductive content analysis [19] and grouped reoccurring themes.

Table 2. Summary of difficulties when participants performed the login tasks, including the source of each difficulty, the average amount of time taken by each participant, and the total number of occurrence for each difficulty.

ID	Difficulty	Source of Difficulty	Average Time (seconds)	Standard Deviation	Total Occurrence
D1	Locating the authentication area on a web page	Accessibility	87	92.1	13
D2	Determining if another user is already logged in on a shared computer	Authentication	133	0.0	1
D3	Waiting for screen reader output to either start or finish speaking in order to find desired information quickly	Accessibility	35	25.3	72
D4	Attempting to verify successful authentication	Accessibility & Authentication	79	49.0	3
D5	Entering passwords correctly due to design of screen reader software	Accessibility & Authentication	14	4.9	2
D6	Receiving insufficient audio feedback from JAWS about error messages	Accessibility	89	33.8	3
D7	Proper finger placement over fingerprint recognition system	Authentication	11	0.0	1
D8	Determining if mobile browser successfully stored login credentials	Authentication	16	0.0	1
D9	Encountering unexpected distractions (i.e. pop-ups, dialog boxes, new windows) while attempting to authenticate	Accessibility	33	0.0	1
D10	Answering security questions correctly	Accessibility & Authentication	166	0.0	1

While analyzing the data following the contextual inquiry process, we noticed some participants who failed to complete some of their login tasks or took a relatively long time to complete them (see Table 1). To help pinpoint which aspects or steps of the authentication process that are time-consuming and/or challenging, we conducted a hierarchical task analysis [20] to identify the steps that were taken and what actually went wrong in those circumstances. For each authentication task, we began by watching the respective video and listening to the audio recordings to identify the steps the participant took to complete the task. We created a high-level task flow diagram from the steps we identified while reviewing the relevant parts of the audio and/or video recording. We then broke down the high-level steps in need of further analysis into one or more separate, more detailed flowcharts in the same diagram with the goal of outlining the specific sub-steps the participant took to complete the higher-level steps. To triangulate different sources of data, we included example quotes and comments from the observation notes and interview transcripts that were relevant to the sub-steps. We also noted the time, in seconds the participant took to complete each step and sub-step on the diagram to understand how time-consuming they were. A few example task flow diagrams are included in the Appendix (Figure 7).

4. RESULTS

By triangulating data from the observations, interviews, and task analysis, we identified a number of difficulties (see Table 2) our participants faced in their authentication experiences. Furthermore, our findings show most of these difficulties can be attributed to a general lack of knowledge and experience of the websites and assistive technologies (e.g. screen readers) they are using, as well as the way in which web designers and software developers have implemented such technologies. Next we discuss these difficulties in detail.

4.1 Locate or identify login elements on page

Participants expressed confusion and frustration over where to find the appropriate area on a web page to log into their accounts. We found this process to be significantly time-consuming and hinder participants' abilities to access the secure services that can only be accessed via successful authentication. P1 felt that websites should be designed to include the most critical information, such as the login area on the top of the page in order for users to quickly access relevant information as soon as the page loads. Results of the hierarchical task analysis also revealed that P7 struggled the most while attempting to identify the area she needed to authenticate into the PayPal website (See Figure 7E in Appendix). She completed this step in 73.5 seconds. P7 was unsure whether she needed to find a link or button to log in, describing her actions and explaining her confusion by saying: "I didn't know it was a button. I thought it was a link, so, that's the trick. If you, if you don't find it one way, you look for it another way." Her resourcefulness showed her ability to adapt to different web interface environments and use alternative methods if her original plan does not work successfully. P7's PayPal task shared similar characteristics with her email task as well as P1's email task in that all three tasks involved spending the most time locating the login area. She depended on using a keyboard shortcut to list all of the links on the page in her email task helped her identify the clickable and interactive webpage elements. She relied on past visits to the website to find the authentication area as a link and then was confused as she realized the login element was actually a button. P7 used her instinct to try and find any authentication-related links and was puzzled as to why she couldn't find any. For example, she became perplexed while locating links beginning with the letter 'L' but no links saying 'log in': "No, that's not there, either…oh, let's see…" Furthermore, P7 expressed the same frustration while failing to find any links beginning with the letter 'S' related to 'sign in': "it's not here, I don't know why not."

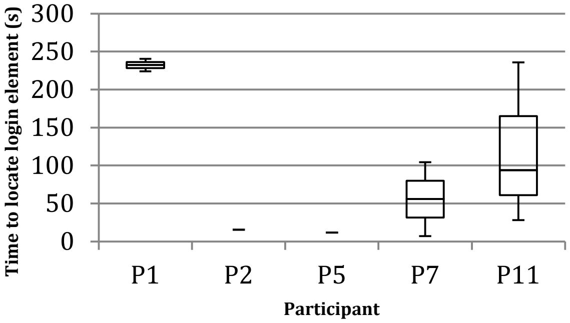


Figure 1: The distribution of time participants spent while attempting to locate the authentication area on a page. All other participants either did not complete certain tasks, or they completed the tasks but did not need to locate the authentication area (e.g., log into a computer), or no video recordings were available to depict them performing such tasks, and therefore are not shown on this graph.

Participants expressed confusion over which “Sign In” or “Log In” buttons or links to use because multiple buttons or links of the same type are placed on one web page. P11 entered the URL of the bank directly into the Google search bar, as opposed to entering search terms. P11 spent the most time struggling in an effort to identify the sign-in link for her bank's website on the Google search results page. Each attempt to locate her bank’s “Sign In” link from the Google search results page took 140.5 seconds and 100 seconds respectively to complete, totaling 240.5 seconds. Both of P11’s attempts to find this link produced no luck. She may have unintentionally skipped the link as the speech output reported, *“this browser does not support inline frames”* right before announcing its existence of the link. P11 cut off the second word and continued to press the Down Arrow key to sift through the rest of the search results without catching it.

During her second attempt, she became more frustrated as she continued down the search results page, still not being able to find the desired link: *“Come on...why doesn’t it ask me to sign in? It wants me to get into the Rewards thing, you know? I’ve gotta find out...*” P11 continued to express disgruntlement during her second attempt as the screen reader identified every other link belonging to her bank’s website except for Sign In: *“wants me to follow on Twitter, and yada, yada... [sighs]... Yep, they’ve changed this. Uh, let’s see.”* P11's failure to successfully authenticate into her online banking site can also be attributed to confusion over both the layout of the search results as well as which service she was actually logging in to use. She frustratingly skimmed through the search results after encountering an unfamiliar link and noticed, *“they must have changed the way it was set up since I last used it.”* P11 found a "Sign In" link on the page, which she perceived to be that of her online banking site, yet in reality the link directed her to a Google Account settings page. She realized this upon hearing her screen reader prompt her to log into a different service altogether with an account she does not have: *“I don’t understand. I should have an [online banking] account, not a Google account.”*

Participants spent an average of 96 seconds attempting to locate the authentication area of the web pages they accessed, as shown in Figure 1. The hierarchical task analysis results show that this step alone was the most time-consuming part of the authentication process for three participants: P1, P7 and P11. Not all participants are included in Figure 1 because some participants did not locate an authentication area on a webpage while performing their tasks or no video recordings were available for the research team to determine the completion time.

4.2 Logging in as another user

One participant struggled to determine whether or not another user was already logged into Gmail on the public library’s Dell desktop computer he was using before he could locate the authentication area (see D2 in Table 2). According to the hierarchical task analysis, P1 inferred that another user had previously logged into this computer. However, he needed to find the name of the other user to confirm and finally did so after frustratingly combing his way through the Gmail Sign-In page in an effort to locate the other user’s account: “*OK, there it is... so that’s her email*.” After finding the name of the other user, P1 then struggled to find the button he needed to log into his own account because he was unsure of the terminology used to describe the login area: *“sometimes it’s ‘log in as another user’, sometimes it’s ‘sign in as another user’, sometimes it’s ‘change user’.”* He feels constantly changing the terminology of login elements introduces a new learning curve regarding how to locate the authentication area quickly and efficiently: *“unfortunately, this is somethin’ that we run into a lot, is, you don’t know what they call things, and every time they update the website, you have to re-learn how to do it.”* Standardization of terminologies would greatly aid users of screen-reader technology.

4.3 Delay in finding necessary information

By default, screen readers such as JAWS read contents of a web page in a linear fashion, beginning with text located at the top and then gradually moving down to the bottom of the page. Websites are generally designed with authentication forms placed closer to the center of the page beneath a considerable amount of graphics and text. This practice posed a significant challenge to participants who depend on screen readers to access the information they need quickly. Users with visual impairments are further impeded from efficiently accessing the authentication area they are attempting to locate because they must weave through a complex layout of webpage elements to access the login form.

We found our participants used an array of keyboard shortcuts to cut down on the time waiting for screen readers

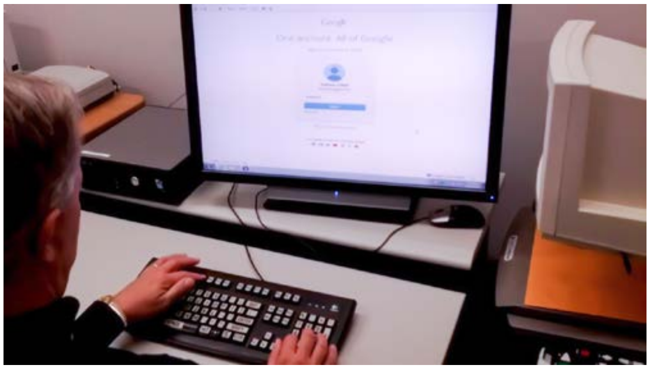


Figure 2: P1 diligently continued to troubleshoot through an authentication error by finding an alternate way to log into his email account using a variety of keyboard shortcuts.

to identify the information they need (as illustrated in Figure 2). However, these shortcuts do not always work for them and can sometimes lead to additional obstacles inhibiting their ability to authenticate.

According to the hierarchical task analysis results, participants spent an average of 39 seconds waiting for the JAWS speech output to start speaking, finish reading all the elements on a web page or both as shown in Figure 3 and D3 in Table 2. This waiting period significantly added to the total completion time of each task. For example, P7 took the most amount of time just waiting for JAWS to read the information she needed to perform the necessary steps to log into her PayPal account (see Figure 7D in Appendix). While waiting for the PayPal page to load after entering the URL, P7 waited and listened for the presence of any buttons, links or text. After she entered the URL into the Open dialog box in the Internet Explorer browser, the JAWS output read: *"Search the catalog,"* indicating the browser had not left the home page yet. P7 then indicated the page was taking a little extra time to load than usual: *"OK, it hasn't loaded yet...should load."* She eventually remarked*: "Oh, we are loaded"* after pressing multiple keyboard commands to obtain information from the screen reader, confirming she was now on the PayPal home page. This entire process lasted for a total of 25 seconds.

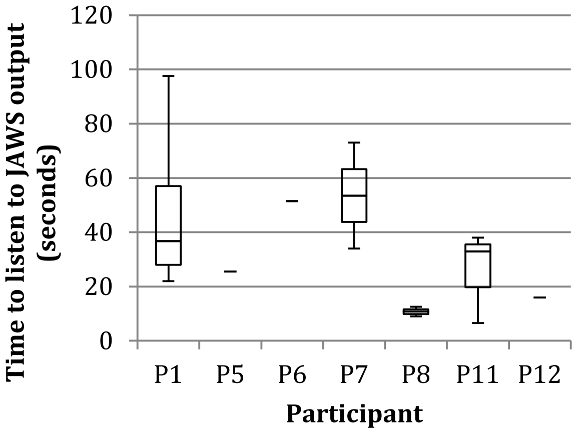


Figure 3: The distribution of time participants spent either waiting for their screen reader to begin speaking or listening to their screen reader finish reading web page elements aloud. All other participants are not shown in this graph because they did not use a screen reader when performing their tasks or no video recordings were available.

4.4 Verifying successful authentication

Three participants were uncertain about whether their authentication attempts were successful (see D4 in Table 2). They searched for specific web elements or textual cues to infer their authentication status. The results of the hierarchical task analyses showed that P11 attempted to authenticate into Amazon and encountered account management links usually associated with post-authentication activities that are present even if users are not logged in at all. Some of these links, for example included *"Your Account," "Manage Your Content and Devices," "Manage Your Cloud Subscriptions," "Your Games and Software Library"* and *"Your Watch List."* This process of locating the "Sign In" link from the search results after typing the URL into the Google search bar took her 56 seconds to complete. P11 was confused as to why there was no "sign in" link in the search results and assumed Amazon had already recognized her credentials: *"See, it put me already right into, uh… this isn't helping you, because it must have remembered my password, which I was very willing to enter in."* She was unsure when the screen reader announced a link in the Google search results that she heard called "Try Prime Cart" (this is actually a combination of two links, one inviting users to evaluate Amazon's premium subscription content service called Prime and another for the user to manage his/her shopping cart) after hearing the "Shop by Department," "Sign In" and "Your Account" links. P11 curiously selected the link to find out what it was but continued to stray further away from her desired destination: "*'Try Prime Cart'? I don't know what that is. Let's see."* From the Amazon pages, she continued to express frustration and confusion as she encountered unnecessary links for managing her Amazon account such as *"Your Amazon Music Settings," 'Your Video Library" and "Your Games and Software Library,"* rather than ways to authenticate into the website: *"I don't want that right now. I wanna sign in for you."* P11 ultimately gave up her attempt at authenticating into Amazon, expressing her ultimate confusion as to why she couldn't successfully log in: *"I don't know why I'm not getting into the 'Sign In' thing."*

P1 did not encounter any account management links associated with post-authentication, but was unsure whether he successfully authenticated into Gmail after eventually finding the fields necessary to do so. He inferred successful entry of his login credentials when he browsed around the user interface of the actual Gmail client. When P1 found his email address on the Gmail page after submitting the login form confirming successful authentication, he expressed:

"Yes, it did. It logged me in." This entire process took him 24 seconds to complete, which added to the total completion time of 351 seconds for this task.

P7 shared similar difficulties along with P1 in terms of self-validating her successful attempt at logging into her PayPal account (see Figure 7F in Appendix). This step took the longest for P7 to complete, totaling 118.5 seconds. She attempted to locate her name on the page that loaded after entering her credentials and remarked about the amount of time taken to find the information she desired: *"huh, that's not what I want...must take a while to load. Sometimes it does."* When failing to find her name on the screen, P7 gave up on her own efforts and asked the research team to confirm for her whether or not she had successfully completed this task. She asked to start over before making a decision whether or not to actually repeat the process of authenticating into PayPal, which she ultimately decided against, since Researcher 2 had notified her of a successful login. Unsure of this fact, P7 asked him a second time and Researcher 2 again reassured her successful login.

4.5 Limitations of assistive technology

4.5.1 Password masking

Our users depended on JAWS to assist them with using their computers to navigate through elements on any given web page. However, these screen readers did little to ameliorate their authentication experiences (see D5 in Table 2). For example, P1 expressed frustration at how JAWS verbally concealed passwords as he and other users he assists type them into the field: *"[As a librarian], I show the public how to log into websites and how to do searches, and they're sure that they've typed it in right but all they hear when they type is the screen reader say 'star, star, star,' so they don't know if they hit the wrong key, or if the caps lock key happened to be on or something. They don't know."*

This design choice was purportedly made to prevent shoulder surfing attacks (i.e., someone standing next to the user and overhearing the password). However, P1 had no way of confirming whether or not he correctly typed in the password until either a verification or error screen propagated from the field submission. To accommodate for this difficulty, P1 suggested the following design modification of screen readers such as JAWS: *"Give people options. If they want to mask the password, then they can choose to do that, but if they don't want to, if there was a checkbox that you could check and say, 'don't mask the passwords for me logging in,' so then you could hear it and know if you did it right or not. That would make it easier."*

4.5.2 Lack of feedback using case-sensitive passwords

One participant was also mystified when trying to determine the correct capitalization for entering in case-sensitive credentials. The confusion contributes to whether or not users mistyped their usernames and passwords. P3 expressed uncertainty in figuring out whether or not she enabled the caps lock function on her keyboard. Even though she activated ZoomText, an accessibility software application that enlarges everything displayed on a computer screen with increased clarity—to assist her with navigating her computer, the screen reader portion of ZoomText does not indicate to her the case of the letters she typed. P3 is concerned about her uncertainty when entering in her passwords: *"I'll try two or three times. Sometimes, I'll lock myself out, 'cuz I don't see that right away."*

4.5.3 Lack of screen reader output for error messages

Participants experienced the most difficulties when attempting to log into their computer systems because they were unaware of an error message that obstructed them from successfully authenticating into these systems (see D6 in Table 2). For two participants, JAWS provided no speech output when the error message appeared on the computer screen. P1 had attempted to enter his credentials and received an error message from Windows stating one or both of his credentials were incorrect. He was unsure of whether or not he was successful after hesitantly entering his password while attempting to log into the computer's Administrator account. He does not normally sign in and out of this computer on a regular basis because the computer he was using is programmed by the IT department to log into Windows automatically upon initial boot-up. While anticipating the available users to appear after initiating the "Switch user" command on the Windows Start Menu, P1 remarked: *"Now, I'm waiting...sometimes the screen reader program reads the new screen automatically, sometimes it doesn't."* As the screen reader indicated the Administrator account was currently selected, P1 confirmed this: *"Now that said 'Administrator account.' Let's see."* After prompting him to enter his password, the screen reader he was using did not read this error message aloud. P1 was unsure what to do because of the silence created from the lack of audio feedback: *"It's not talking to me. So I'm waiting. I'm sitting here thinking, 'OK.' Either it's gonna do something in a few seconds or it's not', but I don't know."* He then desperately used various keyboard shortcuts to elicit some response from the computer, but this trick did not work*: "I haven't got any...it's not talking to me."*

P1 wondered whether the computer logged in or if something went wrong: *"So at this point, I don't know if it switched or not."* He then asked the research team for assistance: *"if you can see, but the screen's on, right?"* He continued to express confusion and began to explore alternate methods of solving his problem by saying, "*So I don't know if that worked. So what I would have to do then would be start over, unless you can see something else for me to click on there.*" P1 attempted to log in again after shutting down and restarting his computer. This second attempt was successful and did not require him to enter in any login credentials because the computer automatically logged in and loaded the Windows desktop. P1 remarked about making this computer as accessible as possible for anyone who uses it by simplifying the authentication process: *"We have such a variety of users that our technical staff tried to streamline things and so they write this little automatic login just for the boot-up part."* Researcher 2 confirmed successful login as did P1, who noted the JAWS output: *"so the screen reader started automatically."* The presence of this initial speech output indicates the computer successfully bypassed the Windows login dialog and loaded the Windows desktop. Each restart attempt took 150 seconds and 83.5 seconds, respectively to complete.

P8's case was similar to that of P1, except he used a fingerprint recognition system to log into his Lenovo laptop computer, attempting to do so eight times before ultimately giving up and authenticating using his username and password instead. This fingerprint-based authentication attempt took him 88 seconds to complete. For a few attempts, he had to take the time to position his finger over the fingerprint reader and was not sure whether or not his first swipe registered. He pointed out: *"well, it didn't do it yet"* after not receiving a response from the computer. P8 seemed to become more frustrated after the following unsuccessful attempts and began to wonder whether or not his placement over the fingerprint reader was a contributing factor to this (see D7 in Table 2): *"I might not be touching it in the right place. I'm never quite sure where to touch it."* P8 continued to express his utmost frustration after three more attempts as he lowered his head, sighed, grunted and explained: *"See, it isn't responding. But if it had…didn't seem to respond. I wonder why. It, maybe, I don't think it's forgotten it."*

The computer did respond after every attempt P8 made to swipe his finger and authenticate, yet this response was an error message displayed on the screen that P8 was not able to see due to a lack of any notification from JAWS. While our video recording did not capture the error message text, it is very likely that the error message returned a visual notification to the user that the computer could not recognize his fingerprint. The error message appeared on P8's computer screen 3.5 seconds after his first attempt and remained there for all attempts following. In addition, a second error message appeared on top of the existing message after four (non-consecutive) attempts and disappeared a few seconds later. JAWS did not provide any speech output as those two messages popped onto the screen and therefore P8 was unaware of such a notification.

Usually, he would power up his computer and JAWS starts up just after the operating system loads and just before the Windows logon box appears on the computer screen. The screen reader announces, *"JAWS for Windows is ready,"* indicating to P8 the machine is ready for authentication. The screen reader provided him with this notification on start-up, but did not give any feedback during his attempts to authenticate using his computer's built-in fingerprint reader. When attempting to log in using text-based credentials, JAWS notified P8 of successful startup and automatically positioned the cursor at the password field since his username had already been filled.

One participant took advantage of password-saving mechanisms like auto-fill features on common Internet browsers. P2 voiced an issue associated with keeping track of multiple credentials: "*It's kind of a pain because I have to remember all these passwords*." There were times when P2 was concerned this browser-saving trick would not always work. P2 used one to help her keep track of her password for the business she manages online. She felt concerned if the browser did not remember her credentials but was ultimately relieved to discover the browser had indeed remembered them (see Figure 4 and D8 in Table 2): *"Let me see. Oh, it does remember me!"* If this were not the case, she may not have been able to log in because she said she was unsure she could remember them herself. In the event the password manager had failed to remember her credentials, P2 referenced alternative password recovery mechanisms, making small modifications to the same base password to create a new one. She used this strategy to her advantage in the event of a forgotten password, sometimes making attempts at a login area to crack her code. She limited the number of times she attempted this stunt in an effort to prevent sites from locking herself out after a number of unsuccessful attempts.

4.5.4 Difficulty with password recovery mechanisms

Other participants we observed showed the most difficulty using screen readers to recover their login credentials in case they lost or forgot those they originally created. P7 provided the most prominent example of this, explaining how screen readers were not always capable of reading new passwords provided by the system via email after attempting to reset her login credentials. P7 referred to instances outside the authentication tasks in our study where she clicked the "lost/forgot password" link on websites. She further explained how users must request assistance from elsewhere to obtain the necessary information: *"You're not always able to get the information on the screen, and you have to get somebody to come in and read you the temporary password. You know, 'H-J-3-9-4-8-*

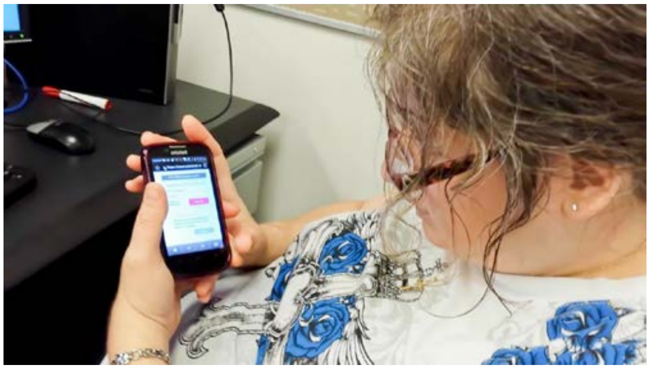


Figure 4: P2 expressed relief when finding out her login credentials were successfully saved into her small business account when using her mobile phone's browser.

4-9-6-9-1,' etc., and then, you got to try to remember it." The entire process of asking others for help, remembering the characters they tell users and attempting to enter in those credentials takes additional effort and adds to the frustration of multiple login failures.

Another overlooked aspect of authentication systems involved the actual terms used to identify specific login credentials. Some users may not understand the difference between the meaning or purpose of a username and a password. P8 provided valuable insight into his mental model he used to distinguish between the two: *"You know, if you called me 'ugly', you know, that would be a name, or 'handsome' or whatever name. So I'm never sure the difference between a user-name and a password. I guess a word is a name, but anyway, but that's my confusion about it."* P8 went further on to clarify what would fit his definition of an appropriate password as he logged into his computer: "I could have said that, OK…if it was '*hound dog', that's a password to me.*" This distinction between user<u>names</u> and pass<u>words</u> is an interesting way of looking at how most sites requiring these credentials may trap users who contemplate specific mental models of what credentials they tend to create for themselves and what each credential means to them personally.

4.6 Other unexpected distractions

During the authentication process, users with visual impairments may encounter additional obstructions that may further hinder their ability to log in either on the webpages or software applications they use (see D9 in Table 2). Navigating around these obstacles continues to add to the confusion and frustration users with visual impairments face. For example, P6 had attempted to locate his email client, which had already been logged in. However, when pressing the TAB key to navigate to the email client, he encountered a Dropbox software application, which was already open on his desktop PC. P6 expressed frustration to the research team about encountering this unexpected obstacle: "*[sighs] I hate that. I actually only got into Dropbox somehow and I'm gonna turn that off.*" Since the research team analyzed this task using the audio recording, the presence of a "yes" or "no" button is unknown. However, upon downloading the Dropbox application and re-constructing this step, the research team can determine the existence of an "OK" and "Cancel" button with the cursor positioned over the latter. There was no mention from either JAWS or P6 regarding the outcome of completing this step. P6 eventually was able to continue and successfully get into his Juno email client. As his computer screen displayed a browser window associated with Juno, P6 can also infer he is where needs to be: "*I think I'm back in my email.*" After pressing ALT + Tab two steps later, the computer screen displayed Juno's main window. P6 confirmed he had indeed accessed the client after hearing the JAWS speech output mention the name of the email client. The research team verbally notified P6 he was automatically logged into Juno 4.5 seconds after he verbally confirmed it himself.

4.7 Hassles authenticating into mobile phones

Six participants owned a smartphone, but only two participants password-protect their devices. They disliked adding an additional layer of security to their mobile phones because most of them cannot see the keys or characters required to authenticate into them. For example, P3 owned an iPhone and used many of its accessibility features. However, when asked if she password-protected her phone, P3 stated: *"No, it is not password-protected, and that's only because I can't see what's on the dim phone screen in bright areas. I won't be able to see it if I'm outside. Like, every time you get a text, you have to put your password in, I get confused."* P3 felt she is at more of a security risk by not password-protecting her phone, referring to the potential risks associated with one of her children having his phone stolen. However, she must accommodate for her visual impairment when password-protecting her smartphone.

5. DISCUSSION

5.1 Reflection of key findings

Our findings illustrate the challenges participants face as a result of accessibility issues hindering them from successfully authenticating into the websites and services they use, while also shedding insightful light on these challenges. In addition, we highlighted the strategies they use to overcome them as reported in the literature we reviewed. We situate these challenges in the specific experiences our participants faced in order to provide novel awareness of how they contribute to the holistic authentication experience. The authentication experience involves numerous stakeholders in the process, all of whom play a significant role in users' efficient and timely access to the services and information they want and need. Finally, our results show that participants take a relatively long time to access the authentication area by struggling to find the login fields themselves and/or waiting for the screen reader to provide them with enough information to proceed. This is significantly longer than the average time taken by the general population to authenticate.

According to Table 2, four of the ten difficulties participants faced arose from general accessibility issues, while three derived from issues related to the underlying authentication mechanisms themselves. Three were associated with a combination of issues related to both. The most common difficulties include: screen readers failing to notify users of error messages; participants struggling to efficiently locate the authentication fields on a web page; participants expressing uncertainty when verifying their attempts to log in; and participants waiting an unnecessarily long amount of time for the screen reader to either start or finish reading webpage elements aloud. With the exception of waiting for JAWS to finish reading the webpage elements aloud and some aspects of inefficiently locating the authentication area, these common difficulties mentioned above create significant barriers to accessing the areas of websites they use that require them to authenticate.

E-commerce websites such as Amazon providing account management links associated with post-authentication such as “Your Account” and “Your Orders” misleads users with visual impairments with a false sense of logging into their accounts. Users who click these account management links are taken to an authentication page where they are supposed to enter their credentials and submit the form. This compounds the frustrating task of finding their way to the website’s authentication mechanism for users with visual impairments because encountering the misleading post-authentication account management links before landing on a login page would trick these users into thinking they have already logged into the site when indeed the opposite is the case. Bonneau, et al. demonstrated how numerous aspects of password implementation lack standardization [21]. In our study, we did not look into actual password implementation, but we did find inconsistency regarding the names of login fields.

Our findings reveal key accessibility-related issues that create significant obstacles not reported in accessibility-related communities such as ASSETS. Additionally, no security literature discusses the difficulties we encountered in the context of authentication. We present this novel, insightful evidence in the form of difficulties that participants experienced in our contextual inquiry study. These difficulties directly impact authentication as well as directly relate to accessibility since these issues related to the design and implementation of web content render authentication systems nearly unusable by those with visual impairments, regardless of the level of security they may provide to their users. The security community must seriously consider these accessibility difficulties and contemplate how the empirical evidence we present here directly corresponds to the usability of authentication systems and mechanisms, similar to the way we critically examine key usability issues we feel relate to security. The most advantageous form of authentication is one that can both be utilized by and accommodate users of all needs, including users with visual impairments.

5.2 Sources of difficulties

5.2.1 Socioeconomic conditions

System designers need to be cognizant of the various socioeconomic hurdles that financially burden users with visual impairments, as they often cannot afford the latest technology on the market. Several participants in our study explicitly stated how they could only afford lower end models of electronic devices and services. In many instances, these devices and services either come with no accessibility support (e.g., a feature phone instead of a smartphone) or corporate providers discontinue support for legacy systems altogether (e.g., Windows XP). Thus, assistive technologies such as screen readers should be backwardly compatible with older operating systems. Furthermore, screen readers such as JAWS are becoming more expensive to purchase. However, the availability of open-source screen reading applications such as Non-Visual Disabled Access (NVDA) is increasing in popularity. This provides reasonable means for users with visual impairments to access to the software they crucially depend on in order to operate their computers without sacrificing any necessary expenses. However, our findings show that none of the participants used any of these open-source screen readers.

5.2.2 Technical learning curves

All participants informed us of the sharp learning curves that came with using a screen reader for the first time. They expressed that it took a considerable amount of patience and practice to use assistive technology efficiently. Users must know which particular elements they want to find and determine their location in relation to their current point of control on the screen. This takes a great mental skill of trial-and-error and reasonable deduction using repetitive up-and-down-arrow keystrokes and actively listening to the auditory output JAWS provides. As a result of these technical learning curves, users with visual impairments may take a significantly long time to perform simple tasks using their computers.

5.3 Implications for design

Based on our findings, we propose four concrete suggestions to address the difficulties our participants faced when authenticating into the systems they use. First, we suggest web designers should improve accessibility to the authentication areas (i.e., login forms). Placing fields for credentials and submit buttons to an easier location on the page closer to the top or changing the code would allow screen readers to say where the login form is located. Placing only one sign-in element on a page at a time reduces the confusion of users locating their desired authentication field and removing any links referencing “your account" also reduces the possibility for users to enter a false sense of being logged in after encountering links associated with post-authentication and multiple points of authentication. Developers of web services should also provide confirmation messages to users with visual impairments indicating the success of their authentication attempts by creating a page or prompt simply stating whether or not users have successfully logged in, which can be launched immediately after users submit the login form on a page. Furthermore, we suggest screen reader developers add an additional keyboard shortcut allowing for users to immediately identify any authentication fields on the page, which immediately takes them to the authentication area. Finally, we suggest the introduction of web design standards regarding consistent terminology related to authentication mechanisms, which will reduce the amount of confusion users with visual impairments may face when trying to locate any authentication area on a webpage. Doing this can potentially reduce this confusion and provide some stability across various websites.

Most of the difficulties we found specifically apply to users with visual impairments. However, some of our suggestions can also apply to the general population of users. For example, developing guidelines related to accessible authentication elements would allow users with visual impairments to quickly find the authentication fields they need while also reducing the time for sighted users to navigate through a cluttered page. A variety of users can also benefit from the concrete assurance from a confirmation page or message notifying them whether or not their login attempts are successful. Taking into account this notion of universal design allows web developers to address issues that may help one specific marginalized population of users overcome these difficulties while also making significant changes that will benefit all users.

Designers of assistive technology should include users with visual impairments as part of the design, evaluation and testing process. They should encourage users who are the most affected by their designs to test their prototypes themselves. This would allow those with visual impairments the opportunity to provide insightful feedback regarding the strengths, weaknesses and potential improvements that could be made. Actively involving users in the design, evaluation and testing process would allow them to better understand their needs and help influence future designs. At the same time, however, designers of assistive technologies as well as web designers should be aware of any security and privacy risks associated with any suggestions made by users with visual impairments before implementing them (e.g., using only security questions as suggested by some of our participants).

5.4 Considerations for alternative authentication practices

Our participants either used or commented on alternative authentication methods that they preferred over the traditional username/password scheme. However, these alternatives are not silver bullets, either.

5.4.1 Using password managers to remember login credentials

Participants can use password managers provided with their browsers to remember login credentials. This mechanism reduces the remembering of multiple sets of usernames and passwords. For example, P2 used one to help her keep track of her password for the business she manages online. If this were not the case, she may not have been able to log in because she said she was unsure she could remember them herself. She relied on browsers and other password-saving mechanisms to help reduce this burden. However, she may increase her vulnerability to hackers and cybercriminals and put herself at risk for identity theft if attackers target the master password. Even if the master password remains safe from such attacks, the original web passwords remain as vulnerable as before [22].

5.4.2 Using biometrics

In order to further reduce the frustration and confusion associated with conventional login systems, most of our participants expressed interest in seeing biometric authentication become more widely used by society. They perceived biometrics would reduce the need of memorizing and entering in their login credentials. P2 felt that authenticating into the systems she used would be easier if she could just *"put [her] hand up to the computer. It's going to know it's [her] and it's going to let [her] into everything."* Other users, such as P7, were more skeptical of using biometric authentication systems because they are *"starting in the wrong direction"* and will become more intrusive as this type of emerging technology evolves. She illustrated, *"Once something like that starts, everybody's going get a chip implant when they're born and they'll know where everybody is all the time."* P12 argues that biometric systems may not work for all users with disabilities, especially those whose natural physical traits have been replaced by artificial ones: *"It could be as simple as having no motor skills or having had your fingerprints damaged as a result of a fire or some kind of body injury. Or if biometrics becomes basically retina scans and somebody has prosthetic eyes, and same with biometrics using fingerprints and somebody has prosthetic limbs. That would be problematic. So you will always have to be able to design systems of authentication that account for the possibility that there will be a subset of the population that can't access everything through biometrics."* Using an ability based-design approach allows systems to adapt to users' needs rather than their disabilities [23]. We should build systems that use this approach so that we work upon users' abilities instead of their disabilities.

5.4.3 Using security questions

Typically, security questions are used as an additional layer of verifying users' identities after entering in their usernames and passwords [24]. They could also be used as a way to replace them as a set of authentication credentials. P7 suggested doing so as an alternative to using passwords for routine authentication. While she provided this alternative to simplify the authentication experience, employing this mechanism creates more security issues than using the conventional practice of using login credentials. For example, answering security questions just swaps out the need to remember one set of information for another (i.e. passwords as opposed to answers to security questions). In addition, security questions are mostly used as a secondary authentication scheme in password-reset situations where users attempt to answer them, and if successful, must enter a new set of credentials.

Screen readers do not mask answers to security questions, as they do for passwords. We observed P9 attempting to answer security questions when attempting to log into her online banking account and noticed JAWS speaking out her answers to the security questions. This poses significant risks to participants because others can use this additional information to gain unauthorized access for fraudulent purposes if they successfully authenticate using these security questions. Supplementing usernames and passwords with answers to security questions ensures users are who they claim to be by providing the system with an additional layer of uniquely identifiable information. This practice, however does pose significant security risks as opposed to using login credentials to authenticate. Users may not easily remember the answers to the security questions they created after not using them in a long time.

5.5 Study limitations

We did not aim to report on a representative capture of all possible variations, but we rather aimed to gain a deeper understanding of analyzed cases. As we were only able to recruit participants who used technology, we did not study those who were afraid of using technology or those who refused to use computers altogether. Our study did not collect data of our participants performing authentication tasks captured from their browsing history. Most of our participants described themselves as living with blindness or having a visual impairment, therefore the generalizability of our findings to other types of conditions was limited. We note that most of our participants are 50 years or older, therefore the difficulties we observed might also be due to their age. It is difficult to disentangle the effect of their visual impairments with age as a confounding factor.

The timings of the authentication tasks that we reported are not intended to be a precise, quantitative measure of the exact amount of time participants took to complete each task as well as their various steps and sub-steps. Instead, these timings are intended to be indicative of which tasks or steps and sub-steps are relatively time-consuming for participants to complete. Calculating the time participants took to locate the authentication area reveals that certain steps are quite time-consuming for many participants, as shown in Figure 1. We note that the timings we measured could be affected by many factors related to individual participants such as their self-described conditions, skills, use of assistive technology, setting in which the computer is being used (i.e., shared public terminal vs. home machine), computer configuration including the hardware and software (i.e., browser) installed on the machine as well as previous knowledge and experience. Since participants sometimes spoke aloud describing what they were thinking or doing during the authentication tasks to the research team, the timings we measured might be longer than if they did not think aloud. Nevertheless, our evidence suggests that a few specific steps, such as locating login area and waiting for screen reader output are particularly challenging and for some users and need to be improved.

6. CONCLUSION

Current authentication interfaces are difficult to use for users with disabilities. This causes frustration and leads to insecure behavior. Our study provides a nuanced account of various difficulties these users encounter with authentication. Our recommendations aim to inform future related research and authentication system design. As the security community actively creates new authentication mechanisms, they should take into account the various characteristics of users and potential challenges they may face. New authentication mechanisms should be fast to use and work well with assistive technologies such as screen readers. We hope the authentication community can use our study insights to make their authentication mechanisms more accessible.

7. ACKNOWLEDGEMENTS

The contents of this paper were developed under a grant from the National Institute on Disability, Independent Living, and Rehabilitation Research (NIDILRR grant number 90DP0061-01-00). NIDILRR is a Center within the Administration for Community Living (ACL), Department of Health and Human Services (HHS). The contents of this paper do not necessarily represent the policy of NIDILRR, ACL, HHS, and readers should not assume endorsement by the United States Federal Government.

We would like to thank our participants for sharing their experiences and insights. We also acknowledge Natã Barbosa, Huichuan Xia, Yun Huang, Kevin Du, Joon Park and other colleagues in the Social Computing Systems (SALT) Lab at Syracuse University as well as Markel Vigo, Mike Just, Jeffery Bigham, Amy Hurst, Aaron Steinfeld, Sonia Chiasson and anonymous reviewers for their feedback and help.

8. REFERENCES

[1] L. Fritsch, K. S. Fuglerud, and I. Solheim, “Towards inclusive identity management,” *Identity in the Information Society*, vol. 3, no. I, pp. 515–538, 2010.

[2] B. Cassidy, G. Cockton, and L. Coventry, “A haptic ATM interface to assist visually impaired users,” *Proceedings of the 15th International ACM SIGACCESS Conference on Computers and Accessibility*, pp. 1–8, 2013.

[3] K. Helkala, “Disabilities and authentication methods: Usability and security,” *2012 Seventh International Conference on Availability, Reliability and Security*, pp. 327–334, 2012.

[4] E. Murphy, R. Kuber, G. McAllister, P. Strain, and W. Yu, “An empirical investigation into the difficulties experienced by visually impaired internet users,” *Universal Access in the Information Society*, vol. 7, no. 1–2, pp. 79–91, 2008.

[5] G. Al-Hudhud, M. Abdulaziz Alzamel, E. Alattas, and A. Alwabil, “Using brain signals patterns for biometric identity verification systems,” *Computers in Human Behavior*, vol. 31, pp. 224–229, 2014.

[6] J. Holman, J. Lazar, and J. Feng, “Investigating the security-related challenges of blind users on the Web,” in *Designing Inclusive Futures*, Springer, 2008, pp. 129–138.

[7] N. Saxena and J. H. Watt, “Authentication technologies for the blind or visually impaired,” *Proceedings of the USENIX Workshop on Hot Topics in Security*, p. 7, 2009.

[8] S. Azenkot and K. Rector, “PassChords: Secure multi-touch authentication for blind people,” *Proceedings of the 14th International ACM SIGACCESS Conference on Computers and Accessibility*, pp. 159–166, 2012.

[9] T. Ahmed, “Privacy concerns and behaviors of people with visual impairments,” *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, 2015.

[10] M. M. Haque, S. Zawoad, and R. Hasan, “Secure techniques and methods for authenticating visually impaired mobile phone users,” *2013 IEEE International Conference on Technologies for Homeland Security*, pp. 735–740, 2013.

[11] G. Kristin and B. Johansen, “e-Me Mobile: Accessible authentication for mobile devices Table of Contents,” *Mobile Information Systems*, 2011.

[12] Y. Borodin, J. P. Bigham, G. Dausch, and I. V Ramakrishnan, “More than meets the eye: A survey of screen-reader browsing strategies,” *Proceedings of the 2010 International Cross Disciplinary Conference on Web Accessibility*, pp. 1–10, 2010.

[13] A. Whitten and J. D. Tyger, “Why Johnny can’t encrypt: A usability evaluation of PGP 5.0,” *USENIX Security*, 1999.

[14] M. Vigo and S. Harper, “Coping tactics employed by visually disabled users on the web,” *International Journal of Human Computer Studies*, vol. 71, no. 11, pp. 1013–1025, 2013.

[15] J. P. Bigham and A. C. Cavender, “Evaluating existing audio CAPTCHAs and an interface optimized for non-visual use,” *Proceedings of the 27th International Conference on Human factors in Computing Systems*, p. 1829, 2009.

[16] M. E. Raven and A. Flanders, “Using contextual inquiry to learn about your audiences,” *ACM SIGDOC Asterisk Journal of Computer Documentation*, vol. 20, pp. 1–13, 1996.

[17] K. Holtzblatt and S. Jones, “Contextual inquiry: a participatory technique for system design,” *Participatory Design: Principles and Practice*, 1993.

[18] A. Strauss and J. M. Corbin, *Basics of qualitative research: Grounded theory procedures and techniques*. Sage Publications, Inc., 1990.

[19] Mayring, “Qualitative content analysis,” *Forum Qualitative Sozialforschung*, vol. 1, no. 2, p. 10, 2000.

[20] J. Annett, “Hierarchical task analysis,” *Handbook of Cognitive Task Design*, pp. 17–35, 2003.

[21] J. Bonneau, “The password thicket: Technical and market failures in human authentication on the web,” *Information Security*, vol. 8, no. 1, pp. 230–237, 2010.

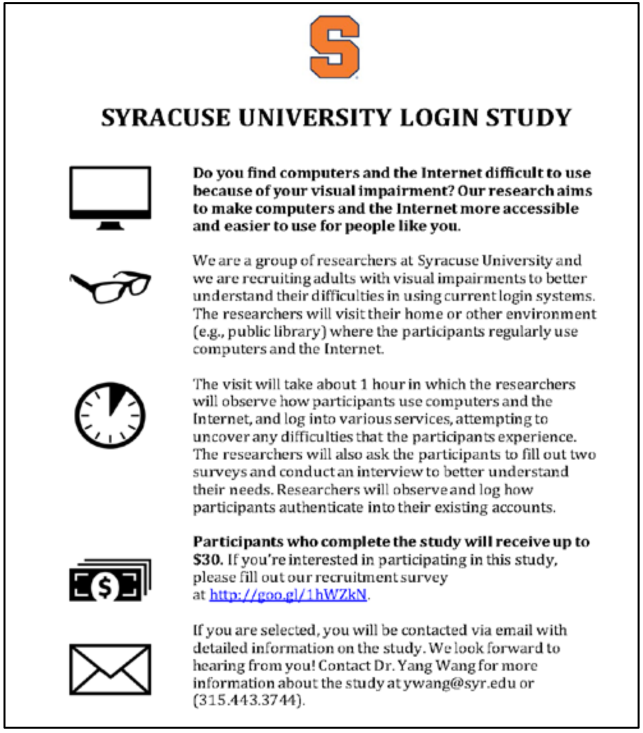
[22] J. Bonneau, C. Herley, P. C. Van Oorschot, and F. Stajano, “The quest to replace passwords: A framework for comparative evaluation of web authentication schemes,” *IEEE Symposium on Security and Privacy*, pp. 1–15.

[23] J. O. Wobbrock, S. K. Kane, K. Z. Gajos, S. Harada, and J. Froehlich, “Ability-Based Design,” *ACM Transactions on Accessible Computing*, vol. 3, no. 3, pp. 1–27, 2011.

[24] S. Schechter, a. J. B. Brush, and S. Egelman, “It’s no secret: Measuring the security and reliability of authentication via ‘secret’ questions,” *IEEE Symposium on Security and Privacy*, pp. 375–390, 2009.

9. APPENDIX

Figure 5: Recruitment Flyer



SYRACUSE UNIVERSITY LOGIN STUDY

Do you find computers and the Internet difficult to use because of your visual impairment? Our research aims to make computers and the Internet more accessible and easier to use for people like you.

We are a group of researchers at Syracuse University and we are recruiting adults with visual impairments to better understand their difficulties in using current login systems. The researchers will visit their home or other environment (e.g., public library) where the participants regularly use computers and the Internet.

The visit will take about 1 hour in which the researchers will observe how participants use computers and the Internet, and log into various services, attempting to uncover any difficulties that the participants experience. The researchers will also ask the participants to fill out two surveys and conduct an interview to better understand their needs. Researchers will observe and log how participants authenticate into their existing accounts.

Participants who complete the study will receive up to $30. If you're interested in participating in this study, please fill out our recruitment survey at http://goo.gl/1hWZkN.

If you are selected, you will be contacted via email with detailed information on the study. We look forward to hearing from you! Contact Dr. Yang Wang for more information about the study at ywang@syr.edu or (315.443.3744).

Figure 6: Contextual Inquiry Script

Introduction

Let me start by telling you a bit about this project and what we are trying to do. Our research team is trying to understand the challenges and difficulties that people with visual impairments face in using current authentication systems (e.g., logging into a computer or website). We want to understand your thought process so that we can develop technology that enhances current authentication systems.

We consider you the expert at so there are no wrong answers to any of our questions. While you answer questions or guide us through tasks, please focus on the details of how you actually log into your computer and online accounts. It may help to think about the last time you performed the task and explain it to us as if we are going to need to perform the task just as you did.

To backup my notes, I'd like to tape record our session. My research team will be the only users to listen to this. Are you okay with me recording the conversation? Thanks.

Please review and sign the consent form before we proceed.

Do you have any questions before we begin? Let's get started. (Make observations of the interviewee's work environment.)

Observation

I'll be observing you and when it won't disrupt your flow, I'll stop you when I see something interesting and ask questions. Or, I'll wait until there is a break or talk to you between tasks. I'll also share my observations so you can tell me if I really understand what you do.

So, let's start by getting a bit of an overview of what you do that involves authentication systems. Keep in mind that although I will be making observations about your log in activities, I will not be recording your passwords nor will I be watching what you type in password fields. Please think out loud and verbally guide me through your thought processes and actions.

- Can you please turn on/restart your machine and walk me through how you log into it?
- Is your computer system password protected? Why or why not?
- Do you face any challenges with logging into your machine? For example, do you frequently forget your username or password, enter one or both incorrectly and don't know what to do about it?
- How frequently do you change your password? When was the last time you changed it?
- Describe your thought processes as you change your usernames and/or passwords or create new ones. Do you have any particular strategies for creating your usernames and/or passwords? If so, please describe them.
- Who knows this password? Is it just you?

Let's go over any of the collaboration and coordination tasks you have to do. I'd like you to go over with me what you do on your computer on a daily basis. Let's first see how you check your email. Again, please think out loud and verbally guide me through your thought processes and actions.

- Do you use any desktop icons or browser bookmarks to access your email client?
- If so, do you find these shortcuts convenient for you? How so?
- Do you have your passwords automatically saved on your email client, or do you manually enter your password each time to check your email?
- Is the login text easy for you to read and understand?
- Are there times where you've typed in the incorrect password? Have you ever been given a warning for typing the incorrect password too many times?

I'm now curious to learn how you manage your personal finances online. All of these sites have strict authentication systems, and I want to understand how you navigate through their web interfaces.

- Have you signed up for any online banking systems to track your balance? If so, which ones? Let's log into the one you check most frequently.
- Do you use different passwords for various online services, or do you generally stick with one or two for signing into multiple sites? Why?
- Are you comfortable making online purchases? Or would you prefer to conduct transactions offline and in person?
- Do you ever run into problems with verifying your online identity?

- Do you feel like authentication systems for money are more or less strict than authentication systems for checking email?

Let's move onto how you communicate with family and friends through your computer. Just a reminder, please continue to think out loud and verbally guide me through your thought processes and actions.

- Are you connected in any social media networks (e.g. Facebook, Twitter, Tumblr)?
- Can you walk me through logging into Facebook?
- Why did you choose to save/not save your password as a cookie on your browser? Do you find this convenient?
- How does logging in through social media sites differ from checking your financial activity online?
- Do you share any of your passwords with family or friends?

(Skip this section if user does not own a smartphone, tablet or other portable device besides their personal computer that they normally use.) I think you've given me a good overview of the work that you do on your computer. What I'd like now is for you to start logging into other sites that you normally check on a routine basis through your smart phone. I'd like for you to walk me through this process as well by thinking out loud.

- Is your smart phone password-protected?
- Does the smaller screen pose any additional/new challenges for you?
- Are you familiar with two-factor authentication? Using two-factor authentication provides an additional layer of security by asking you to enter an additional piece of information, such as a verification code, to log in after entering your username and password. Have you ever used your smart phone as a token for verifying your identity?
- Do you use any of the disability/accessibility features on your smartphone (e.g. iPhone's "Assistive Touch")?
- Have you downloaded any applications on your smart phone that help accommodate for your disability? If so, can you walk me through opening these applications and enabling them?

Great. I just have several more questions to ask you before we wrap up here.

- Do you have any suggestions to improve these login experiences?
- Aside from passwords, what would be the ideal way to log into these services?

Wrap up

I really appreciate all the time you've given me. As we wrap up, let me summarize some of the key points I've learned about your role here.

- Create a large interpretation of your learning about the user's role. The wrap-up is an opportunity to summarize what you learned about the user's role and work. It is a way for you to check your high-level understanding with the user.
- Clear up any thought processes that need further clarification.
- Ask the user to reflect on his or her experience after completing the test. Clear up any thought processes that need further clarification.
- Ask if there is anything else regarding the usability of authentication systems the user would like to add and whether or not this test has changed his or her perspectives and/or attitudes towards current authentication practices.
- Can the user suggest another interested person of disability who would like to get involved with the study?
- Thank the user for his/her time and give the user a gift card. Exchange contact information so that the user/researcher can ask any follow up with any questions.

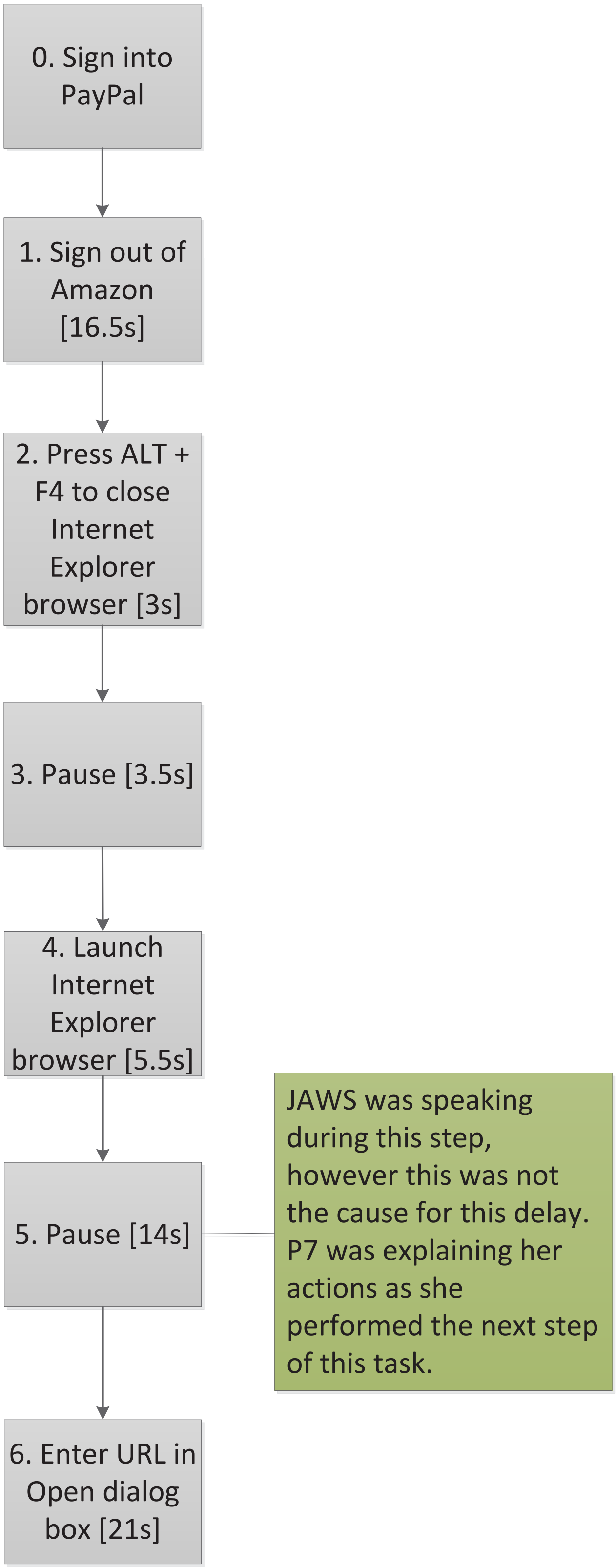
Figure 7: Task Flow Diagrams

We assigned a light-gray color to each step and sub-step on the diagram that the participant actually completed, while white-colored boxes indicate steps that participants did not take. In the process of doing this, we also included any quotes, comments and/or observations from the notes and interview transcripts that were of interest and relevant to the sub-steps involved in completing the task. We then placed each annotation next to the applicable step in the diagram and assigned them a different color in order to distinguish the annotations from the actual steps and sub-steps.

We calculated the total task completion time by adding all of the timings in each step in the high-level diagram and placed the final sum on the final step of this diagram.

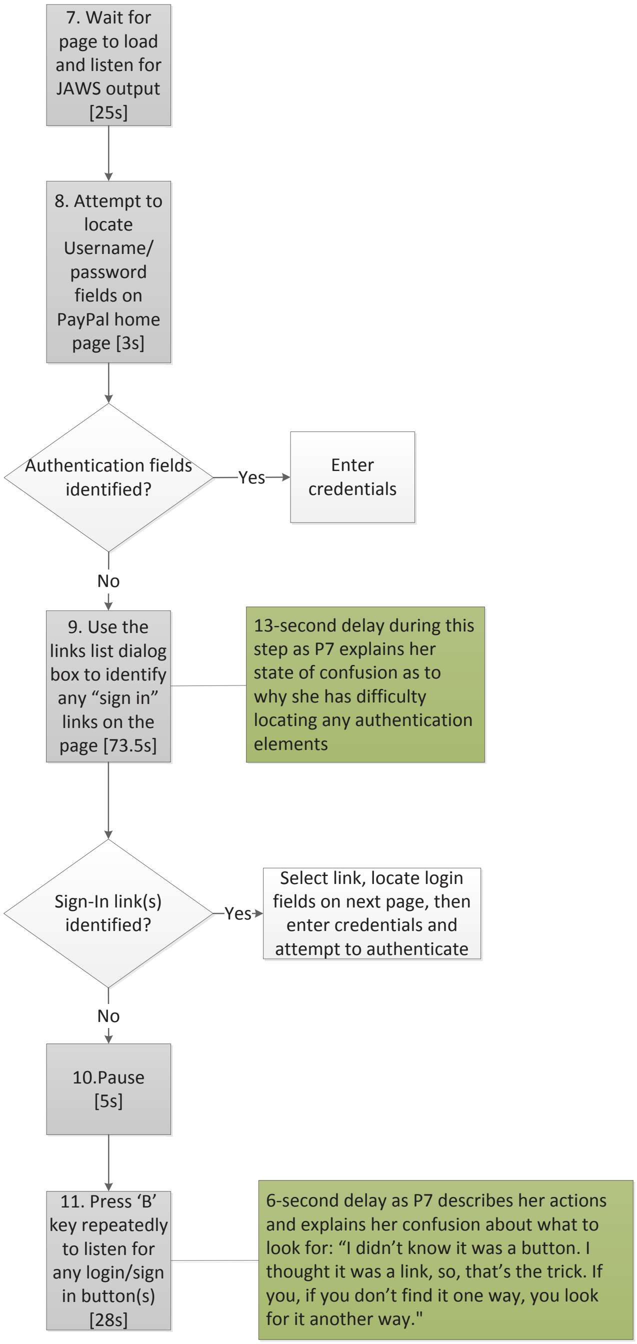
We include an example of P7's PayPal task in Figure 7, which took a total of 400 seconds to complete. The high-level diagram contained too many steps to legibly fit on one page; therefore we split this diagram into three separate parts. Furthermore, we then selected two particular steps we felt were the most complex and time-consuming for P7 to complete her entire task. Specifically, these complex steps are Step 7 and Step 9.

Figure 7A. PayPal (P7) HTA High-Level Diagram (part 1)


0. Sign into PayPal
1. Sign out of Amazon [16.5s]
2. Press ALT + F4 to close Internet Explorer browser [3s]
3. Pause [3.5s]
4. Launch Internet Explorer browser [5.5s]
5. Pause [14s]
6. Enter URL in Open dialog box [21s]

JAWS was speaking during this step, however this was not the cause for this delay. P7 was explaining her actions as she performed the next step of this task.

Figure 7B. PayPal (P7) HTA High-Level Diagram (part 2)


7. Wait for page to load and listen for JAWS output [25s]
8. Attempt to locate Username/ password fields on PayPal home page [3s]
Authentication fields identified?
Yes
Enter credentials
No
9. Use the links list dialog box to identify any “sign in” links on the page [73.5s]
Sign-In link(s) identified?
Yes
Select link, locate login fields on next page, then enter credentials and attempt to authenticate
No
10.Pause [5s]
11. Press ‘B’ key repeatedly to listen for any login/sign in button(s) [28s]

13-second delay during this step as P7 explains her state of confusion as to why she has difficulty locating any authentication elements

6-second delay as P7 describes her actions and explains her confusion about what to look for: “I didn’t know it was a button. I thought it was a link, so, that’s the trick. If you, if you don’t find it one way, you look for it another way."

Figure 7C. PayPal (P7) HTA High-Level Diagram (part 3)

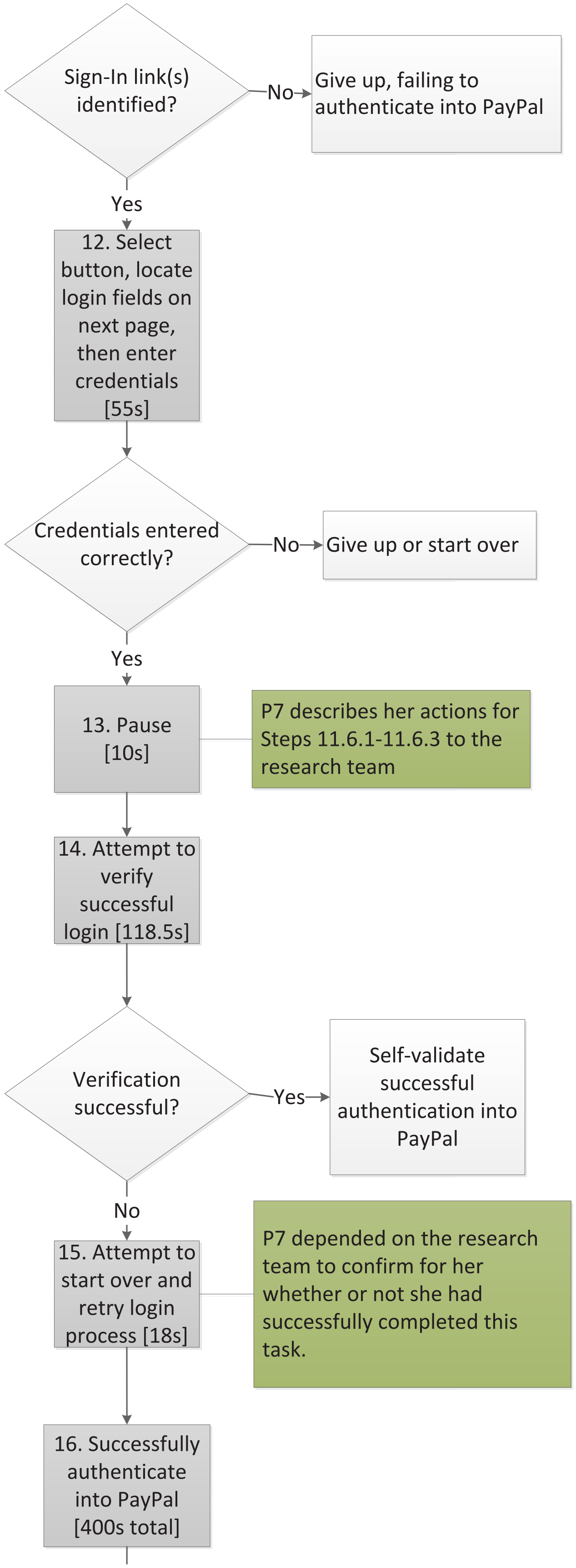

Sign-In link(s) identified?
No
Give up, failing to authenticate into PayPal
Yes
12. Select button, locate login fields on next page, then enter credentials [55s]
Credentials entered correctly?
No
Give up or start over
Yes
13. Pause [10s]
P7 describes her actions for Steps 11.6.1-11.6.3 to the research team
14. Attempt to verify successful login [118.5s]
Verification successful?
Yes
Self-validate successful authentication into PayPal
No
15. Attempt to start over and retry login process [18s]
P7 depended on the research team to confirm for her whether or not she had successfully completed this task.
16. Successfully authenticate into PayPal [400s total]

Figure 7D. PayPal (P7) HTA Step 7 Diagram

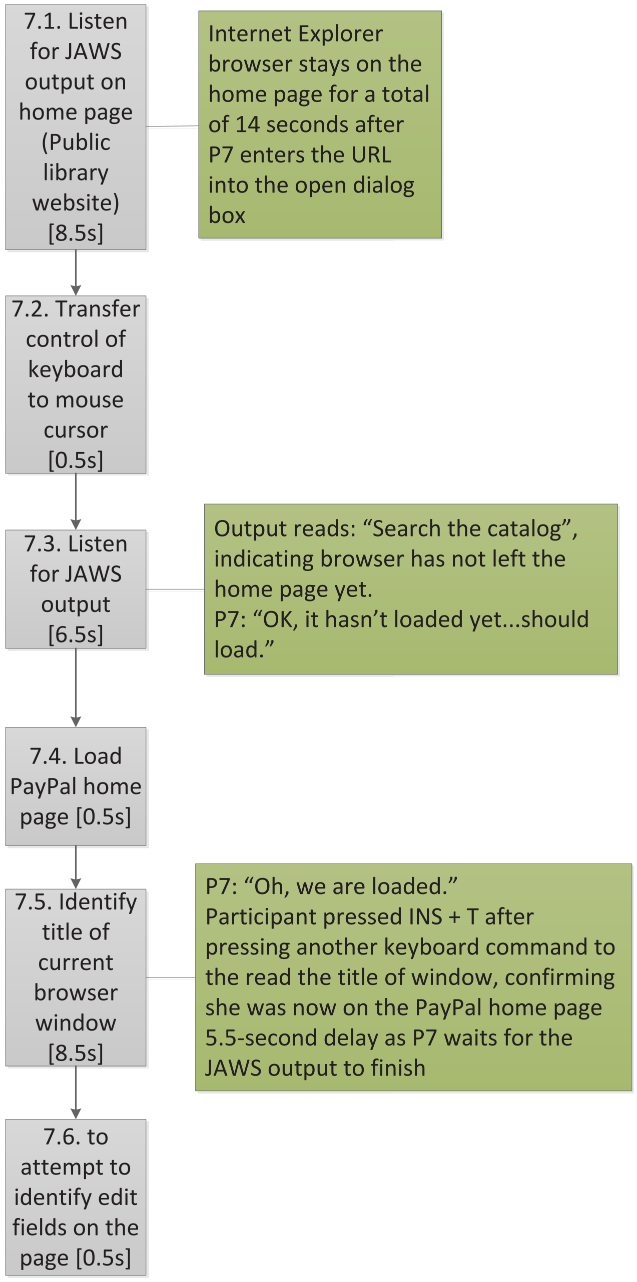
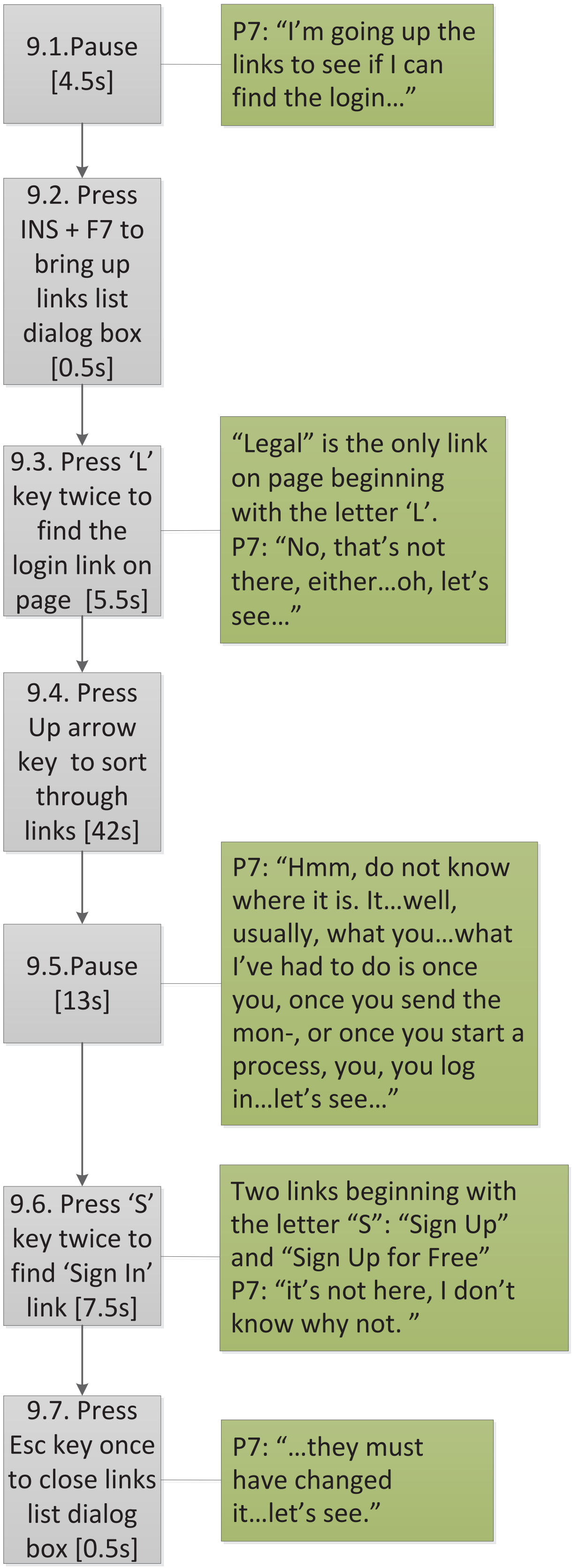

7.1. Listen for JAWS output on home page (Public library website) [8.5s]
Internet Explorer browser stays on the home page for a total of 14 seconds after P7 enters the URL into the open dialog box
7.2. Transfer control of keyboard to mouse cursor [0.5s]
7.3. Listen for JAWS output [6.5s]
Output reads: "Search the catalog", indicating browser has not left the home page yet.
P7: "OK, it hasn't loaded yet...should load."
7.4. Load PayPal home page [0.5s]
7.5. Identify title of current browser window [8.5s]
P7: "Oh, we are loaded."
Participant pressed INS + T after pressing another keyboard command to the read the title of window, confirming she was now on the PayPal home page
5.5-second delay as P7 waits for the JAWS output to finish
7.6. to attempt to identify edit fields on the page [0.5s]

Figure 7E. PayPal (P7) HTA Step 9 Diagram



9.1.Pause [4.5s]

P7: "I'm going up the links to see if I can find the login…"

9.2. Press INS + F7 to bring up links list dialog box [0.5s]

9.3. Press 'L' key twice to find the login link on page [5.5s]

"Legal" is the only link on page beginning with the letter 'L'. P7: "No, that's not there, either…oh, let's see…"

9.4. Press Up arrow key to sort through links [42s]

9.5.Pause [13s]

P7: "Hmm, do not know where it is. It…well, usually, what you…what I've had to do is once you, once you send the mon-, or once you start a process, you, you log in…let's see…"

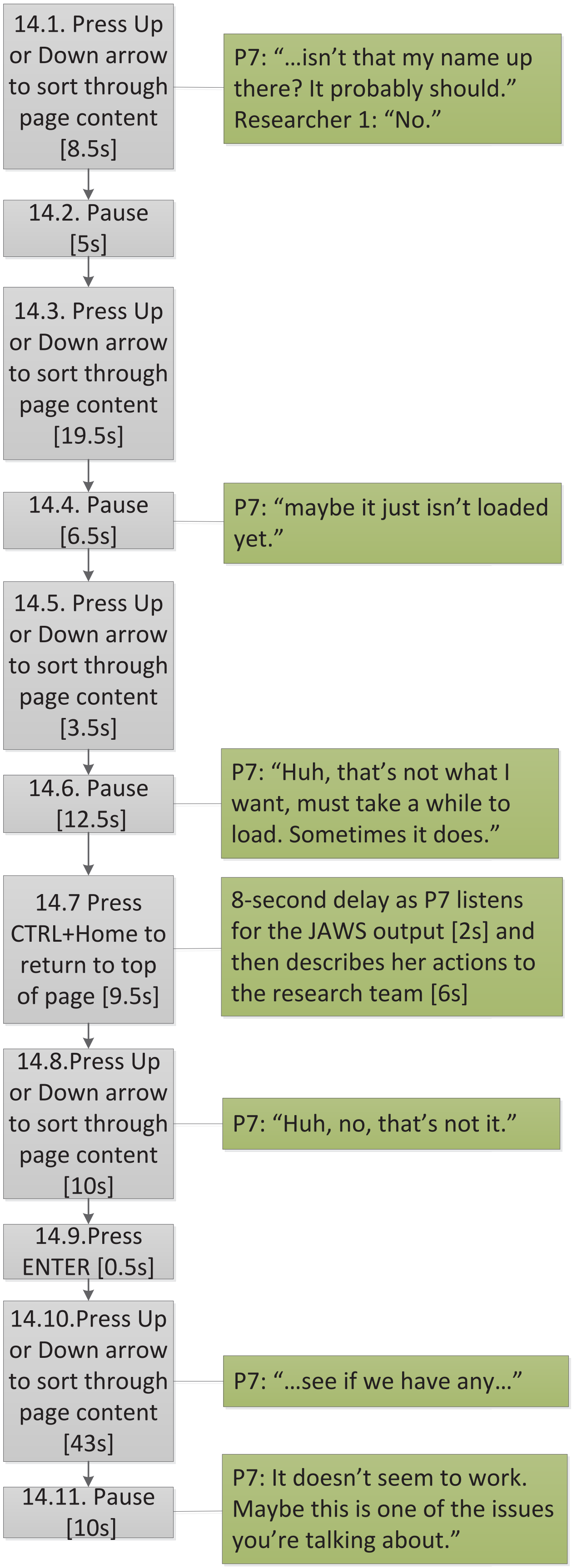
9.6. Press 'S' key twice to find 'Sign In' link [7.5s]

Two links beginning with the letter "S": "Sign Up" and "Sign Up for Free" P7: "it's not here, I don't know why not. "

9.7. Press Esc key once to close links list dialog box [0.5s]

P7: "…they must have changed it…let's see."

Figure 7F. PayPal (P7) HTA Step 14 Diagram



14.1. Press Up or Down arrow to sort through page content [8.5s]

P7: "…isn't that my name up there? It probably should." Researcher 1: "No."

14.2. Pause [5s]

14.3. Press Up or Down arrow to sort through page content [19.5s]

14.4. Pause [6.5s]

P7: "maybe it just isn't loaded yet."

14.5. Press Up or Down arrow to sort through page content [3.5s]

14.6. Pause [12.5s]

P7: "Huh, that's not what I want, must take a while to load. Sometimes it does."

14.7 Press CTRL+Home to return to top of page [9.5s]

8-second delay as P7 listens for the JAWS output [2s] and then describes her actions to the research team [6s]

14.8.Press Up or Down arrow to sort through page content [10s]

P7: "Huh, no, that's not it."

14.9.Press ENTER [0.5s]

14.10.Press Up or Down arrow to sort through page content [43s]

P7: "…see if we have any…"

14.11. Pause [10s]

P7: It doesn't seem to work. Maybe this is one of the issues you're talking about."

Where Have You Been? Using Location-Based Security Questions for Fallback Authentication

Alina Hang
Media Informatics Group
University of Munich (LMU)
Germany
alina.hang@ifi.lmu.de

Alexander De Luca
Media Informatics Group
University of Munich (LMU)
DFKI GmbH
Germany
alexander.de.luca@ifi.lmu.de

Matthew Smith
Usable Security & Privacy Lab
University of Bonn
Germany
smith@cs.uni-bonn.de

Michael Richter
Media Informatics Group
University of Munich (LMU)
Germany
michael.richter@campus.lmu.de

Heinrich Hussmann
Media Informatics Group
University of Munich (LMU)
Germany
hussmann@ifi.lmu.de

ABSTRACT

In this paper, we propose and evaluate the combination of location-based authentication with security questions as a more usable and secure fallback authentication scheme. A four weeks user study with and additional evaluation after six months was conducted to test the feasibility of the concept in the context of long-term fallback authentication. The results show that most users are able to recall the locations to their security questions within a distance of 30 meters, while potential adversaries are bad in guessing the answers even after performing Internet research. After four weeks, our approach yields an accuracy of 95% and reaches, after six months, a value of 92%. In both cases, none of the adversaries were able to attack users successfully.

1. INTRODUCTION

Passwords still have a prevalent role in today's world, where they are mostly used in combination with usernames to protect the users' accounts and data. However, the number of these accounts is steadily increasing, confronting users with the challenge to define distinct and secure passwords [1]. When users forget passwords, fallback authentication schemes are required to enable users to regain access to their account and data. While authentication schemes such as passwords have received a lot of attention in the usable security and privacy community, fallback authentication schemes have not seen the same amount of attention.

Most common approaches for fallback authentication rely on email-based password resets or security questions (e.g. [18]). In general, email-based password resets work well, but are not appropriate in all circumstances (i.e. when users forget the password to their email account). Therefore, security questions are often used as an alternative. They take advantage of personal information, assuming that such information are easily remembered by users and at the same time hard to guess by others. However, previous research has shown that the use of security questions comes with a variety of shortcomings with respect to usability and security (e.g. [10]).

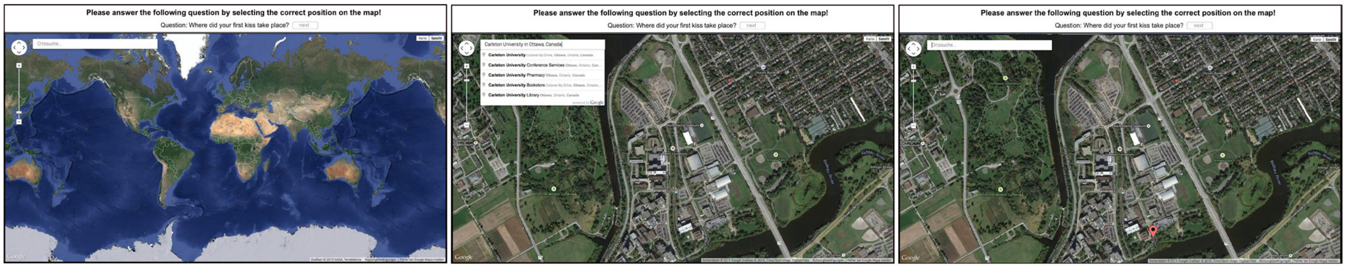
To overcome these shortcomings, we propose location-based security questions as an alternative design. Our questions are similar to traditional security questions in the sense that they are based on personal information, but different as they focus on questions about episodic memories with a spatio-temporal context [17] (e.g. *"Where did your first kiss take place?"*) and thus, also differ in the way the answers are provided. Instead of entering them as text, which often comes with issues like repeatability [8], users submit their answers by selecting a location on a map.

Our hypothesis is that location-based questions are easier to recall as they are remembered more vivid than personal facts [17]. Furthermore, using maps for answer input can serve as helpful memory hooks for users to recall their answers to questions (e.g. street crossings, buildings, etc.). In order to test the usability and security of the proposed approach, we conducted a user study over a period of four weeks and evaluated three types of location-based questions: predefined, guided and open questions. All questions were tested with different types of adversaries: close adversaries (i.e. persons that know the user well) and strangers. We also performed an additional evaluation after six months to test the memorability of the presented approach.

The results of our study show that it is hard for persons close to the user as well as strangers to guess or even research the answer to a question. Social networks and search engines do not provide sufficient hints and even if they do, it is difficult to be close enough to the actual location (i.e. to be within a distance of 30 meters). In turn, users are very good in answering their questions. The accuracy values (95% after four weeks; and 91% after six months) of location-based questions are promising, but leave room for improvements with respect to the usability of the approach.

The main contributions of this paper are twofold: (1) we

Symposium on Usable Privacy and Security (SOUPS) 2015, July 22–24, 2015, Ottawa, Canada.



Please answer the following question by selecting the correct position on the map!

Question: Where did your first kiss take place?

Please answer the following question by selecting the correct position on the map!

Question: Where did your first kiss take place?

Carleton University in Ottawa, Canada

Carleton University Colonel By Drive, Ottawa, Ontario, Canada

Carleton University Conference Services Ottawa, Ontario, Can...

Carleton University Pharmacy Ottawa, Ontario, Canada

Carleton University Bookstore Colonel By Drive, Ottawa, Ontario, ...

Carleton University Library Ottawa, Ontario, Canada

Please answer the following question by selecting the correct position on the map!

Question: Where did your first kiss take place?

Figure 1: Screenshot of the prototype during authentication for the exemplary question *"Where did your first kiss take place?"*. The authentication always starts with the world map (left). Users then can zoom in on a map section by using the mouse, map controls or the provided text field (center), but the answer must be provided by selecting it on the map using the mouse (right). Translated from German.

present a detailed usability (i.e. memorability) and security analysis of location-based security questions over a period of six months to simulate fallback scenarios and (2) we provide reasons why location-based questions work better than traditional security questions and discuss the potentials and risks of different suggestions for further optimization.

2. RELATED WORK

Fallback authentication usually consists of two phases. In the first phase, the enrollment, users have to provide various information, such as email-addresses, phone numbers or answers to security questions. This information is needed for the second phase when password reset/retrieval is required (e.g. forgotten passwords). The time that elapses between those two phases can be very long.

A commonly used method for fallback authentication is the email-based password reset. In case of password loss, a new password or a reset link is sent to the user's email address. According to Simson Garfinkel [3], this approach works well, but comes with certain shortcomings. For example, it makes the email account a single point of failure. Furthermore, the email address that the user has provided during enrollment might be out of date and thus, not accessible anymore.

Some service providers offer users the possibility to associate their mobile phone number with their accounts. Upon request for password reset, a one-time password is sent to this number, with which the users can temporarily log into their account to define a new password. However, mobile phone numbers are sensitive information that not everyone wants to share with every service provider [5].

Another popular approach is the use of security questions. Such questions and their answers can either be fixed, controlled or open [7]. While fixed questions (i.e. predefined questions) leave little room for the user to make changes, users often lack creativity to handle open questions and thus, come up with questions that are similar to fixed ones. Controlled questions are a combination of both. For example, users can define hints for a question that will be shown when they have to answer the respective question. However, these hints will also be visible for potential adversaries.

Most service providers rely on fixed questions about personal facts (e.g. *"What is your mother's maiden name?"*). In the past, such questions were assumed to be easy to remember by the user and hard to answer by others. The reality is that questions that are easy to remember are often easy to guess, while questions that are hard to guess are also hard to remember [10]. Thus, security questions come with numerous insufficiencies in terms of usability and security. Inapplicability, memorability and ambiguity are one of the key issues worth mentioning with respect to usability [10]. In terms of security, many predefined security questions are researchable (e.g. [4]), can easily be answered by close persons like family and friends (e.g. [6]) or can even be guessed by choosing the most popular answers [12].

In order to overcome these insufficiencies, various alternative solutions have been proposed. For example, Schechter et al. [13] propose a system called social authentication. In case of password loss, users have to contact two or three contacts to retrieve tokens that are part of the authentication process. However, their studies also showed that after a certain time, users could not recall the names of the social contacts they had provided during enrollment.

Since memorability becomes an issue when the time between enrollment and fallback authentication increases, Babic et al. [2] propose a dynamic approach that uses security questions based on recent browser activities. Using implicit data seems promising, but may evoke privacy issues, as users have no power over which information is used.

In summary, it can be said that the design of security questions is a challenging task that in particular tackles issues like memorability and security. Most research so far has focused on the design on question level, neglecting the way the answer is provided.

3. CONCEPT

We suggest an alternative concept to traditional security questions to address their well-known shortcomings (e.g. memorability or repeatability [7, 10]). Our concept focuses on episodic memories with a spatio-temporal context [17] to generate location-based security questions. Psychological research has shown that these kinds of memories are easier to remember than, for example, personal facts due to their more vivid recall [17].

Although traditional security questions also include questions about locations, our concept is different as the answers are not provided as text, but instead, are entered by selecting a location on a map. The way of entering a location into the system is inspired by GeoPass [15]. However, our approach is not an extension of this existing approach, where an arbitrary location is used as a primary password, but instead, we present a novel alternative that combines security

Predefined Questions			
Whereto was your first travel by plane?	(5)	Where have you been camping for the first time?	(1)
Whereto was your longest travel so far?	(5)	Where was your first car accident?	(1)
Where is your favorite beach?	(3)	Where did you park for your driving test?	(1)
Where did your best friend from elementary school live?	(2)	Where did you injure yourself badly for the first time (e.g. broken leg)	(1)
Where was your first time at the sea?	(2)	Where did your best kindergarten friend live?	(0)
Where did you meet your best friend?	(2)	Where did you spend your first vacation?	(0)
Where did your first kiss take place??	(2)	Whereto did you drive in your first driving lesson?	(0)
Where have you been in a dangerous situation?	(2)	Where was your first party?	(0)
Where does a distant relative of yours live?	(1)	Where was your first breakup?	(0)
Whereto did you travel for your first school trip?	(1)	Where was your most embarrassing moment?	(0)
Where was your first job interview?	(1)	Where was your saddest moment?	(0)

Table 1: Overview of the 22 fixed questions used in the study. The values in brackets depict the number of times a question has been selected during the study. Translated from German.

Guided Questions	
Please define a location-based question that refers to a travel destination/vacation destination.	(7)
Please define a location-based question that refers to a personally experienced sport event.	(5)
Please define a location-based question that refers to an event in your childhood.	(4)
Please define a location-based question that refers to an event during your time at university/apprenticeship.	(4)
Please define a location-based question that refers to one of your party experiences.	(3)
Please define a location-based question that refers to something that you did for the first time.	(3)
Please define a location-based question that refers to an event that during your time in school.	(2)
Please define a location-based question that involves another person.	(1)
Please define a location-based question that refers to one of your favorite places.	(1)
Please define a location-based question that refers to an experience that had a strong impact on your life.	(0)

Table 2: Overview of the 10 guidelines for the guided questions used in the study. The values in brackets depict the number of times a category has been selected during the study. Translated from German.

questions with map-based input. This is an important difference, since we argue that map-based input is not a good option to replace passwords (e.g. due to long authentication times), but it is a good option to replace text-based answers.

The context in which location-based questions are supposed to be used (i.e. fallback authentication) represents another difference and thus, imposes harder requirements on the design and evaluation of location-based questions:

Fallback authentication happens less frequently than primary authentication (about once a month or less [14]) so that users should be able to recall the needed information even after longer periods of time. Therefore, it seems advisable to favor cued-based recall over free recall as previous research has shown the superiority of the former [16]. In this concept, we use questions as cues to trigger episodic memories that are associated with a particular location.

Furthermore, in order to authenticate, users have to answer a sequence of location-based questions on a map (instead of remembering an arbitrary location). This is required to reach a certain level of security.

To find the best trade-off between usability and security, we evaluate the concept in four sessions to simulate fallback authentication. We test the memorability of the concept shortly after enrollment as well as one week, three weeks and six months after the last authentication attempt. We evaluate the security of the approach and test it with different types of human adversaries. We further analyze the number of questions that users should answer in order to reach a certain level of security and discuss the implications when users exhaust the number of authentication attempts.

4. THREAT MODEL

We consider three different types of threats to evaluate the presented concept: a) threats by close adversaries, b) threats by close adversaries that use the Internet for researching the answers and, c) threats by strangers that also use the Internet for research to perform educated guesses.

Close adversaries (e.g. partners) have the advantage to know the user well and thus, do not have to rely on plain luck to guess the correct answers. The threat can be increased, when they use additional tools like social networks or search engines for research. This kind of threat can be considered as one of the worst case scenarios for location-based security questions.Threats by close adversaries were shown to be very likely and thus, interesting to consider [9].

The chances for a stranger to guess the correct answer (without any assisting tools for research) is $(\frac{1}{x})^n$, with x being the number of all possible locations on a world map and n depicting the number of questions asked. The answer space is narrowed down when more targeted attacks are considered (e.g. by limiting the answers to the country where the victim lives). Therefore, we also test the performance of adversaries that do not have any prior knowledge about the user (i.e. strangers), but use the Internet to take advantage of information on social networks, telephone directories or results from search engines to make educated guesses.

In the scope of location-based questions, brute force attacks have to be mentioned, where more sophisticated adversaries have the skills to use automated processes to attack the questions by successively guessing one location after another. In order to undermine these attacks, our concept limits the number of attempts per question to three, which is a common threshold used for fallback authentication systems. A more detailed analysis of the number of attempts will be provided in the result section, while the implications of such a limit will be addressed in the discussion.

5. SECURITY QUESTIONS DESIGN

For the first design of the security questions, we performed a focus group with five participants (all male). They were recruited over bulletin boards, mailing lists and personal communication. Participants were aged between 18-26 years (average: 22 years) and were all students with a background in natural sciences (i.e. computer science, physics and medical engineering).

The participants were invited to our lab and were given a short introduction to fallback authentication and security questions. This was followed by a brief explanation of our concept. We asked participants to discuss advantages and disadvantages of the concept and encouraged them to discuss ideas for location-based security questions.

During the discussion, participants identified promising categories, including *childhood memories* (as these memories lie far in the past so that only few people know about it), *travel / vacation* (as these kinds of questions have a large answer space) and *first time memories* (as they are memorable). Participants also mentioned questions about *big events* (like concerts) or *third parties* (e.g childhood friends).

Since the identified categories are highly individual (not everyone has made similar experiences in the past), participants raised concerns about the applicability of predefined questions. Open questions were also considered as difficult, since users might define questions that are too easy to guess. Therefore, participants suggested to use something in-between those two extremes: guided questions which provide users with a basis to work on, but allow them to personalize the questions (e.g. *define a location-based question that refers to an event in your childhood).* The concerns comply with the problems discussed in [7].

In our study, we used all three question types (predefined, guided and open) and compared them to each other. For each type (except for open questions), we used the insights from the focus group to design the location-based questions. Altogether, we ended up with 22 predefined questions (see table 1) and 10 guided questions (see table 2).

6. PROTOTYPE

The study application used the Google Maps API (in combination with HTML5 and JavaScript) to obtain location-based information and logged all relevant user interactions (e.g *timestamps*, *selected/defined questions*, *latitude*, *longitude*, etc.). It consisted of three main modes: enrollment, authentication and attack.

6.1 Enrollment

In the enrollment phase, users selected their questions and provided the corresponding answers. The way of enrollment varied for the different question types. For predefined questions users had to select three questions from a list of 22 questions. For guided questions users had to select three out of 10 guidelines from a list. In addition to this, three text fields were provided that allowed users to define a question based on each selected guideline. For open questions users were given three text fields and a brief instruction to define three location-based questions.

Once the questions had been selected/defined, they were consecutively shown to the users. Users were asked to provide the answers to the given questions by selecting a location on the map. Since it may be difficult for some users to find the right region on the world map, they had the possibility to enter an address into a given text field to zoom in on the corresponding map section. However, no position marker was set to ensure that users make their own selection by clicking on the map (see figure 1). This was done to make the selection more individual and thus, more difficult to be guessed. Users were allowed to reposition their marker. The answer was submitted by pressing the *save*-button.

6.2 Authentication

In authentication mode, users were presented with the questions they had selected/defined during enrollment. In order to authenticate, users had to provide the answers by selecting the locations on the map. Again, users had the possibility to enter the location into a text field to zoom in on a particular part of the world map, but a position marker had to be set by clicking on a location on the map. Users had three attempts to submit the correct answer. An answer was considered as correct, when the distance between the selected location and the location provided during enrollment was smaller than 30 meters. This threshold was shown to be useful by Thorpe et al. [15].

6.3 Attack

The only difference to the authentication mode was that the answers were provided by potential adversaries (close ones and strangers) instead of the legit user.

6.4 Map and Zoom Level

For each question, the map was initialized at zoom level 2 and was centered at the position 0.0 / 0.0 (latitude / longitude). Participants always saw the whole world map as a starting point. This was done to avoid influencing users during answer selection and helped to prevent hinting possible location areas for answers to potential adversaries.

In order to submit a location as an answer, users were required to obtain a zoom level that was higher than 16. This value was shown to be useful by Thorpe et al. [15]. In case the zoom level was too small, a pop-up notification informed users to zoom in.

7. USER STUDY

The user study consisted of a short-term evaluation of four weeks (with three sessions) and a long-term evaluation after six months.

7.1 Study Design

For the study design, we used a between-groups design with the independent variable *question type* (three levels: predefined, guided and open). A between-groups design was necessary to prevent biasing users during enrollment (e.g. preventing users to define similar questions to the ones that they encounter for predefined questions).

The prerequisite to participate in the study was to come in pairs and to have a close relationship with each other. We gave participants examples of close relationships during recruitment (e.g. partners, best friends). For each pair, the participants took over different roles. One acted as legit user, while the other acted as close adversary who tried to attack the questions. In the remainder of this paper, we will refer to legit users shortly as *users* and to participants who attacked the questions as *close adversaries*. It was also possible for participants to take part in both roles, meaning that they acted as users as well as each other's close adversary.

As incentives, participants received gift vouchers of 20€ for users or 5€ for close adversaries. In case they acted in both roles, they received 25€ . No incentive was provided when not all required sessions of the short-term evaluation were completed. Participants received additional 5€ gift vouchers when they took part in the long-term evaluation.

7.2 Study Procedure

The study was divided into three sessions and a long-term evaluation. For all sessions, participants were invited to our lab. While users had to attend all sessions for memorability testing, close adversaries only had to come for the first session. The long-term evaluation was conducted online.

7.2.1 First Session

The first session started with a brief introduction to fallback authentication (and security questions), the proposed concept and the study procedure. Then, users were assigned to one of the three groups (predefined, guided or open).

Close adversaries were asked to leave the room and wait, while users did the enrollment for their assigned question type (i.e. selecting/defining the corresponding questions and providing the corresponding answers on a map). Users were asked to select/define and answer three location-based questions. Once the enrollment was completed, we gave users a distraction video (duration about six minutes) after which a short-term memorability test was performed. Users were given three attempts to answer the questions they had just selected/defined. Users were informed whether a question was answered correctly/incorrectly.

For the attack, we asked users to leave the room and invited the close adversaries back in. Adversaries also had three attempts to guess the answers to the selected/defined security questions. For all questions that close adversaries did not answer correctly after three attempts, we gave them a second chance for attack, but this time, they were allowed to use the Internet for research.

In case close adversaries also wanted to participate as users, we paid particular attention that both users completed the enrollment and short-term memorability test first (one after another) before performing the actual attack to avoid influencing users during enrollment. Furthermore, both users were assigned to different groups.

At the end of the study, participants were asked to fill out a questionnaire to collect demographic information, qualitative ratings and also to ask participants to state the closeness of their relationship on a 5-point Likert scale to check their level of agreement.

On a separate form, we asked users to fill out the following information: first name, last name, date of birth and place of birth. This kind of information can usually be spied on (e.g. personal ID) or retrieved from public records. We used this information for educated guessing attacks in which adversaries, that did not know the users, tried to research the answers. We informed users about the purpose of collecting this information and told them that providing the information was optional. However, none of them refused. Furthermore, we paid particular attention that our research complies with the federal (privacy) laws in our country.

7.2.2 Second Session and Third Session

One week after the first meeting, we invited users back to perform another memorability test. Again, users had to answer their three questions from the first session within three attempts. Another memorability test was conducted in a third session that took place three weeks after the second one (i.e. four weeks after the first meeting). Users had to complete the same tasks as for the second session.

7.2.3 Long-Term Evaluation

Six months after the third session, we invited user to a long-term evaluation to simulate a realistic fallback scenario in which a long time between enrollment and required fallback authentication had passed. The procedure was similar to the second and third session, but was done online over Skype to spare users from long travel times to our lab (and thus, to encourage them to participate).

7.2.4 Educated Guessing Attacks

Two persons that were strangers to the user were asked to research the answers to the security questions. We provided them with the users' personal information (i.e. first name, last name, date of birth and place of birth) and a list of the security questions grouped by user. The adversaries had two weeks time to use this information for Internet research. For each question, they had to submit three possible locations and state briefly why they had selected a certain answer.

7.3 Participants

Thirty-two participants (15 female) took part in the user study. Twenty-eight of them acted as both, user and close adversary. Two participants acted as user only and another two acted as close adversary only. Participants were aged between 17-55 years (average: 26 years).

Four of them were high school students, 21 of them were students with different backgrounds (e.g. computer science, business or medicine), 5 were employed (e.g. administration or finance).

The relationships between users and close adversaries were manifold. The majority of pairs were good friends, best friends or partners/spouses. In four cases, the stated relationships did not match. For example, while one person described the other person as good/best friend, the assumed good/best friend stated to be only acquainted/good friends. However, the lines between good friend and good acquaintances or best friends and good friends are hard to draw. Altogether, there was a good agreement among pairs about their relationship. We also asked participants to rate how well they knew each other on a Likert scale ranging from not at all (1) to very well (5). Almost all pairs stated to know each other very well (8 pairs) or at least well (4 pairs). One pair stated to know each other a little. For three pairs, the ratings did not match. Two pairs had a mismatch between very well and well, while one pair had a mismatch between well and a little.

Two additional participants (one female) were recruited to perform educated guessing attacks. They were not related to the users or close adversaries from the user study and thus, strangers to them. They were 29/33 years old. Both of them

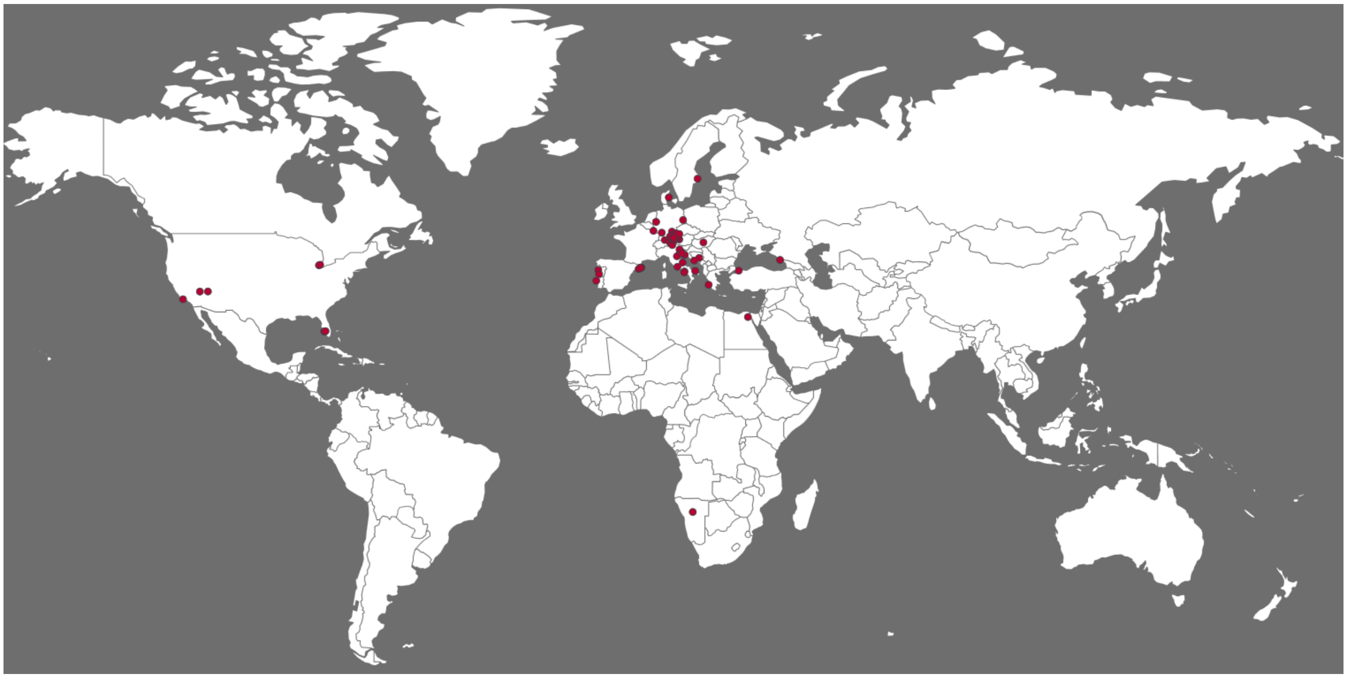


Figure 2: Overview of the distribution of answers provided by users during enrollment on the world map.

were security experts with experience in attacking security systems. They did not receive any incentives, but due to their background, they had a high intrinsic motivation to break the system.

With respect to the long-term evaluation, 24 out of 30 users took part in the experiment. Six users did not reply to our invitation to participate again.

8. RESULTS: SESSIONS 1,2 AND 3

Altogether, users defined or picked 90 location-based security questions (30 predefined questions, 30 guided questions and 30 open questions).

8.1 Question Types

8.1.1 Predefined Questions

Most predefined questions were about travel (e.g. first flight, or longest travel). This is followed by questions that involve a third person (e.g. best friend or first kiss). Table 1 gives an overview of how often each question was selected.

8.1.2 Guided Questions

With respect to guided questions, most users picked out guidelines for questions about travel, followed by questions about sport activities. Table 2 gives an overview of how often a guideline was picked. In terms of content, most sport activities related to first athletic experiences or special achievements (e.g.*"Where did I receive my first sports award?"*, *"Where did I run my first marathon?"*). Examples of the questions about travel are *"Where was your graduation trip?"* or *"Where did I spend my most beautiful summer when I was a child?"*. There was one question that referred to the future (i.e. *"A place where I want to be at least once."*), while all other questions were about the past.

8.1.3 Open Questions

The open questions that our users defined often involved a third person or animal (e.g. *"Where was my tomcat born?"*). They also included special events (e.g. *"Where did I celebrate the victory against Argentina in 2010?"*), first times (e.g. *"Where did your first kiss take place?"*, travel (e.g. *"In which country did I get homesick?"*, education (e.g. *"Where was my final exam?"*) or preferences (e.g. *"Where can I eat my favorite food?"*).

8.2 Amount of Information in a Question

The amount of information that one needs to know to answer a question varied from question to question. For example, the question *"Where is the center of the route to my best childhood friend?"* assumes the knowledge of five pieces of information: Who is the childhood friend? Where does the childhood friend live? Where does the user live? Which route did the user take to his friend (there are probably multiple routes possible)? Where is the center of this route?

All questions require at least the knowledge of one piece of information (i.e. the location of the question). This was the case for 77% of the open and guided questions. Ten questions (17%) required two pieces of information (e.g. the involvement of a third person). Two questions (3%) required three pieces of information, while the remaining two questions (3%) required four or more pieces of information.

8.3 Number of Correct Answers

Users submitted their answers in three sessions (and during the long-term evaluation, but the corresponding results will be reported in another section). Adversaries (close ones and strangers) submitted their answers only in the first session as memorability testing was not relevant for them. Figure 2 gives an overview of the locations that users selected during enrollment. Interestingly, most answers were clus-

	S1	S2	S3
3 Correct Answers	21	20	19
2 Correct Answers	8	9	8
1 Correct Answers	1	1	3
0 Correct Answers	0	0	0
Total	30	30	30

	Questions 1			**Questions 2**			**Questions 3**		
	S1	**S2**	**S3**	**S1**	**S2**	**S3**	**S1**	**S2**	**S3**
3 Attempts	1	0	1	1	0	0	0	0	1
2 Attempts	0	3	1	1	3	1	1	2	1
1 Attempt	26	25	23	25	24	24	25	25	21
Fail	3	2	5	3	3	5	4	3	7
Total	30	30	30	30	30	30	30	30	30

Table 3: The table overviews for all three sessions (S1, S2 and S3) the number of users that had three, two or one of the three questions correct (left). It also shows for each question and session the number of users that needed one, two or three attempts as well as the number of users that failed for the corresponding question (right).

tered in a geographical area. Table 3 and 4 give an overview of the number of correct answers as well as the number of needed attempts for each question. Note that an answer was considered as correct, when its distance to the actual solution did not exceed a threshold of 30 meters.

8.3.1 Users

In the first session, users answered 80 questions (89%) correctly and failed for 10 questions (11%). Most users provided their answers within one attempt and the majority of users had all of their three questions correct. There were users who failed for some questions (one user failed in two questions; eight users failed in one question), but none of them failed completely.

In most cases, the reason for providing the incorrect answer was precision, meaning that users were close to the correct location, but failed to be within the required threshold (i.e. 30 meters). Only in two cases, the distances to the original locations were over 1000 meters. Those users stated to have forgotten the answer or to have used a location that they did not associate strong memories with.

After one week users recalled 79 of the 90 questions (88%) and failed only for 11 questions (12%). Again, most users needed only one attempt to provide the correct answers. Most of them were able to answer all of their three questions correctly. There is no user who failed in all three questions, but nine users had one incorrect answer and another user had two incorrect answers. Incorrect answers were mostly made by users (nine out of ten) who already failed to provide the correct answers to the same questions in the first session.

Four weeks after the first session, users were able to answer altogether 76 of the 90 questions (84%) correctly. Fourteen questions (16%) were answered incorrectly. The majority of users needed only one attempt to answer a question. No user failed in all three questions and the majority had all questions correct. Again, most users who had difficulties to provide the correct answers in the first and second session, could not provide the correct answers for the same questions in the third session. However, two users who submitted incorrect answers in the first and second session, managed to answer correctly in the third session.

8.3.2 Close Adversaries

Close adversaries answered 6 of the 90 questions (7%) correctly, meaning that they failed to provide the correct answers at most times (93%). No close adversary had more than one correct answer within the set of three questions. The number of attempts needed differed from adversary to adversary. One needed three attempts, one needed two attempts and four needed one attempt.

		Question Type			
	S	**P**	**G**	**O**	**Total**
User	**1**	26	27	27	80
	2	26	26	27	79
	3	27	22	27	76
Close Adversary	**1**	1	3	2	6
Close Adversary (R)	**1**	2	3	3	8
Stranger	**1**	1	2	0	3

Table 4: Overview of the number of correct answers by users, close adversaries, close adversaries with research (R) and strangers for each session (S) and for the three question types: predefined (P), guided (G) and open (O). Adversaries only provided answers in the first session.

Exemplary questions that close adversaries answered correctly are: *“In which street did my grandma live?”* (guessed by spouse) or *“In which building was my first lecture?”* (guessed by university friend). Close adversaries were allowed to research the answers to questions that they had previously answered incorrectly. Only two adversaries succeeded. Each of them found the answer to one question. They needed one and two attempts, respectively.

8.3.3 Strangers

Two strangers tried to research the answers to the questions. Both of them failed most of the time. One of them was able to research the answers to two questions, while the other one succeeded only for one question. They needed one to two attempts. None of them had more than one correct answer within the set of three questions of a user. The questions that they attacked successfully were *“Where was the first time I partied when I was a student?”*; *“Where did I meet my best friend?”* and *“In which building was my first lecture?”* .

The reasons why adversaries selected a particular location are all based on different assumptions. For example, for the last question, the corresponding adversary was assuming that the user was a student at the department he was working at. Thus, he selected common university buildings where students usually have classes. For the question about the best friend, the adversary assumed that the user had met the victim in high school and thus, selected various school buildings in the home town of the user.

Despite the availability of social networks and search engines, both adversaries stated that it was very difficult to research the answers.

8.3.4 Comparison between Users, Adversaries and Question Types

For each session, we conducted a two-way ANOVA to examine the effects of *question type* and *user type* (i.e. user, close adversary and stranger) on the number of correctly answered questions. Simple main effects analysis showed that users were significantly better in answering questions than adversaries for all three session (each p < 0.01). No significant effects between the different types of adversaries were found. We also did not find any interaction effects.

8.4 Answer Distances

For the following calculations, we took, for each question, the shortest distance (of all attempts per participant) to the actual solution. This was done to analyze how often users and adversaries were how far from the correct answers. While most answers by users (for all three session) were close to the original location, most answers by adversaries were far from the original location, meaning that it was difficult to guess the correct answer.

8.4.1 Users

The answers of eighty questions from the first session were within a range of 30 meters to the original location (and thus, were answered correctly). For the remaining answers, we found the following distances: Seven had a distance of 40-100 meters; two had a distance of 300-600 meters and one answer had a distance of several kilometers.

Similar observations were made for the second session. Seventy-nine questions were answered correctly and thus in the range of 30 meters. The distances of the incorrect answers from the original location were 40-100 meters for four questions, 100-400 meters for five questions and over one kilometer for two questions.

For the third session, 76 answers were within a distance of 30 meters. In turn, nine answers had a distance between 40-100 meters, two had a distance between 200-600 meters and three had a distance of several kilometers.

8.4.2 Close Adversaries and Strangers

Most close adversaries provided answers that were very far from the actual location. Seventy-two answers had a distance of multiple kilometers (average: 440.4 kilometers). Twelve answers had a distance of 200-900 meters, while six answers were within a distance of 30 meters.

The distance distributions for strangers were similar. The distances were multiple kilometers (average: 1176.7 kilometers) for 167 of the 180 questions (since two adversaries attacked 90 questions each). Nine answers had a distance of 300-800 meters, one answer had a distance of 50 meters and three answers were within 30 meters to the original location.

8.5 Authentication Time

The time measurement for enrollment/authentication started when users pressed the start-button to open the corresponding HTML-page and ended with the submission of the last answer to the last question. On average, users needed four minutes for enrollment. The fastest user needed 1 min 15 s, while the longest enrollment lasted 7 min. In the first session, users needed on average 36 s for authentication (min=12 s; max=214 s). For the second and third session, they needed on average 45 s (min=13 s; max=225 s) and 47 s (min=13 s; max=232 s), respectively.

8.6 Accuracy

Accuracy is a good indicator on how well a system works in terms of usability and security. It takes into account the number of true positives (TP), true negatives (TN), false negatives (FN) and false positives (FP). TP refers to the number of successful authentications by legit users, while TN refers to the number of failed attacks by adversaries. In turn, FN represents the number of unsuccessful authentication attempts by legit users, while FP depicts the number of successful attacks by adversaries. The formula can be described as follows:

$$Accuracy = \frac{\sum TP + \sum TN}{\sum TP + \sum FP + \sum TN + \sum FN}$$

The formula returns a value between 0 and 1 (100 in percent). A value of 0 means that all authentication attempts by users fail, while all attacks by adversaries succeed. A value of 1 means the opposite and is a desirable result. It should be noted that accuracy values should always be interpreted in combination with the number of FP and FN.

For each session we calculated the accuracy values using a distance threshold of 30 meters. We also took into account two different parameters: a) the number of correct answers that are required in order to authenticate successfully [1..3] and b) the number of maximum attempts that one has to answer a question [1..3]. An overview of the calculations can be found in the appendix A.1 - A.3.

8.6.1 Close Adversaries

For the first session, the best accuracy values, when considering attacks by close adversaries only, are yielded when two correct answers were required and when there were two or three attempts allowed per question. These combinations result in an accuracy value of 98.3% (0FP, 1FN). The accuracy value remains the same after one week (i.e. second session), when allowing up to three attempts to submit the answer to a question. Restricting the number of attempts to two, decreases the accuracy to 96.7% (0FP, 2FN).

In the third session, the accuracy values decrease to 95% (0FP, 3FN) for the combinations of two required answers and two/three attempts for each question. However, still none of the close adversaries succeed in their attack.

The best combinations, the corresponding accuracy values as well as the number of FP and FN remain the same for all three sessions, when allowing close adversaries to research the answers to questions.

8.6.2 Strangers

The best accuracy values, when taking into account attacks by strangers only, are found for two required answers and two or three attempts. These combinations yield an accuracy of 98.3 % (0FP, 1FN). The accuracy value remains stable after one week, when three attempts are allowed. Allowing only two attempts, increases the number of FN and decreases the accuracy to 96.7% (2FN). In the third session, the accuracy value is 95% (0FP, 3FN) for a combination of two required questions and two/three attempts.

Based on the increasing number of FN after the third session, for both types of considered adversaries, we had a closer look at those three users and the distances to the actual solution for the questions that they answered incorrectly. The distances of the first users were between 120 m and 250 m. They were <40 m for the second user and <65 m for the third user.

Increasing the distance threshold to over 250 m would result in more FP and thus, is not reasonable. Also, choosing a distance threshold of 65 m would increase the number of FP to one when attacks by strangers are considered and thus, is not appropriate as well. In turn, using a distance threshold of 40 m does not increase the number of FP and reduces the number of FN by one. This results in an accuracy value of 96.7% (0FP, 2FN) for the third session and a combination of two required answers and two/three attempts.

8.7 Perceived Memorability

During the first session we asked users if they think that they could recall the answers to their question after a longer period of time. Users affirmed this for 78 of 90 questions (87%). They did not agree in three cases (3%) and were neutral for nine of the questions (10%).

During the second and third session, we then asked them to state how well they could recall the answers. For the second session, they stated to have no problems at all for 73 of 90 questions (81%). For 6 questions (7%) they had to think for some time before recalling the answer, and for 11 questions (12%) they had forgotten the answer.

Similar results were found for the third session. For 71 of 90 questions (79%) they had no problems at all. For 8 questions (9%) they had to think about the questions for some time. They had no idea for 11 of the questions (12%).

8.8 Perceived Security

In the first session, users were asked to rate the security of their questions with respect to different types of adversaries. Users provided their ratings on a 5-point Likert scale from strongly disagree (1) to strongly agree (5).

When asked, if they think that their questions are guessable or researchable by close adversaries, the opinions were not clear. For 34 of 90 questions (38%), users thought that their answers were not guessable, while others thought for 35 questions (39%) that they were. The remaining users were neutral.

In terms of researchability, users thought for 48 questions (53%) that their questions were not researchable by close adversaries. Other users did not share this opinion and believed for 27 questions (30%) that the answers could be researched. The remaining users had a neutral position.

For almost all questions (99%) users did not believe that they could be guessed by strangers. They also thought for 85 questions (94%) that they were not researchable by strangers.

8.9 Perceived Ease of Guessing/Researching

During the first session we asked close adversaries to state whether they knew or guessed the answers to the questions. For 46 of 90 questions (51%), adversaries had to guess the answer. Some adversaries had speculations for 25 questions (28%), but only in one case the correct answer was provided. There were some adversaries who thought to know the approximate location for 13 questions (14%), but only in one case the answer was correct. For the remaining six questions (7%), the adversaries were sure about the question's answer and thus, all but two submitted the correct answers.

Interestingly, some close adversaries were sure about their answers as they were part of the actual memory. For example, when they had the same favorite vacation destination as their spouse they tried to attack.

Close adversaries who did not manage to guess the correct answer, were asked if they felt like knowing the answer after they were allowed to research the question. Even after research, the adversaries did not know the answer for 45 of 84 questions (54%). For all these questions the incorrect answer was provided. Some adversaries stated to have had some kind of feeling where the answer might be for 26 questions (31%). Despite their feeling, all but one of these questions were answered incorrectly. Other adversaries thought to know the approximate location after research for 11 questions (13%), but had no correct answers. Only few adversaries were sure about the answers of two questions (2%). All these questions were answered correctly. The adversaries who succeeded in researching the answers stated that they had found the location on social networks.

8.10 Rating of System

In general, users liked the presented concept in terms of time consumption, memorability and security. The majority (21 users) found that location-based security questions are not too time consuming. They also felt that it is more secure than traditional security questions (27 users) as well as more memorable (24 users). All users stated that they would use location-based security questions for their accounts. Fourteen users for all of their accounts, 14 users for their important accounts and 2 users at least for their unimportant accounts.

9. RESULTS: LONG-TERM EVALUATION

9.1 Number of Correct Answers

Six months after the last session, users answered 55 out of 72 questions (76%) correctly. Seventeen questions (24%) were answered incorrectly. The number of attempts needed varied among users. Most of them needed one attempt (11 users for the first and third question, 14 users for the second question), while others needed two attempts (5 for the first question, 7 for the second questions and 2 for the third question) or one attempt (one user for the third question). None of the users failed in all three questions. Eleven users had all questions correct, nine had two correct answers and four had one correct answer. Most incorrect answers were caused by imprecise selections where users were close to the original location, but not within the required threshold. Another reason was that users had forgotten their answers.

Similar to the previous sections, we calculated the answer distances for each question by using the shortest distance of all attempts. Most answers were within a distance of 30 meters to the actual location. However, 17 questions were answered incorrectly. The answer distances were as follows: Five answers had a distance between 40-100 meters, five questions had a distance between 100-700 meters and two answers had a distance of multiple kilometers.

9.2 Accuracy

Accuracy calculation was done as explained previously. An overview of the accuracy calculations can be found in the appendix B. When considering attacks of close adversaries only, the best combination yields an accuracy value of 91.7% (0FP, 4FN) and requires users to answer at least two answers correctly and gives them three/two attempts per question to provide the answer. The same values and parameters work best when only attacks by strangers are considered.

9.3 Perceived Memorability

Users were asked to state how well they were in recalling the answers to their questions. For most of the questions (49 out of 72), users had no problems at all. For ten questions they had to think some time before the answer was recalled, while they had forgotten the answers to 13 questions. The self-assessment complies with the actual performance of the users. Only in four cases, users claimed to have recalled the answers, but gave an incorrect answer instead. Analyzing the distances of the corresponding answers showed that those users were close to the original answers (between 80-213 meters), but not within the required threshold.

9.4 User Feedback

In general, users felt positive about the presented system and found it more usable than traditional security questions. Thus, they would consider using location-based questions in a real-world deployment. However, one of the main concerns that users raised was the precision with which an answer had to be provided. This criticism was not related to the threshold of 30 meters, but instead, the knowledge that such a narrow threshold is given. Several users noted that they would have paid more attention during enrollment if they had known that such a threshold was given.

With respect to the ease with which an answer could be recalled during the different sessions, most users stated to have no major difficulties. They also told us that the ease of recall did not change over time, meaning that answers that they found easy were easy to recall over all three sessions and the other way around. These statements comply with the observations we have made.

Further interesting remarks were made by two participants who were caught by surprise as the map section that they needed had been updated since their last authentication attempt. As a consequence, some orientation points were lost (e.g. buildings) so that they had to think some time before the answer could be provided.

10. DISCUSSION

10.1 Question Type

In our user study we tested three question types: predefined, guided and open. The analysis did not reveal any significant differences, which may be due to the small number of participants per group. Nonetheless, guided questions appear to be the most promising ones of the three.

The lack of guidance for open questions may lead users to define weaker (but not meaning weak) questions than for the other types. For example, two users defined the question *"Where is my mother born?"* which reminds of the common security question *"What is your mother's maiden name"*. This kind of information could be researched through public records, providing hints to potential adversaries (though it is still difficult to select the location within a given distance threshold). In turn, predefined questions do not leave room for users to adapt the questions to their personal needs and thus, they miss the opportunity to phrase a more memorable question. Hence, the use of guided questions seems to be a good trade-off between the two extremes.

10.2 Topics of Question

The topics covered for the different question types (i.e. predefined, guided and open) were similar and ranged from travel, third persons to special activities. The topics were close to the ones that users like to choose for traditional security questions (i.e. preferences and questions about family members) [10]. However, in terms of guessability by close persons, our approach yields much better results (9%) than traditional security questions (38%; e.g. [6]).

More interestingly, allowing users to phrase their own location-based questions (i.e. in case of guided or open questions) gives the questions a more personal notion which is mirrored in the amount of information that is required to answer a question (e.g. *"Where is the center of the route to my best childhood friend?"*). Requiring more information makes it probably even more difficult for adversaries to guess or research the answers.

Thus, when designing location-based security questions, one could think of extending the set of guidelines by encouraging users to create more complex questions. For example, asking them to define a question that involves the center of two locations. However, it is also essential not to limit users too much in phrasing their questions to ensure applicability of the selected guidelines. This is important to avoid users phrasing questions that meet the restrictions of the guidelines, but may not be memorable.

10.3 User Performance

Users in our study were very confident that they will be able to recall their answers after a longer period of time and had a good estimation about their future performance. This is encouraging, since a positive and realistic attitude toward a system will motivate users to pay some effort when defining location-based security questions. A contrary example are commonly used security question that most users are not willing to spend time answering, since they think that they will not remember the answers anyway.

With respect to recall, the majority of our users were good in answering their security questions and only few forgot the answer to a question (even after six months). This shows that our approach works very well in terms of memorability.

10.4 Adversary Performance

Our approach showed promising results in terms of security as the adversaries in our study performed badly and could only attack few of the questions. This was mainly because the answers to the security questions were difficult to research and thus, forced adversaries to guess the answers at most times. In particular, strangers had problems during research. Even close adversaries who thought to have found clues during research failed to provide the correct answers at most times. They either drew the wrong conclusions from their research or were close to the location, but not within the required distance threshold. In comparison to the analysis by Ariel Rabkin [10] where 12% of the security questions sample could be attacked through research, our close adversaries and strangers could only succeed in 2% of the cases.

The biggest threats come from adversaries that share the same or similar experiences. For example, when close adversaries and users have traveled together to the location the question is referring to or when the user and adversary (close one or stranger) have attended the same course of studies in the same city. Most of the questions that the strangers guessed successfully, would not have been possible if the corresponding adversary had not been in the same situation in the past and thus, had some advanced knowledge.

10.5 Answer Precision

Most errors were made in terms of precision, meaning that users were close the the actual location, but not within the required range. However, the problem was, in the majority of cases, not caused by memorability reasons or the strict threshold, but by the assumption of the users that the system was more tolerant of imprecise selections. During the interviews, users were confident that if they had known about this requirement, they would have had less difficulties during authentication.

This means that when designing location-based systems for fallback authentication, it is important to inform users about the required precision to reduce the number of false negatives. Since most precision errors were already done shortly after enrollment (and then repeated in other authentication sessions), one could think of improving the enrollment procedure. For example, the system could ask users to re-enter the location to a question when a marker has been set to verify their answer. This approach is similar to the verification of passwords during registration.

10.6 Answer Distances

An answer was considered as correct when it was within a distance of 30 meters to the actual location. This threshold worked well to distinguish between users and adversaries. Despite the fact that most answers were clustered within a geographical region centered around the user's hometown, most adversaries were not able to guess the answer. They were hundreds of kilometers away from the actual location, supporting the assumption that most of them probably just selected random locations. Even in cases where adversaries stated to know the approximate location, they failed most of the time. This shows that while some users know the region in which an answer has to be in, it is still very hard to know which location within this region the user has selected. Thus, our approach has a very good answer space entropy. In turn, traditional security questions often have a very limited answer space (smaller than 25) [13].

10.7 Perceived Security

The perceived security of users strongly depend on the type of adversary. While they think that most strangers will not be able to guess or research their questions, their opinion is not as clear for close adversaries. In general, users seem to consider close adversaries as more likely to know an answer to a question than strangers. If a close adversary is considered as harmful, probably depends on how often they interact with the user and how much information this user is willing to share with friends in general. In comparison to the actual performance of adversaries, there is no difference in the number of answers they are able to answer, thus our approach works equally well against both types of adversaries.

10.8 Accuracy

To analyze the interplay between usability and security, we calculated the accuracy values for our approach. The best combination requires users to answer at least two out of three questions correctly, allowing them three/two attempts per question. This combination yielded an accuracy value of 91.7% with 4 FN and 0 FP after six months (with an increase of only 1FN in comparison to the third session).

This means that in terms of usability, our approach yielded good values, but leaves room for improvement, since still a few users were not able to authenticate under these conditions. In a real-world deployment, one would have to provide these users an alternative for the fallback authentication. This approach is commonly used for web services where users can select from a set of different fallback authentication schemes.

In terms of security, our evaluation obtained a very desirable result, since no adversaries (close ones as well as strangers) were able to attack successfully. However, we must also take into account that the number of attacks we were able to consider in this paper was limited.

Based on the usability and security insights, it would be interesting to study, if increasing the number of attempts decreases the number of FN, while maintaining the number of FP and if these improvements are resistant to a larger number of attacks, since increasing the number of attempts also means to give adversaries more opportunities for guessing the correct answers. As the answer space of location-based security question is huge, we assume that slightly increasing the number of attempts does not have a big impact on the actual security. However, these questions need to be addressed in the future.

10.9 Limitations

The participants of our focus group were all male which could have had an influence on the identified topics for the design of the questions. Literature on gender differences for autobiographical memories are ambiguous. While some assume no differences, others find woman to have more vivid and precise memories. If the latter is the case, we only have a lower bound for location-based questions, which, however, is good for general applicability.

Although a larger study sample would have been desirable, we opted for long-term participants (opposed to many participants for a one time lab-session), since we believe that this allowed us to get better insights into the potentials and shortcomings of location-based questions. In addition to this, the majority of participants were quite young with diverse backgrounds, but mostly students. Thus, it would be interesting to evaluate the concept with a larger and older sample of participants, since younger and older people remember different types of episodic memories [11]. Therefore, we encourage further studies with a more diverse sample.

Though we were able to re-invite the majority of participants, the dropout rate after six months needs to be mentioned as another limitation.

11. CONCLUSION

In this paper, we proposed the use of location-based security questions as a new approach for fallback authentication, and as an alternative to open text-based security questions that are known for their usability and security issues. We presented the design, implementation and evaluation of this approach and tested the location-based security questions under the worst circumstances. The results reported in this paper highlight the potential of the presented approach.

While users are good in recalling the location-answers to their questions, adversaries (close ones as well as strangers) failed most of the time when attacking these questions. Furthermore, the problems reported by our users are helpful guidelines to be considered when designing location-based questions for a real-world deployment. Since the accuracy values as well as the number of false positives and false negatives are promising, we believe that the presented approach has the potential to replace commonly used security questions in the future and thus, encourage further research in this area to optimize the questions and the overall parameters for deployment.

12. REFERENCES

[1] A. Adams and M. A. Sasse. Users are not the enemy. *Commun. ACM*, 42(12):40–46, Dec. 1999.

[2] A. Babic, H. Xiong, D. Yao, and L. Iftode. Building robust authentication systems with activity-based personal questions. In *Proc. SafeConfig 2009*, pages 19–24. ACM Press, 2009.

[3] S. L. Garfinkel. Email-based identification and authentication: An alternative to pki? *IEEE Security & Privacy*, 1(6):20–26, 2003.

[4] V. Griffith and M. Jakobsson. Messin' with texas deriving mother's maiden names using public records. In *Proc. ACNS 2014*, pages 91–103. Springer, 2005.

[5] R. Gross and A. Acquisti. Information revelation and privacy in online social networks. In *Proc. WPES '05*, pages 71–80. ACM Press, 2005.

[6] W. J. Haga and M. Zviran. Question-and-answer passwords: An empirical evaluation. *Information Systems*, 16(3):335 – 343, 1991.

[7] M. Just. Designing and evaluating challenge-question systems. *IEEE Security & Privacy*, 2(5):32–39, 2004.

[8] M. Just and D. Aspinall. Personal choice and challenge questions: A security and usability assessment. In *Proc. SOUPS 2009*, pages 8:1–8:11. ACM Press, 2009.

[9] I. Muslukhov, Y. Boshmaf, C. Kuo, J. Lester, and K. Beznosov. Know your enemy: The risk of unauthorized access in smartphones by insiders. In *Proc. MobileHCI 2013*, pages 271–280. ACM Press, 2013.

[10] A. Rabkin. Personal knowledge questions for fallback authentication: Security questions in the era of facebook. In *Proc. SOUPS 2008*, pages 13–23, New York, NY, USA, 2008. ACM Press.

[11] H. L. Roediger and E. J. Marsh. Episodic and autobiographical memory. In A. F. Healy and R. W. Proctor, editors, *Handbook of psychology (Volume 4: Experimental Psychology)*, pages 475–497. John Wiley and Sons, 2003.

[12] S. Schechter, A. J. B. Brush, and S. Egelman. It's no secret: Measuring the security and reliability of authentication via 'secret' questions. In *Proc. SOUPS 2009*, pages 40:1–40:1. ACM Press, 2009.

[13] S. Schechter, S. Egelman, and R. W. Reeder. It's not what you know, but who you know: A social approach to last-resort authentication. In *Proc.CHI 2009*, pages 1983–1992. ACM Press, 2009.

[14] E. Stobert and R. Biddle. The password life cycle: User behaviour in managing passwords. In *Proc. SOUPS 2014*, pages 243–255. USENIX, 2014.

[15] J. Thorpe, B. MacRae, and A. Salehi-Abari. Usability and security evaluation of geopass: A geographic location-password scheme. In *Proc. SOUPS 2013*, pages 14:1–14:14. ACM Press, 2013.

[16] E. Tulving. Availablity versus accessibility of information in memory for words. *Verbal Learing and Verbal Behavior*, 5:381–391, 1966.

[17] E. Tulving. Episodic and semantic memory. In *Organization of Memory*, pages 381–402. Academic Press, 1972.

[18] Yahoo! Help. Recovering a lost or forgotten password. `https://help.yahoo.com/kb/recovering-lost-forgotten-password-sln2047.html` (Accessed: 02/09/2014).

APPENDIX

A. ACCURACY VALUES: SESSION 1, 2 AND 3

A.1 Session 1

A.1.1 Close Adversaries

Table 5: Overview of the accuracy values (A) for the first session. The calculation uses a distance threshold of 30 and takes into account the attacks by close adversaries as well as the number of required answers (ANS) and the maximal number of allowed attempts (ATT) in order to authenticate successfully.

Answer	**3**			**2**			**1**		
Attempt	3	2	1	3	2	1	3	2	1
TP	21	19	16	29	29	28	30	30	30
TN	30	30	30	30	30	30	24	25	26
FP	0	0	0	0	0	0	6	5	4
FN	9	11	14	1	1	2	0	0	0
Accuracy	85,0%	81,7%	76,7%	98,3%	98,3%	96,7%	90,0%	91,7%	93,3%

A.1.2 Stranger

Table 6: Overview of the accuracy values (A) for the first session. The calculation uses a distance threshold of 30 and takes into account the attacks by strangers as well as the number of required answers (ANS) and the maximal number of allowed attempts (ATT) in order to authenticate successfully.

Answer	**3**			**2**			**1**		
Attempt	3	2	1	3	2	1	3	2	1
TP	21	19	16	29	29	28	30	30	30
TN	30	30	30	30	30	30	27	27	29
FP	0	0	0	0	0	0	3	3	1
FN	9	11	14	1	1	2	0	0	0
Accuracy	85,0%	81,7%	76,7%	98,3%	98,3%	96,7%	95,0%	95,0%	98.3%

A.2 Session 2

A.2.1 Close Adversaries

Table 7: Overview of the accuracy values (A) for the second session. The calculation uses a distance threshold of 30 and takes into account the attacks by close adversaries as well as the number of required answers (ANS) and the maximal number of allowed attempts (ATT) in order to authenticate successfully.

Answer	**3**			**2**			**1**		
Attempt	3	2	1	3	2	1	3	2	1
TP	20	20	16	29	28	27	30	30	30
TN	30	30	30	30	30	30	24	25	26
FP	0	0	0	0	0	0	6	5	4
FN	10	10	14	1	2	3	0	0	0
Accuracy	83,3%	83,3%	76,7%	98,3%	96,7%	95,0%	90,0%	91,7%	93,3%

A.2.2 Stranger

Table 8: Overview of the accuracy values (A) for the second session. The calculation uses a distance threshold of 30 and takes into account the attacks by strangers as well as the number of required answers (ANS) and the maximal number of allowed attempts (ATT) in order to authenticate successfully.

Answer	**3**			**2**			**1**		
Attempt	3	2	1	3	2	1	3	2	1
TP	20	20	16	29	28	27	30	30	30
TN	30	30	30	30	30	30	27	27	29
FP	0	0	0	0	0	0	3	3	1
FN	10	10	14	1	2	3	0	0	0
Accuracy	83,3%	83,3%	76,7%	98,3%	96,7%	95,0%	95,0%	95,0%	98,3%

A.3 Session 3

A.3.1 Close Adversaries

Table 9: Overview of the accuracy values (A) for the third session. The calculation uses a distance threshold of 30 and takes into account the attacks by close adversaries as well as the number of required answers (ANS) and the maximal number of allowed attempts (ATT) in order to authenticate successfully.

Answer	**3**			**2**			**1**		
Attempt	3	2	1	3	2	1	3	2	1
TP	19	18	16	27	27	25	30	30	30
TN	30	30	30	30	30	30	24	25	26
FP	0	0	0	0	0	0	6	5	4
FN	11	12	14	3	3	5	0	0	0
Accuracy	81,67%	80,00%	76,67%	95,00%	95,00%	91,67%	90,00%	91,67%	93,33%

A.3.2 Stranger

Table 10: Overview of the accuracy values (A) for the third session. The calculation uses a distance threshold of 30 and takes into account the attacks by strangers as well as the number of required answers (ANS) and the maximal number of allowed attempts (ATT) in order to authenticate successfully.

Answer	**3**			**2**			**1**		
Attempt	3	2	1	3	2	1	3	2	1
TP	19	18	16	27	27	25	30	30	30
TN	30	30	30	30	30	30	27	27	29
FP	0	0	0	0	0	0	3	3	1
FN	11	12	14	3	3	5	0	0	0
Accuracy	81,7%	80,0%	76,7%	95,0%	95,0%	91,7%	95,0%	95,0%	98,3%

B. ACCURACY VALUES: AFTER SIX MONTHS

B.1 Close Adversaries

Table 11: Overview of the accuracy values (A) after six months. The calculation uses a distance threshold of 30 and takes into account the attacks by close adversaries as well as the number of required answers (ANS) and the maximal number of allowed attempts (ATT) in order to authenticate successfully.

Answer	**3**			**2**			**1**		
Attempt	3	2	1	3	2	1	3	2	1
TP	11	10	5	20	20	14	24	24	21
TN	24	24	24	24	24	24	19	19	20
FP	0	0	0	0	0	0	5	5	4
FN	13	14	19	4	4	10	0	0	3
A	72,9%	70,8%	60,4%	91,7%	91,7%	79,2%	89,6%	89,6%	85,4%

B.2 Strangers

Table 12: Overview of the accuracy values (A) after six months. The calculation uses a distance threshold of 30 and takes into account the attacks by strangers as well as the number of required answers (ANS) and the maximal number of allowed attempts (ATT) in order to authenticate successfully.

Answer	**3**			**2**			**1**		
Attempt	3	2	1	3	2	1	3	2	1
TP	11	10	5	20	20	14	24	24	21
TN	24	24	24	24	24	24	22	22	24
FP	0	0	0	0	0	0	2	2	0
FN	13	14	19	4	4	10	0	0	3
A	72,9%	70,8%	60,4%	91,7%	91,7%	79,2%	95,8%	95,8%	93,8%

The Impact of Cues and User Interaction on the Memorability of System-Assigned Recognition-Based Graphical Passwords

Mahdi Nasrullah Al-Ameen, Kanis Fatema, Matthew Wright, Shannon Scielzo
The University of Texas at Arlington
Arlington, TX, USA
{mahdi.al-ameen, kanis.fatema}@mavs.uta.edu, mwright@cse.uta.edu, scielzo@uta.edu

ABSTRACT

User-chosen passwords reflecting common strategies and patterns ease memorization, but offer uncertain and often weak security. System-assigned passwords provide higher security, and thus in commercially deployed graphical-password systems (e.g., Passfaces), images are randomly assigned by the system. It is difficult, however, for many users to remember system-assigned passwords. We argue that this is because existing password schemes do not fully leverage humans' cognitive strengths, and we thus examine techniques to enhance password memorability that incorporate scientific understanding of long-term memory. In our study, we examine the efficacy of *spatial cues* (fixed position of images), *verbal cues* (phrases/facts related to the images), and employing *user interaction* (learning images through writing a short description at registration) to improve the memorability of passwords based on face images and object images. We conducted a multi-session in-lab user study with 56 participants, where each participant was assigned seven different graphical passwords, each representing one study condition. One week after registration, participants had a 98% login success rate for a scheme offering spatial and verbal cues, while the scheme based on user interaction had a 95% login success rate for face images and a 93% login success rate for object images. All of these were significantly higher than the control conditions representing existing graphical password schemes. These findings contribute to our understanding of the impact of cues and user interaction on graphical passwords, and they show a promising direction for future research to gain high memorability for system-assigned random passwords.

Keywords

System-assigned graphical password, memorability,
cued-recognition, user interaction

Symposium on Usable Privacy and Security (SOUPS) 2015, July 22–24, 2015, Ottawa, Canada.

1. INTRODUCTION

Traditional user-chosen textual passwords suffer from security problems because of password reuse and predictable patterns [13, 42]. Users bear the responsibility of ensuring security of their account by creating a password that should be chosen with creativity and intelligence so that it achieves satisfactory security and memorability. For many users, this is a lot of work, and in many cases they compromise on security and create a weak but memorable password.

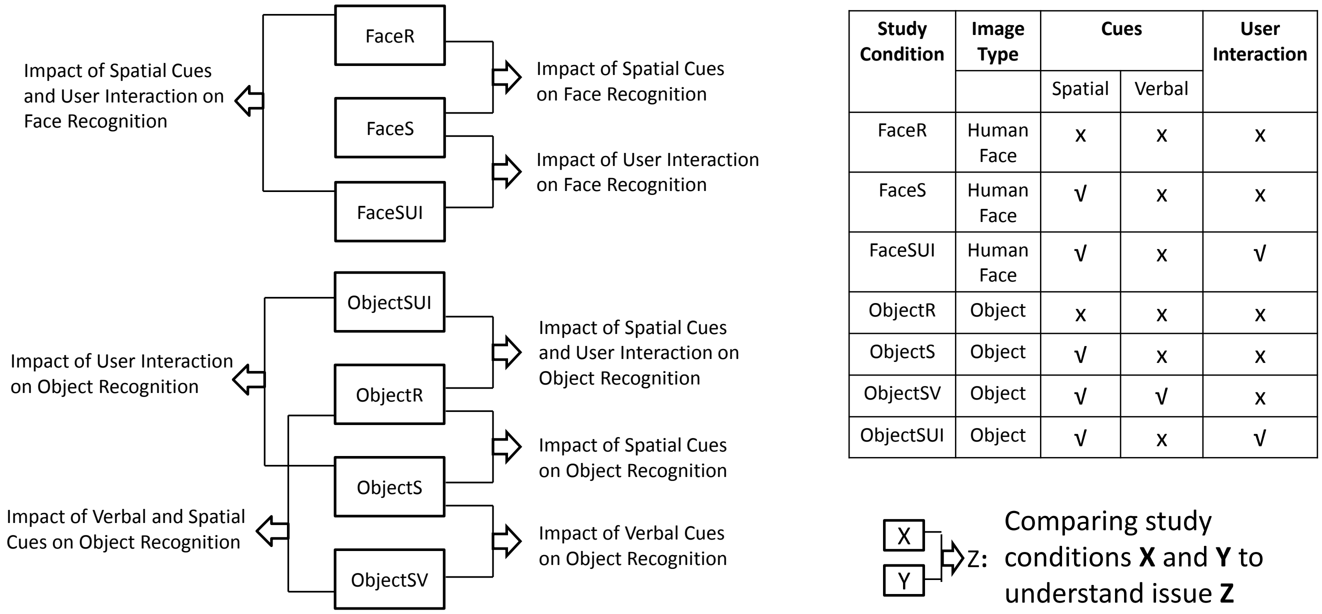
A recent study [40] reveals that with the advancement of digital technology and widespread use of the Internet, users now more than ever realize the importance of strong passwords, and many of them intend to create secure passwords but still fail to achieve a good balance between security and memorability. Policies requiring users to create longer passwords with different character types do not necessarily lead to more secure passwords, but they do adversely affect memorability in some cases [36, 42].

A number of important cognitive propensities have been considered by researchers to explore better alternatives to traditional textual passwords. For example, recognition is an easier memory task than recall [6, 46, 47], and due to the *picture superiority effect*, the human brain is better at memorizing graphical information as compared to textual information [33, 35]. These are the core ideas behind the design of recognition-based graphical passwords, such as Passfaces [1], which is now commercially available and deployed by a number of large websites.[1]

In recognition-based graphical passwords, such as Passfaces [1], users are shown several portfolios of faces (e.g., four portfolios of nine faces each), and one face per portfolio serves as the authentication secret that they have to recognize during login. Previously, users in Passfaces could select images from the portfolio for their authentication secret. Davis et al. [14], however, found that users select predictable images. As a result, the commercial Passfaces [1] product now assigns a random image for each portfolio instead of allowing users to choose.

With system-assigned passwords, the user does not have to guess whether a password is secure, and the system can ensure that all passwords offer the desired level of security. Additionally, while password reuse could pose a serious security threat [13], using system-assigned passwords ensures that users do not reuse a password (or modification thereof) already used on another account. Unfortunately, it is dif-

[1] http://www.realuser.com/ shows testimonials about Passfaces from customers, such as banks.


FaceR
FaceS
FaceSUI
Impact of Spatial Cues and User Interaction on Face Recognition
Impact of Spatial Cues on Face Recognition
Impact of User Interaction on Face Recognition
ObjectSUI
ObjectR
ObjectS
ObjectSV
Impact of User Interaction on Object Recognition
Impact of Verbal and Spatial Cues on Object Recognition
Impact of Spatial Cues and User Interaction on Object Recognition
Impact of Spatial Cues on Object Recognition
Impact of Verbal Cues on Object Recognition
X
Y
Z:
Comparing study conditions **X** and **Y** to understand issue **Z**

Study Condition	Image Type	Cues		User Interaction
		Spatial	Verbal	
FaceR	Human Face	x	x	x
FaceS	Human Face	√	x	x
FaceSUI	Human Face	√	x	√
ObjectR	Object	x	x	x
ObjectS	Object	√	x	x
ObjectSV	Object	√	√	x
ObjectSUI	Object	√	x	√

Figure 1: The Study Model to Understand the Impact of Cues and User Interaction on Recognition-based Graphical Authentication

ficult for most people to memorize system-assigned passwords [17, 41, 49]. Thus, it still remains a critical challenge to design an authentication scheme that offers satisfactory memorability for system-assigned random passwords.

The commercial deployment of recognition-based graphical passwords (e.g., Passfaces [1]) and its demonstrated potential mean that improvements to such schemes would be very valuable contributions. In this paper, we aim to incorporate the scientific understanding of long-term memory to advance the memorability of system-assigned recognition-based graphical passwords.

1.1 Contributions

To this end, we draw upon several prominent theories of cognitive psychology to enhance the memorability of system-assigned recognition-based graphical passwords. In particular, we examine the impact of using memory cues, including *spatial cues* in which images in a portfolio are shown in the same position each time and *verbal cues* in which each image is presented with a phrase or fact related to the image. The use of cues facilitates a detailed encoding that helps to transfer the authentication information (e.g., assigned images) from the working memory to long-term memory at registration [7], helping users recognize their images when logging in later. We call this approach *cued-recognition* [5].

We also explore the efficacy of requiring user interaction at registration, in which we have users apply their observation and imagination to type a short description about assigned images. In the course of such observations and thinking on the assigned images, users get more familiar with them and consequently succeed to recognize those images from the set of decoys during login. This process engages users' action-event memory [29], in addition to their visual memory [33, 35], and aids in the elaborate encoding of the authentication secret in long-term memory [7]. We provide a detailed discussion on these memorization processes in §3.

Considering both human faces and objects as images, along with cues and interaction, we design seven different study conditions (see Figure 1). In our within-group study with 56 participants, every participant was assigned seven different graphical passwords, each representing one study condition. The major findings from our study include:

- Verbal cues make a significant contribution in improving the memorability for object-recognition-based graphical passwords.
- Spatial cues do not contribute significantly to improve memorability for either face or object recognition.
- User interaction is an effective approach to enhance memorability for both face and object recognition.

We organize this paper as follows: In §2, we give an overview of the notable authentication schemes with a discussion on their limitations and the scope for possible improvements. In §3, we explain from the perspective of cognitive psychology how the design choices for our study conditions are set up. We then describe our study procedure in §4 and present the results in §5. In §6, we discuss the findings from our study and highlight the possible directions for future research, followed by a conclusion in §7.

2. RELATED WORK

In this section, we give a brief overview of notable textual and graphical password schemes in which we highlight why existing schemes are insufficient. A possible exception is the CuedR scheme of Al-Ameen et al. [5], which inspires the deeper investigation that we undertake in this paper. We end this section by describing our distinct contributions from their work.

2.1 Textual Password Schemes

2.1.1 Traditional passwords

Traditional user-chosen textual passwords are fraught with security problems because of password reuse and predictable patterns [13,42]. Different password restriction policies (e.g., increasing the minimum password length, requiring a combination of different types of characters, and using password strength meters) have been deployed to get users to create stronger passwords [19, 42]. However, in separate studies, Proctor et al. [36] and Shay et al. [42] report that such policies do not necessarily lead to more secure passwords but do adversely affect memorability in some cases.

2.1.2 Mnemonic Passwords

Kuo et al. [30] studied passwords based on mnemonic phrases, in which the user chooses a memorable phrase and uses a character (often the first letter) to represent each word in the phrase. Their results show that user-selected mnemonic passwords are slightly more resistant to brute-force attacks than traditional passwords. However, mnemonic passwords are found to be more predictable when users choose common phrases to create their passwords. A properly chosen dictionary may further increase the success rate in guessing mnemonic passwords [30].

2.1.3 System-assigned passwords

System-assigned random textual password schemes are more secure but fail to provide sufficient memorability, even when natural-language words are used [41, 49]. Wright et al. [49] compared the usability of three different system-assigned textual password schemes: Word Recall, Word Recognition, and Letter Recall. None of these schemes had sufficient memorability rates.

2.1.4 PTP

Forget et al. [20, 21] proposed the Persuasive Text Passwords (PTP) scheme, in which the user first creates a password, and PTP improves its security by placing randomly-chosen characters at random positions into the password. PTP is resilient against attacks exploiting password reuse and predictable patterns. Unfortunately, the memorability for PTP is just 25% when two random characters are inserted at random positions [20].

2.1.5 Cognitive questions

Furnell et al. [23] revealed the potential of cognitive questions and reported a high level of user satisfaction in using them for primary authentication. However, Just and Aspinall [28] showed the usability and security problems of using cognitive questions for authentication, and several other studies [37, 39] point out the vulnerability of this approach to targeted guessing attacks.

2.2 Graphical Password Schemes

Graphical password schemes can be divided into three categories [8], based on the kind of memory leveraged by the systems: i) Drawmetric (recall-based), ii) Locimetric (cued-recall-based), and iii) Cognometric (recognition-based).

2.2.1 Drawmetric

The user is asked to reproduce a drawing in this category of graphical passwords. In *Draw-a-Secret (DAS)* [27], a user draws on top of a grid, and the password is represented as the sequence of grid squares. Nali and Thorpe [32] have shown that users choose predictable patterns in DAS that include drawing symmetric images with 1-3 pen strokes, using grid cell corners and lines (presumably as points of reference) and placing their drawing approximately in the center of the grid.

BDAS [16] intends to reduce the amount of symmetry in the user's drawing by adding background images, but this may introduce other predictable behaviors such as targeting similar areas of the images or image-specific patterns [8]. DAS and BDAS have recall rates of no higher than 80%.

2.2.2 Locimetric

The password schemes in this category present users with one or more images as a memory cue to assist them selecting their particular points on the image(s). In the *Passpoints* [9] scheme, users select a sequence of click-points on a single image as their password. *Cued Click-Points (CCP)* [12] is a modified version of Passpoints, where users sequentially choose one click-point on each of five images. Dirik et al. [15] developed a model that can predict 70-80% of users' click positions in Passpoints. To address this issue, Chiasson et al. proposed *Persuasive Cued Click-Points (PCCP)* [11, 22], in which a randomly-positioned viewport is shown on top of the image during password creation, and users select their click-point within this viewport. The memorability for PCCP was found to be 83-94%.

In a follow-up study, Chiasson et al. [10] found predictability in users' click points, showing that in Passpoints, the click points are roughly evenly spaced across the image, in straight lines starting from left to right, and either completely horizontal or sloping from top to bottom. The authors [10] indicate that predictability is still a security concern for PCCP.

2.2.3 Cognometric

In this recognition-based category of graphical passwords, the user is asked to recognize and identify their password images from a set of distractor images. *Passfaces* [1] is the most studied cognometric scheme as it is commercially deployed by a number of large websites. The commercial Passfaces [1] product assigns a random set of faces instead of allowing users to choose, since the research [14] has found that users select predictable faces, biased by race, gender, and attractiveness of faces. However, Everitt et al. [17] show that users have difficulty in remembering system-assigned Passfaces.

Davis et al. [14] proposed the *Story* scheme, in which users select a sequence of images as their password and, to aid memorability, are encouraged to mentally construct a story to connect those images. During login, users have to identify their images in accurate order from a panel of decoy images. Though the user choices in Story are found to be more varied than the face-recognition-based scheme, the results still display some exploitable patterns, and the user study showed a memorability rate of about 85% [14].

2.3 Cued-recognition

All prior graphical password schemes show either deficit in memorability, security, or both. A cognometric scheme called Cued-recognition (CuedR) was recently proposed by Al-Ameen et al. [5]. CuedR includes spatial and verbal cues designed to aid recognition of the images of objects, and in a lab study with 37 participants, it had 100% memorability one week after registration. This suggests that the use of cues is very promising and motivates further study. In particular, their study relied on user feedback to discern the relative importance of different cues. They did not actually study the impact of different cues in an experiment. Our deeper investigation on this issue, through direct comparisons between schemes offering different combinations of cues, indicates that relying solely on user feedback might not be reliable in this context (see §6 for detailed discussion). Further, the commercially deployed Passfaces scheme uses face images instead of object images, and it is unclear which should be used. We also examine this issue in our study.

3. SYSTEM DESIGN

Passfaces [1] provides *PIN-level* security (13 bits of entropy), while an authentication scheme should offer at least 20 bits of entropy to attain *password-level* security [8]. Hlywa et al. [26] provide a guideline to design recognition-based graphical authentication schemes with password-level security, where the user is assigned five images at registration and has to recognize each of the assigned images from a distinct portfolio of 16 images during login. We follow this guideline to design our study conditions, where a successful authentication requires the user to recognize all five images correctly. For an unsuccessful login, the user is shown an error message at the end of the login attempt but not informed on which portfolio the mistake was made.

In this section, we explain from the perspective of cognitive psychology how our study design is set up to understand the impact of cues and user interaction (at registration) in improving memorability for system-assigned recognition-based (i.e., cognometric) graphical password schemes. We illustrate our study model in Figure 1.

3.1 Visual Memory

In our study, we leverage the *picture superiority effect* [35], which points out that the human brain is better at memorizing graphical information as compared to textual information [33, 35]. Several explanations for this effect have been proposed in psychology research, where *dual-coding theory* [35] is the most widely accepted. According to this theory [35], images are encoded not only visually and remembered as images, but they are also translated into a verbal form (as in a description) and remembered semantically in human memory. Another explanation for the picture superiority effect is *sensory-semantic model* [33], which postulates that images are accessed more easily than the textual information because they are accompanied by more distinct sensory codes.

3.2 Memory Retrieval

Users are required to perform a recognition task in our study, since it is easier to identify the correct item among a set of distractors (i.e., recognition) than reproducing the item from memory (i.e., recall) [46]. This ease in recognition is explained through *Strength theory* [47] and *Generate-recognize theory* [6]. Strength theory [47] simply states that although the same memory tasks are involved in both recall and recognition, recognition requires less effort. According to generate-recognize theory [6], recall consists of two phases:

Generate phase: A list of candidate words is formed by searching long-term memory.

Recognize phase: The list of words (formed in generate phase) is evaluated to see if they can be recognized as the sought-out memory.

Generate-recognize theory postulates that recognition tasks are faster and easier to perform since they do not utilize the generate phase. This can be illustrated by considering exam questions—having the correct answer available for recognition makes multiple-choice questions easier than short-answer questions.

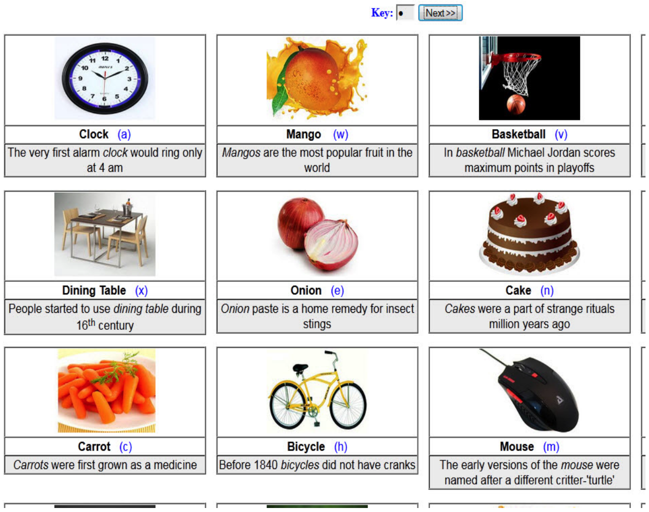
3.3 Face Recognition

In our study, we consider face and object images separately, to understand the impact of cues and user interaction on each image type for recognition-based graphical authentication. Passfaces [1] uses face images, and prior research [24,31] has shown evidence that there may be regions of the human brain dedicated to processing facial information and recognizing faces. Minnebusch et al. [31] demonstrate that three important regions of human brain, *fusiform face area (FFA)*, *superior temporal sulcus (STS)*, and *occipital face area (OFA)*, are activated (recruited) bilaterally (with some right hemisphere bias) while processing facial information. The results of functional MRI show that FFA in the brain gets activated more strongly while viewing faces as compared to other visual objects. STS is sensitive to dynamic aspects of face stimuli, such as gaze or expression. OFA is another important area of human brain, and it deals with the physical features of faces. The findings of Minnebusch et al. [31] are in agreement with the evidence shown by Haxby et al. [24] that suggest that face recognition is functionally different than recognizing other visual objects.

3.4 Long-Term Memory

We incorporate the scientific understanding of long-term memory to advance the usability properties of recognition-based authentication. The cognitive memory model proposed by Atkinson and Shiffrin [7] postulates that users learn new information through the sensory organs, which is then transmitted to their short-term memory (STM). The elaborate processing and encoding of the information, which is held in STM as *memory codes*—mental representations of selected parts of the information, contributes to transferring that information from STM to long-term memory (LTM). This encoding helps people to remember and retrieve the processed information efficiently over an extended period of time. To motivate this encoding, we examine two different approaches in our study:

Cued-recognition: Providing memory cues (e.g., spatial, verbal) with the images, which would be shown both at registration and login.



Clock (a)	**Mango** (w)	**Basketball** (v)
The very first alarm *clock* would ring only at 4 am	*Mangos* are the most popular fruit in the world	In *basketball* Michael Jordan scores maximum points in playoffs
Dining Table (x)	**Onion** (e)	**Cake** (n)
People started to use *dining table* during 16th century	*Onion* paste is a home remedy for insect stings	*Cakes* were a part of strange rituals million years ago
Carrot (c)	**Bicycle** (h)	**Mouse** (m)
Carrots were first grown as a medicine	Before 1840 *bicycles* did not have cranks	The early versions of the *mouse* were named after a different critter-'turtle'

Figure 2: A partial screen shot of ObjectSV scheme during login. The facts corresponding to each image appear below that image. Users enter the key, a lowercase letter shown in parentheses, in the password field (on top) to select the corresponding image. The keys are randomly assigned to images each time the portfolio is loaded, where no two images share the same key. During login, users are shown five such portfolios, where each presents a distinct set of 16 images including one of the five assigned images.

User interaction: Asking users to write a short description about the assigned images during registration. These descriptions would not be shown at login nor stored by the system.

To explore the impact of cues and user interaction on graphical recognition, we design a control condition for face recognition, in which the images in a portfolio remain the same but randomly positioned each time that portfolio is loaded, as in Passfaces [1]. In this paper, we term this control condition *FaceR* (**Face** images with **R**andom positioning). We design a similar control condition for object recognition that we call *ObjectR*.

3.4.1 Cued-Recognition

Based on psychology research [6, 46], we argue that password schemes should ease the memory retrieval of authentication information through providing users with cues, since it is difficult to remember information spontaneously without memory cues. In this regard, the most effective cues are those that are present at the time of remembering [45]. In this paper, we aim to understand the impact of spatial and verbal cues in improving the memorability of cognometric graphical passwords.

Spatial Cues. *Semantic priming* refers to recognizing an object through its relationship with other objects around it [1]. Semantic priming thus eases the recognition task [1], which is augmented by having a fixed set of objects in a certain place. In a graphical password scheme offering spatial cues, the images in a portfolio remain the same and are presented at a fixed position whenever that a portfolio is loaded. For example, in Figure 2, the clock is not only in the upper-left-hand corner each time, but it is always next to the mango and above the dining table. This establishes a relationship between the objects and reinforces semantic priming. Thus, the schemes (except control conditions: FaceR and ObjectR) in our study offer spatial cues, while we design *FaceS* (**Face** Images with **S**patial cues) and *ObjectS* (**Object** Images with **S**patial cues) schemes to understand the precise impact of spatial cues on graphical recognition.

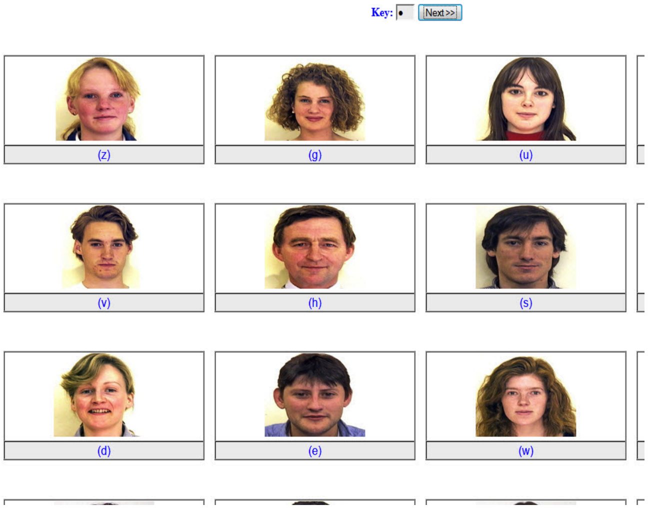


Figure 3: A partial screen shot of FaceSUI during login. Users are shown five such portfolios, and each presents a distinct set of 16 images including one of the five assigned images.

In FaceS and ObjectS, the images in a portfolio remain at the same position each time that portfolio is loaded. We compare FaceS with FaceR and ObjectS with ObjectR to show the impact of spatial cues on face recognition and object recognition, respectively.

Verbal Cues. If the system provides verbal cues, i.e., phrases/facts related to the images, then users may focus their attention on associating the images with the corresponding cues, which should help to process and encode the information to store them in long-term memory. The cues would also assist users to recognize the images in the future and thus enhance their memorability.

In our study, the *ObjectSV* (**Object** images with **S**patial and **V**erbal cues) scheme provides users with verbal cues. For example, the image of a 'Dining Table' is provided with the name of this object ('Dining Table'), and a corresponding phrase/fact ("People started to use dining table during 16th century."). Yan et. al. [50] examined the influence of phrases in increasing the memorability of passwords, which inspires us to accommodate a common phrase or fact for each image as a verbal cue. See Figure 2 showing a partial screen shot of the login screen for ObjectSV.

We typically do not provide a physical description of an image (e.g., "A dining table has four legs.") as a phrase or fact, since it is already visible in the image. Rather, ObjectSV offers an additional fact corresponding to the object (in image) as a verbal cue for helping users to better remember the image through correlating it with the given cues. Thus, we did not accommodate verbal cues for face recognition, since it is not possible to provide users with facts about the anonymous face images.

We compare ObjectSV with ObjectR to examine the memorability gain of combining spatial and verbal cues, and we compare ObjectSV with ObjectS to examine the more precise impact of verbal cues.

3.4.2 User Interaction

In our study, we implement user interaction through the schemes FaceSUI (**Face** images with **S**patial cues and **U**ser **I**nteraction) and ObjectSUI (**Object** images with **S**patial cues and **U**ser **I**nteraction)), in which the system asks users to describe each assigned image during registration. The user interface includes a text field for users to type a short description about the assigned image. The descriptions, provided by the users during registration, are not stored by the system nor shown in any form at login. The sole purpose of this approach is to make random images more familiar to the users through motivating their deeper observations. See Figure 3 showing a partial screen shot of the login screen for FaceSUI scheme.

Unlike existing graphical password schemes, such as Passfaces [1], where users just use their visual memory to memorize the given images, FaceSUI and ObjectSUI schemes leverage both visual memory and action-event memory [29] for a more elaborate encoding of authentication information (e.g., assigned images), which help users to remember and retrieve the processed information efficiently over an extended period of time.

We compare FaceSUI with FaceR and ObjectSUI with ObjectR to examine the memorability gain through combining spatial cues with user interaction, while the comparisons of FaceSUI with FaceS and ObjectSUI with ObjectS reveal the more precise impact of user interaction on face recognition and object recognition, respectively.

3.5 Variant Response

In existing cognometric graphical password schemes [1, 26], mouse input is used to select an image, where the images in a portfolio remain the same but are positioned randomly each time that a portfolio is loaded to compensate for shoulder surfing risk during login. Since the existing recognition-based schemes [26] are presented through our control conditions, FaceR and ObjectR schemes also use mouse input to select images. In fixed-position schemes, i.e., schemes offering spatial cues, mouse input is badly susceptible to shoulder surfing and should not be used. Instead, we use keyboard input, where each time a portfolio is loaded, a distinct lowercase letter `a-z` is assigned randomly as a *key* to one image on the page, and the user inputs the key letter corresponding to her assigned image into a single-character password field to move on to the next portfolio (see Figure 2 and Figure 3). The user-entered letter in the password field is shown as an asterisk to reduce the risk of shoulder surfing.

Keyboard input offers the ability to use *variant response*, in which the user's responses (typed characters) vary for each login session [8]. Tari et al. performed a shoulder-surfing study that showed that cognometric schemes with keyboard input and variant response provide higher resilience to shoulder surfing than schemes with mouse input [43]. Thus, we used keyboard input with variant response for a fair test of reasonably secure conditions compared with the control conditions.

4. USER STUDY

We now present the design of our user study to explore the impact of cues and user interaction on the memorability of recognition-based graphical authentication. In this study, we used a within-subjects design consisting of seven experimental conditions (see Figure 1). Using a within-subjects design controls for individual differences and permits the use of statistically stronger hypothesis tests. The Institutional Review Board (IRB) at the University of Texas at Arlington (UT Arlington) approved the procedures of our user study.

4.1 Participants, Apparatus and Environment

For this experiment, we recruited 56 students (40 women, 16 men) through our university's Psychology Research Pool. Participants came from diverse backgrounds, including majors from Nursing, Psychology, Business, Environmental Science, Biochemistry, and Spanish Language. The age of the participants varied between 18 to 51 with a mean age of 21. Participants received course credit as a compensation for participating in our study. They were aware that the amount of compensation would not be affected by their performance or feedback in this study.

We conducted lab studies in an isolated room on campus, free from outside distractions. The studies were conducted with one participant at a time to allow the researchers to observe the users' interactions with the system, where the participants did not perceive any real risk. We used mock sites, each with distinct look-and-feel to distinguish between multiple schemes. In particular, we created seven realistic and distinct websites, including sites for banking, social networking, email, movie streaming, and shopping. The sites used the images and layouts from familiar commercial sites, and each of them was equipped with one of our seven graphical password schemes.

In our study, each of the five portfolios in a scheme consists of unique set of images that are not repeated in any other portfolio nor in any other scheme. In other words, we did not reuse any images. Users were shown the same set of images for a given portfolio in a scheme, where the passwords were randomly assigned by the system. We collected the images and phrases/facts (verbal cues) from free online resources.

4.2 Procedure

In password studies with multiple sessions, a one-week delay is a common interval (e.g., [2,4,5,16,34,49]), and Hayashi and Hong [25] showed that a one-week delay is larger than the maximum average interval for a user between subsequent logins to any of her important online accounts. So to test users' memorization of the assigned passwords in our study, each participant sat in two sessions, each lasting around 30 minutes, with the second session one week after the first one.

Session 1. After signing a consent form, the participants were given an overview of our study. Then they performed registration for each of the seven sites, each outfitted with a distinct scheme. The sites were shown to the participants in a random order during registration. After registering with each scheme, participants performed a practice login with that scheme. They performed another practice login with each scheme after completing registration for all of the seven sites. We did not collect data for these practice trials. They were asked to not record (e.g., write down or take a picture of) their authentication secrets.

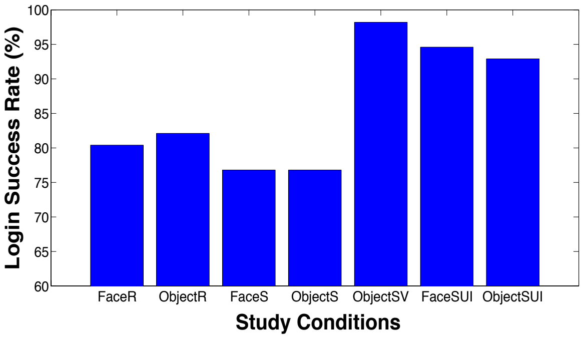


Figure 4: Login success rates for the study conditions [Number of participants=56]

Session 2. The participants returned one week after registration and logged into each of the seven sites using the assigned graphical passwords. The sites were shown to the participants in random order, and they could make a maximum of five attempts for a successful login. After the participants had finished, they were compensated and thanked for their time.

4.3 Ecological Validity

In our study, most of the participants were young and all of them were university educated. This participant pool may not generalize to the entire population. However, they are still representative of a large number of frequent Web users. They also came from diverse majors. As the study was performed in a lab setting, we were only able to gather data from 56 participants. However, lab studies have been preferred to examine brain-powered memorability of passwords [18]. Since lab studies take place in a controlled setting, it helps to establish performance bounds and figure out whether field tests are worthwhile in future research. We believe that 56 provides a suitable sample size for a lab study as compared to the prior studies on password memorability [2,4,5,11,12,44,48].

5. RESULTS

We now discuss the results of our user study. To analyze our results, we use statistical tests and consider results comparing two conditions to be significantly different when we find $p < 0.05$. When comparing two conditions where the variable is at least ordinal, we use a Wilcoxon signed-rank test for the matched pairs of subjects and a Wilcoxon-Mann-Whitney test for unpaired results. Wilcoxon tests are similar to t-tests, but make no assumption about the distributions of the compared samples, which is appropriate to the datasets in our conditions. Whether or not a participant successfully authenticated is a binary measure, and so we use either a McNemar's test (for matched pairs of subjects) or a chi-squared test (for unpaired results) to compare login success rates between two conditions. Here, we tested the following hypotheses:

Hypothesis 1

$H1_a$: *The login success rate for FaceS would be significantly higher than that for FaceR.*

$H1_b$: *The login success rate for ObjectS would be significantly higher than that for ObjectR.*

In a graphical password scheme offering spatial cues, the images in a portfolio remain the same and presented at a fixed position whenever that portfolio is loaded, which establishes a relationship between them and reinforces semantic priming (see §3 for details). Thus, we hypothesized that FaceS and ObjectS, offering spatial cues, would have significantly higher login success rates than FaceR and ObjectR, respectively, in which the position of images in a portfolio are randomly changed each time that a portfolio is loaded.

Our results show that out of 56 participants in our study, 45 participants (80%) succeeded to log in using FaceR, while 43 participants (77%) logged in successfully with FaceS. For ObjectR and ObjectS schemes, 46 participants (82%) and 43 participants (77%) succeeded to log in, respectively (see Figure 4). Thus, $H1_a$ and $H1_b$ are not supported by these results.

Whether or not a participant successfully authenticated is a binary measure, so we compare login success rates between conditions using McNemar's test. We did not find a significant difference in login success rate between FaceS and FaceR, $\mathcal{X}^2(1, N = 56) = 0.08$, $p = 0.77$, nor between ObjectS and ObjectR, $\mathcal{X}^2(1, N = 56) = 0.36$, $p = 0.55$.

Hypothesis 2

$H2_a$: *The login success rate for ObjectSV would be significantly higher than that for ObjectS.*

$H2_b$: *The login success rate for ObjectSV would be significantly higher than that for ObjectR.*

The ObjectSV scheme offers spatial and verbal cues (i.e., phrase or facts related to the images), where cues are shown both at registration and login. So, the users could memorize their graphical passwords through associating them with the corresponding cues, which should help to process and encode the information to store them in long-term memory (see §3 for detailed discussion). Moreover, the cues would assist users to recognize the images in the future, which should enhance their memorability. Thus, we hypothesized that ObjectSV scheme would have significantly higher login success rate than ObjectS and ObjectR schemes.

We observed a 98% login success rate for ObjectSV scheme, while 55 out of 56 participants could log in successfully one week after registration. As we compare the login success rate for ObjectSV scheme with that for ObjectS (77%) and ObjectR (82%), the results for McNemar's test show that ObjectSV had a significantly higher login success rate than ObjectS, $\mathcal{X}^2(1, N = 56) = 10.08$, $p < 0.05$ and ObjectR, $\mathcal{X}^2(1, N = 56) = 7.11$, $p < 0.05$. Hence, $H2_a$ and $H2_b$ are supported by these results.

Hypothesis 3

$H3_a$: *The login success rate for FaceSUI would be significantly higher than that for FaceS.*

$H3_b$: *The login success rate for FaceSUI would be significantly higher than that for FaceR.*

$H3_c$: *The login success rate for ObjectSUI would be significantly higher than that for ObjectS.*

$H3_d$: *The login success rate for ObjectSUI would be significantly higher than that for ObjectR.*

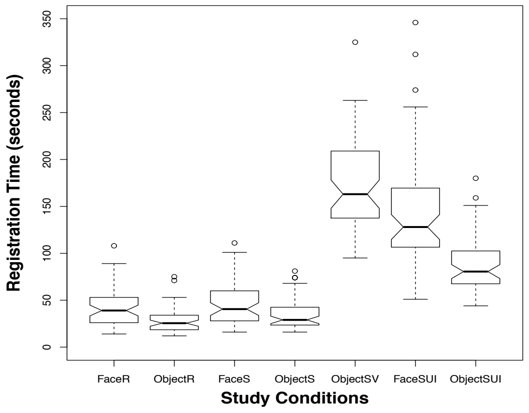


Figure 5: Registration time for the study conditions

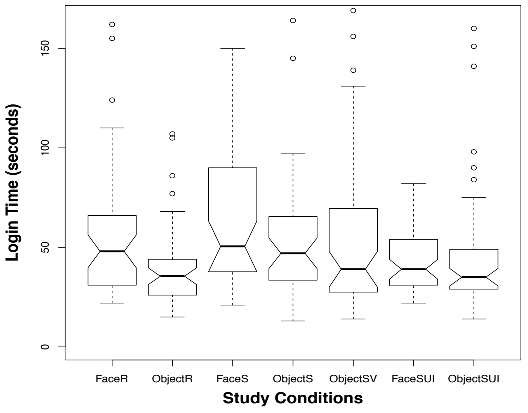


Figure 6: Login time for the study conditions

FaceR and ObjectR schemes represent existing graphical password schemes that use just the visual memory of users [1,26]. FaceSUI and ObjectSUI schemes leverage both visual memory and action-event memory [29], which contributes to an elaborative encoding of the assigned images and thus assists users with memorizing the processed information. In addition, FaceSUI and ObjectSUI schemes offer spatial cues. So, we hypothesized that the login success rate for FaceSUI would be significantly higher than that for FaceS and FaceR schemes, while ObjectSUI would have a significantly higher login success rate than ObjectS and ObjectR schemes.

Our results show that 53 participants (95%) in FaceSUI and 52 participants (93%) in ObjectSUI scheme logged in successfully one week after registration. As we compare the login success rate for FaceSUI with that for FaceS (77%) and FaceR (80%), the results for McNemar's tests show that FaceSUI had a significantly higher login success rate than FaceS, $\mathcal{X}^2(1, N = 56) = 8.1$, $p < 0.05$ and FaceR, $\mathcal{X}^2(1, N = 56) = 4.9$, $p < 0.05$.

We also found that the login success rate for ObjectSUI was significantly higher than that for ObjectS, $\mathcal{X}^2(1, N = 56) = 7.11$, $p < 0.05$ and ObjectR, $\mathcal{X}^2(1, N = 56) = 4.17$, $p < 0.05$. Thus, $H3_a$, $H3_b$, $H3_c$, and $H3_d$ are supported by these results.

5.1 Registration Time

We illustrate the results for registration time in Figure 5. We found that the median registration times for FaceR and FaceS were 39 seconds and 41 seconds, respectively, while FaceSUI scheme had a median registration time of 128 seconds. We use a Wilcoxon signed-rank test (appropriate for matched pairs of subjects) to evaluate two schemes in terms of registration time. The results show that the registration time for FaceR ($V = 1596$, $p < 0.05$) and FaceS ($V = 1596$, $p < 0.05$) were significantly less than that for FaceSUI scheme. We did not find a significant difference in registration time between FaceR and FaceS ($V = 789.5$, $p = 0.69$).

Our results show that the median registration time for ObjectSUI scheme was 81 seconds, while ObjectR and ObjectS schemes had median registration time of 26 seconds and 29 seconds, respectively. The results for Wilcoxon signed-rank tests show that the registration time for ObjectR ($V = 1595$, $p < 0.05$) and ObjectS ($V = 1582.5$, $p < 0.05$) were significantly less than that for ObjectSUI. We also found that the registration time for ObjectR was significantly less than that for ObjectS ($V = 1074.5$, $p < 0.05$). In our study, ObjectSV scheme had a median registration time of 163 seconds, while the registration time for ObjectR ($V = 1596$, $p < 0.05$), ObjectS ($V = 1596$, $p < 0.05$), and ObjectSUI ($V = 1566.5$, $p < 0.05$) were significantly less than that for ObjectSV scheme.

We intend to see if the image type had any impact on the required time for learning system-assigned images (i.e., registration time). Our results show that the registration time for ObjectR was significantly less than that for FaceR ($V = 147.5$, $p < 0.05$). We also found that the registration time for ObjectS was significantly less than that for FaceS scheme ($V = 249$, $p < 0.05$), while the registration time for ObjectSUI was significantly less than that for FaceSUI ($V = 39$, $p < 0.05$).

5.2 Login Time and Number of Attempts

In this paper, *number of attempts* and *login time* respectively refer to the required attempts and time for successful logins only, unless otherwise specified. We do not get matched pairs of subjects while comparing two schemes in terms of login time or number of attempts for successful logins, since some participants who logged in successfully for one scheme failed in the other scheme. So, we use a Wilcoxon-Mann-Whitney test (appropriate for unpaired results) to evaluate two schemes in terms of login time and the number of attempts for successful logins.

5.2.1 Login Time

We illustrate our results for login time in Figure 6. We found that the median login time for FaceR and FaceS were 48 and 51 seconds, respectively, while FaceSUI had a median login time of 39 seconds. The results for Wilcoxon-Mann-Whitney tests show that the login time for FaceSUI scheme was significantly less than that for FaceS ($W = 746.5$, $p < 0.05$). We did not find a significant difference in login time between FaceSUI and FaceR ($W = 1017$, $p = 0.21$), nor between FaceR and FaceS ($W = 1096$, $p = 0.20$).

Our results show that the median login time for ObjectSUI scheme was 35 seconds, while ObjectR and ObjectS schemes had median login times of 36 and 47 seconds, respectively.

The results for Wilcoxon-Mann-Whitney tests show that the login time for ObjectSUI ($W = 814$, $p < 0.05$) and ObjectR ($W = 1310.5$, $p < 0.05$) were significantly less than that for ObjectS. We did not find a significant difference in login time between ObjectSUI and ObjectR ($W = 1285$, $p = 0.53$).

We found a median login time of 39 seconds for ObjectSV. No significant difference was found in terms of login times when we compared ObjectSV with ObjectR ($W = 1489$, $p = 0.13$), ObjectS ($W = 1020.5$, $p = 0.25$), and ObjectSUI ($W = 1560.5$, $p = 0.42$).

We compared ObjectR and FaceR to see if image type had any impact on login time given random image positioning. The results for Wilcoxon-Mann-Whitney tests show that the login time for ObjectR was significantly less than that for FaceR ($W = 1312.5$, $p < 0.05$). However, we did not find a significant difference in login time between ObjectS and FaceS ($W = 1033$, $p = 0.26$), nor between ObjectSUI and FaceSUI ($W = 1498.5$, $p = 0.44$).

5.2.2 Number of Attempts

The mean number of attempts for a successful login was less than two for each of the seven schemes, while the median was one in each case (see Table 1). The results for Wilcoxon-Mann-Whitney tests found no significant difference between any pair of study conditions in terms of the number of attempts for a successful login.

6. DISCUSSION

Cognometric graphical passwords (e.g., Passfaces [1]) are now commercially available and deployed by a number of large websites, in which the images are assigned by the system to provide reasonable security guarantees. They fail, however, to gain satisfactory memorability [17], since it is difficult for most people to memorize system-assigned passwords. Our study explores a promising new direction to improve memorability for these passwords by leveraging humans' cognitive abilities through cues and interaction.

6.1 Cued-Recognition

We accommodate the scientific understanding of long-term memory to improve the memorability of system-assigned cognometric passwords. As noted by Atkinson and Shiffrin [7], any new information is transferred from short-term memory to long-term memory, when it is duly processed and encoded. In our study, we explored the impact of spatial and verbal cues for an elaborate encoding of authentication information to ease recognition during login.

Table 1: Number of Attempts for Successful Logins [SD: Standard Deviation]

Study Conditions	Mean	Median	SD
FaceR	1.3	1	0.7
ObjectR	1.2	1	0.5
FaceS	1.3	1	0.6
ObjectS	1.3	1	0.7
ObjectSV	1.4	1	0.9
FaceSUI	1.1	1	0.3
ObjectSUI	1.1	1	0.4

Al-Ameen et al. [5] show the potential of combining multiple cues to aid recognition, where the participants were asked to rate the efficacy of each type of cue. The participants rated spatial cues to be more effective than verbal cues to aid recognition. In our study, however, we made a deeper investigation of this issue through a direct comparison between schemes offering different combinations of cues, and we found that spatial cues did not significantly contribute to enhance memorability, while verbal cues made a significant contribution in this regard. Thus, the findings from our study make an important contribution to understand the effectiveness of memory cues (e.g., spatial and verbal cues) and indicate that relying solely on user feedback might not be a reliable approach to understand the impact of cues on password memorability.

6.1.1 Spatial Cues

To understand the efficacy of spatial cues, we compared FaceS and ObjectS schemes with fixed positions of all images in a portfolio with FaceR and ObjectR schemes with random repositioning of the images. Our results show that spatial cues did not contribute to improve the login success rate for either face recognition or object recognition.

In theory, spatial cues reinforce semantic priming and thus ease the recognition task [1]. Further, the survey results from Al-Ameen et al. [5] suggest that users found them important. It is possible that spatial cues are less effective when remembering multiple images, in which case it might create confusion when a user attempts to recognize the images using spatial cues. In our future work, we would perform a field study to explore if a higher login frequency could lead to training effects that could help users to benefit from spatial cues.

6.1.2 Verbal Cues

We compared ObjectSV, which has spatial and verbal cues, with both the object-based control condition (ObjectR) and ObjectS, which has spatial cues but not verbal ones. We found a 98% login success rate for ObjectSV, which was significantly higher than those for ObjectS and ObjectR. Given that we also found no benefit in spatial cues alone, we conclude that providing verbal cues with the images played a significant role in improving memorability.

During registration with ObjectSV, the participants may have learned the assigned images by correlating them with the verbal cues. This then assisted them with a more elaborate processing of the authentication information, but it also contributed to the higher registration time compared to ObjectR and ObjectS. No significant difference was found in terms of login time or number of attempts for successful logins between ObjectSV and either ObjectR or ObjectS.

We observed an interesting anecdotal case from one participant. In the first session, he told us that he used to struggle in memorizing new information because of a severe injury on his head. At this point, we were interested to see his login performances in the second session, and we found that he could log in successfully only with the schemes offering verbal cues (e.g., ObjectSV) and leveraging user interaction (e.g., FaceSUI). At the end of second session, he said, "It is much too difficult to manage with just images. The fact [verbal cue] attached with an image help to encode the images." Indeed, the benefit of verbal cues may not be important to all users, but may instead help users who struggle with graphical information alone, like the approximately 20% of participants who failed to login with ObjectS and ObjectR. If so, it may be good to individually tailor a scheme with different types of information for different users.

6.2 User Interaction

We tested user interaction with the FaceSUI and ObjectSUI schemes, in which we have users describe their assigned images at registration so as to deepen users' processing of authentication information. For clarity, we again note that the descriptions would be immediately destroyed and need not even be transferred to the server. In our study, we stored the user-written descriptions for the purpose of analysis. Our manual inspection shows that users made meaningful descriptions. If deployed, systems could use automated checks to partially enforce this.

We compare FaceSUI with FaceR and ObjectSUI with ObjectR to examine the memorability gain from combining spatial cues with user interaction, while comparisons of FaceSUI with FaceS and ObjectSUI with ObjectS reveal the more precise impact of user interaction on face recognition and object recognition, respectively. Our results show that the login success rate for FaceSUI (95%) was significantly higher than that for FaceS and FaceR and the login success rate for ObjectSUI (93%) was significantly higher than for ObjectS and ObjectR. It appears that user interaction played the major role to improve the success rates, since spatial cues by themselves did not help.

During registration with user interaction based schemes (e.g., FaceSUI and ObjectSUI), the participants wrote descriptions about the assigned images, which required significantly higher registration time for FaceSUI (in comparison to FaceS and FaceR) and ObjectSUI (in comparison to ObjectS and ObjectR).

6.3 Cued-Recognition vs. User Interaction

In our study, we found a significant improvement in login success rate through cued-recognition (ObjectSV) and user interaction (FaceSUI, ObjectSUI). The login success rate in ObjectSV was higher than the ObjectSUI and FaceSUI schemes, but no significant difference was found in this regard. We did not find a significant difference in login time or number of attempts for successful logins, when comparing ObjectSV with ObjectSUI and FaceSUI schemes. However, the registration times for ObjectSUI and FaceSUI were significantly less than that for ObjectSV scheme, indicating that the participants required less time to learn the assigned images through writing a description as compared to memorizing images through correlating them with the given verbal cues.

The deployment of ObjectSV scheme may require more effort as compared to other cognometric graphical password schemes that present users with images only, since ObjectSV requires writing verbal cues in addition to the images.

The success of the user-interaction-based schemes depends on the involvement of users in describing the assigned images. Our observations during the study reveal that all of the participants in ObjectSUI and FaceSUI schemes put in effort to describe the assigned images. Users in a lab study, however, are naturally open to take on requested tasks, while users in real life may get lazy. We plan to explore this issue deeper through a field study in a real-life setting and identify more ways for actively compelling users to engage with the interaction activity.

6.4 Face recognition vs. Object Recognition

Hlywa et al. [26] conducted a study to examine the effect of image type on the usability of recognition-based graphical passwords, in which they focused on exploring that impact for randomly-positioned images (similar to FaceR and ObjectR in our study).[2] In this paper, we provide a deeper understanding on this issue, while our investigation about the efficacy of cues and user interaction for face and object images lets us compare the face and object recognition in terms of registration time, login success rate, and login time for three different conditions: i) The images in a portfolio are randomly positioned each time that a portfolio is loaded (FaceR, ObjectR), ii) The images in a portfolio remain at the same position each time that a portfolio is loaded (FaceS, ObjectS), iii) The images in a portfolio are placed at the same position each time that a portfolio is loaded, and users learn the graphical passwords through interaction, e.g., writing a description about the assigned images at registration (FaceSUI, ObjectSUI).

The registration time for ObjectSUI scheme was found to be significantly less than that for FaceSUI scheme, which indicates that it was less time consuming for the participants to describe object images in comparison to face images. We also found that the registration time for ObjectR and ObjectS were significantly less than for FaceR and FaceS, respectively. Thus, users seem to need less time for object images. We speculate that since object images can be selected to be rather distinct from each other within a portfolio both visually and semantically (see Fig. 2), it takes less time to memorize a particular assigned object than for faces, which have distinct details but a basic similarity.

In login performance, we found no significant difference in login success rates as we compared ObjectR with FaceR, ObjectS with FaceS, and ObjectSUI with FaceSUI. ObjectR had a significantly lower login time than FaceR, but no significant difference was found in login time as we compared ObjectS with FaceS and ObjectSUI with FaceSUI. Random positioning may make visually distinctiveness more important to quick login times.

6.5 Input Type

We note that we used mouse input for our control conditions and keyboard input for the other conditions. As explained in Sec. 3.5, this was done to keep the control conditions the same as existing cognometric schemes while ensuring reasonable protection from shoulder surfing in the spatial-cue conditions. The input type, however, could affect memorability and login time. For example, the additional effort of selecting the key letter for typing may help with memorization at registration. On the other hand, using the mouse may provide greater focus on the visual elements, perhaps including cues. Further exploration of input options may be of interest if spatial cues are abandoned in favor of random image placement; in spatially fixed schemes, mouse input is likely too vulnerable to shoulder surfing for practical use.

[2] For randomly-positioned images, our findings about the impact of image type in terms of login time and login success rate are similar to those of Hlywa et al. [26]. Their study did not evaluate the effect of image type on registration time.

6.6 Future Work

Now that lab-study results show promise for implementing verbal cues and user interaction, it would be interesting to evaluate the approaches through a long-term field study with larger and more diverse populations, where we would explore the training effects on login performances over time. A recent field study [3] reveals that login time significantly decreases with the frequent use of a scheme due to training effects.

Although graphical passwords leverage the picture superiority effect, not all users may have a strong visual memory. Additionally, many graphical password schemes require good vision and motor skills, which elderly users [38] may lack. Thus, providing verbal cues for the images could assist users with memorizing their graphical passwords. We would further explore this issue in our future work through a user study with participants from different age groups. We would also make a deeper investigation to understand the impact of cues and user interaction in improving the memorability of passwords for the people with different cognitive limitations.

7. CONCLUSION

In our study, we aimed to better understand the impact of cues and user interaction on system-assigned recognition-based graphical passwords, and designed seven different study conditions to achieve this goal. In a study with 56 participants, we had a 98% login success rate for a scheme offering spatial and verbal cues (ObjectSV), while a scheme based on user interaction had a 95% login success rate for face images (FaceSUI) and a 93% login success rate for object images (ObjectSUI). Our analysis show that verbal cues and user interaction made an important contribution to gain significantly higher login success rate as compared to the control conditions representing existing graphical password schemes. Contrary to the suggestions of user feedback from a prior study [5], we found that spatial cues were not effective. These findings shed light on a promising research direction to leverage humans' cognitive ability through cues and interaction in gaining high memorability for system-assigned random passwords.

8. ACKNOWLEDGEMENT

This material is based upon work supported by the National Science Foundation under Grant No. CNS-1423163 and CAREER Grant No. CNS-0954133.

9. REFERENCES

[1] Passfaces corporation. The science behind Passfaces. White paper, `http://www.passfaces.com/enterprise/resources/white_papers.htm`.

[2] M. N. Al-Ameen, S. M. T. Haque, and M. Wright. Q-A: Towards the solution of usability-security tension in user authentication. Technical report, arXiv:1407.7277 [cs.HC], 2014.

[3] M. N. Al-Ameen and M. Wright. A comprehensive study of the GeoPass user authentication scheme. Technical report, arXiv:1408.2852 [cs.HC], 2014.

[4] M. N. Al-Ameen and M. Wright. Multiple-password interference in the geopass user authentication scheme. In *USEC*, 2015.

[5] M. N. Al-Ameen, M. Wright, and S. Scielzo. Towards making random passwords memorable: Leveraging users' cognitive ability through multiple cues. In *CHI*, 2015.

[6] J. R. Anderson and G. H. Bower. Recognition and recall processes in free recall. *Psychological Review*, 79(2), 1972.

[7] C. R. Atinkson and M. R. Shiffrin. Human memory: A proposed system and its control processes. *K.W. Spence and J.T. Spence (eds), Advances in the psychology of learning and motivation, New York academic press*, 1968.

[8] R. Biddle, S. Chiasson, and P. van Oorschot. Graphical passwords: Learning from the first twelve years. *ACM Computing Surveys*, 44(4), 2012.

[9] S. Chiasson, R. Biddle, and P. C. van Oorschot. A second look at the usability of click-based graphical passwords. In *SOUPS*, 2007.

[10] S. Chiasson, A. Forget, R. Biddle, and P. van Oorschot. User interface design affects security: Patterns in click-based graphical passwords. *International Journal of Information Security*, 8(6), 2009.

[11] S. Chiasson, E. Stobert, R. Biddle, and P. van Oorschot. Persuasive cued click-points: design, implementation, and evaluation of a knowledge- based authentication mechanism. *IEEE TDSC*, 9, 2012.

[12] S. Chiasson, P. C. van Oorschot, and R. Biddle. Graphical password authentication using cued click points. In *ESORICS*, 2007.

[13] A. Das, J. Bonneau, M. Caesar, N. Borisov, and X. Wangz. The tangled web of password reuse. In *NDSS*, 2014.

[14] D. Davis, F. Monrose, and M. Reiter. On user choice in graphical password schemes. In *USENIX Security*, 2004.

[15] A. E. Dirik, N. Memon, and J.-C. Birget. Modeling user choice in the passpoints graphical password scheme. In *SOUPS*, 2007.

[16] P. Dunphy and J. Yan. Do background images improve "Draw a Secret" graphical passwords? In *CCS*, 2007.

[17] K. Everitt, T. Bragin, J. Fogarty, and T. Kohno. A comprehensive study of frequency, interference, and training of multiple graphical passwords. In *CHI*, 2009.

[18] S. Fahl, M. Harbach, Y. Acar, and M. Smith. On the ecological validity of a password study. In *SOUPS*, 2013.

[19] D. Florencio and C. Herley. Where do security policies come from? In *SOUPS*, 2010.

[20] A. Forget. *A World with Many Authentication Schemes*. PhD thesis, Carleton University, 2012.

[21] A. Forget, S. Chiasson, P. van Oorschot, and R. Biddle. Improving text passwords through persuasion. In *SOUPS*, 2008.

[22] A. Forget, S. Chiasson, P. C. van Oorschot, and R. Biddle. Persuasion for stronger passwords: Motivation and pilot study. In *PT*, 2008.

[23] S. Furnell, I. Papadopoulos, and P. Dowland. A long-term trial of alternative user authentication technologies. *Information Management and Computer Security*, 12(2), 2004.

[24] J. V. Haxby, E. A. Hoffman, and M. I. Gobbini. The distributed human neural system for face perception. *Trends in Cognitive Science*, 4:223, 2000.

[25] E. Hayashi and J. I. Hong. A diary study of password usage in daily life. In *CHI*, 2011.

[26] M. Hlywa, R. Biddle, and A. S. Patrick. Facing the facts about image type in recognition-based graphical passwords. In *ACSAC*, 2011.

[27] I. Jermyn, A. Mayer, F. Monrose, M. Reiter, and A. Rubin. The design and analysis of graphical passwords. In *USENIX Security*, 1999.

[28] M. Just and D. Aspinall. Personal choice and challenge questions a security and usability assessment. In *SOUPS*, 2009.

[29] M. Knopf, A. Mack, S. Lenel, and S. Ferrante. Memory for action events: Findings in neurological patients. *Scandinavian Journal of Psychology*, 46, 2005.

[30] C. Kuo, S. Romanosky, and L. F. Cranor. Human selection of mnemonic phrase-based passwords. In *SOUPS*, 2006.

[31] D. A. Minnebusch, B. Suchan, O. Koster, and I. Daum. A bilateral occipitotemporal network mediates face perception. *Behavioural Brain Research*, 198 (1):179, 2009.

[32] D. Nali and J. Thorpe. Analyzing user choice in graphical passwords. Technical Report TR-04-01, School of Computer Science, Carleton University, 2004.

[33] D. L. Nelson, V. S. Reed, and C. L. McEvoy. Learning to order pictures and words: A model of sensory and semantic encoding. *Journal of Experimental Psychology: Human Learning and Memory*, 3(5), 1977.

[34] J. Nicholson, L. Coventry, and P. Briggs. Age-related performance issues for PIN and face-based authentication systems. In *CHI*, 2013.

[35] A. Paivio. *Mind and Its Evolution: A Dual Coding Theoretical Approach*. Lawrence Erlbaum: Mahwah, N.J., 2006.

[36] R. W. Proctor, M.-C. Lien, K.-P. L. Vu, E. E. Schultz, and G. Salvendy. Improving computer security for authentication of users: Influence of proactive password restrictions. *Behavior Research Methods, Instruments, and Computers*, 34(2), 2002.

[37] A. Rabkin. Personal knowledge questions for fallback authentication: Security questions in the era of Facebook. In *SOUPS*, 2008.

[38] K. Renaud. A visuo-biometric authentication mechanism for older users. In *British HCI*, 2005.

[39] S. Schechter, A. J. B. Brush, and S. Egelman. It's no secret: Measuring the security and reliability of authentication via 'secret' questions. In *IEEE S&P*, 2009.

[40] R. Shay, I. Ion, R. W. Reeder, and S. Consolvo. "my religious aunt asked why i was trying to sell her Viagra": Experiences with account hijacking. In *CHI*, 2014.

[41] R. Shay, P. G. Kelley, S. Komanduri, M. L. Mazurek, B. Ur, T. Vidas, L. Bauer, N. Christin, and L. F. Cranor. Correct horse battery staple: Exploring the usability of system-assigned passphrases. In *SOUPS*, 2012.

[42] R. Shay, S. Komanduri, P. G. Kelley, P. G. Leon, M. L. Mazurek, L. Bauer, N. Christin, and L. F. Cranor. Encountering stronger password requirements: User attitudes and behaviors. In *SOUPS*, 2010.

[43] F. Tari, A. Ozok, and S. Holden. A comparison of perceived and real shoulder-surfing risks between alphanumeric and graphical passwords. In *SOUPS*, 2006.

[44] J. Thorpe, B. MacRae, and A. Salehi-Abari. Usability and security evaluation of GeoPass: A geographic location-password scheme. In *SOUPS*, 2013.

[45] E. Tulving and D. M. Thompson. Encoding specificity and retrieval processes in episodic memory. *Psychological Review*, 80(5), 1973.

[46] E. Tulving and M. Watkins. Continuity between recall and recognition. *American Journal of Psych*, 86(4), 1973.

[47] W. A. Wickelgren and D. A. Norman. Strength models and serial position in short-term recognition memory. *Journal of Mathematical Psychology*, 3, 1966.

[48] S. Wiedenbeck, J. Waters, J. Birget, A. Brodskiy, and N. Memon. Authentication using graphical passwords: Effects of tolerance and image choice. In *SOUPS*, 2005.

[49] N. Wright, A. S. Patrick, and R. Biddle. Do you see your password? Applying recognition to textual passwords. In *SOUPS*, 2012.

[50] J. Yan, A. Blackwell, R. Anderson, and A. Grant. Password memorability and security: Empirical results. *IEEE Security and Privacy*, 2 (5):25, 2004.

On the Memorability of System-generated PINs: Can Chunking Help?

Jun Ho Huh*
Honeywell ACS Labs
Golden Valley, MN USA
junho.huh@honeywell.com

Hyoungschick Kim
Sungkyunkwan University
Suwon, Korea
hyoung@skku.edu

Rakesh B. Bobba*
Oregon State University
Corvallis, OR USA
Rakesh.Bobba@oregonstate.edu

Masooda N. Bashir
University of Illinois
Urbana-Champaign, IL USA
mnb@illinois.edu

Konstantin Beznosov
University of British Columbia
Vancouver, BC Canada
beznosov@ece.ubc.ca

ABSTRACT

To ensure that users do not choose weak personal identification numbers (PINs), many banks give out *system-generated random PINs*. 4-digit is the most commonly used PIN length, but 6-digit system-generated PINs are also becoming popular. The increased security we get from using system-generated PINs, however, comes at the cost of memorability. And while banks are increasingly adopting system-generated PINs, the impact on memorability of such PINs has not been studied.

We conducted a large-scale online user study with 9,114 participants to investigate the impact of increased PIN length on the memorability of PINs, and whether number *chunking*[1] techniques (breaking a single number into multiple smaller numbers) can be applied to improve memorability for larger PIN lengths. As one would expect, our study shows that system-generated 4-digit PINs outperform 6-, 7-, and 8-digit PINs in long-term memorability. Interestingly, however, we find that there is no statistically significant difference in memorability between 6-, 7-, and 8-digit PINs, indicating that 7-, and 8-digit PINs should also be considered when looking to increase PIN length to 6-digits from currently common length of 4-digits for improved security.

By grouping all 6-, 7-, and 8-digit chunked PINs together, and comparing them against a group of all non-chunked PINs, we find that chunking, overall, improves memorability of system-generated PINs. To our surprise, however, none of the individual chunking policies (e.g., 0000-00-00) showed statistically significant improvement over their peer non-chunked policies (e.g., 00000000), indicating that chunking may only have a limited impact. Interestingly, the top performing 8-digit chunking policy did show noticeable and statistically significant improvement in memorability over shorter 7-digit PINs, indicating that while chunking has the potential to improve memorability, more studies are needed to understand the contexts in which that potential can be realized.

Categories and Subject Descriptors

D.4.6 [**Security and Protection**]: Authentication; H.1.2 [**User/Machine Systems**]: Human factors

General Terms

Experimentation, Human Factors, Measurement, Security

Keywords

Security, Usability, PINs, Passwords, Policy, Chunking

1. INTRODUCTION

A personal identification number (PIN) is a numeric password that is used to authenticate users. PINs are commonly used in banking systems and on handheld devices (e.g., mobile phones and tablets) that require quick and easy yet sufficiently secure access. Many banks use 4-digit PINs to authenticate debit card (and sometimes credit card) transactions. Mobile phones often require users to enter 4-digit PINs to authenticate and unlock the screen.

To strengthen PIN security, some banks and others have recently started using 6-digit PINs, to take advantage of the larger PIN space of 10^6 possible entries. That could provide a significant improvement in security. However, if users generate their own 6-digit PINs, the improvement in entropy will be marginal as there tends to be a small pool of commonly selected 6-digit PINs [20]. Also, people find it harder to remember 6-digit PINs.

To get around the problem of low entropy in user generated PINs, many banks are adopting *system-generated PINs*, asking users to remember randomly generated 4- or 6-digit PINs. Banks in Switzerland, for example, assign 6-8 digit PINs; Canadian banks use both 4- and 6-digit PINs. It is

*Part of this work was done while Dr. Huh and Dr. Bobba were at the University of Illinois.

[1] Note that our notion of chunking differs from the traditional notion in that we do not chunk numbers into semantically meaningful pieces.

Symposium on Usable Privacy and Security (SOUPS) 2015, July 22–24, 2015, Ottawa, Canada.

common practice for banks to send physical mail to customers that contains a randomly generated PIN and instructions on how that PIN should be used and protected. System-generated PINs ensure that users do not use common PINs like "0000" and increase the entropy of PINs, making it more difficult for an attacker to guess them. Randomly generated PINs, when used together with an account lock-out policy (e.g., a user's account should be locked after 5 failed authentication attempts), can be highly effective against online brute-force attacks.

The biggest drawback with system-generated PINs (as with system-generated passwords [5]), however, is their memorability. Although many banks are moving to system-generated 6-digit PIN, the impact on memorability is not clearly known. Are the banks making the right decision in moving toward 6-digit PINs? Why should they not consider 7- or 8-digits?

We conducted a large-scale online user study[2], recruiting a total of 9,114 participants to understand the memorability of system-generated PINs of varying lengths, from 4 to 8 digits. As one would expect, our study shows that system-generated 4-digit PINs outperform 6-, 7-, and 8-digit PINs in long-term memorability. Interestingly, however, we find that there is no statistically significant difference between long-term memorability of 6-, 7-, and 8-digit PINs (see Section 4.4).

To investigate ways of improving memorability, we applied different "*chunking*" policies [14] on system-generated PINs, and studied their impact on memorability through the same online study. It is important to note that the notion of *chunking* used in this paper is different from the traditional notion of chunking. Traditional notion of *chunking* refers to the practice of breaking a single number into multiple smaller numbers that are *semantically meaningful*. Phone numbers are a good example of chunked numbers. In the United States a ten-digit phone number is chunked into smaller chunks of 3-3-4 (000–000–0000) that represent area code, exchange code and subscriber number respectively, to help people remember it easily. Based on what is already working well in the real world when using semantically meaningful chunks, we hypothesized that breaking longer system-generated PINs into smaller chunks, even if they did not have semantic meaning, would improve their long-term memorability.

We investigated a a variety of chunking combinations (referred to as *chunking policies*) to see how different arrangements of smaller chunks can affect memorability. In total, we investigated 12 different chunking policies. To the best of our knowledge, this is the first large-scale study on the impact of applying different variations of chunking techniques to randomly generated information and specifically to PINs; previous studies [9] often focused on showing that chunking is useful for information that has some meaning to a user.

A summary of the study findings is as follows:

- Our empirical evaluation of the relative memorability of 4-, 6-, 7-, and 8-digit system-generated PINs showed that 4-digit PINs outperform all other larger length PINs in long-term memorability as expected. Interestingly, however, the evaluation showed that there is no statistically significant difference between memorability of 6-, 7-, and 8-digit PINs.
- Our large-scale empirical evaluation of a wide variety of chunking policies for system-generated PINs did not find statistically significant improvement in long-term memorability when individual chunking policies (e.g., 0000-00-00) were compared against their non-chunked peers of same PIN length (e.g., 00000000).
- However, chunking policies did show statistically significant improvement in memoraiblity when grouped together and compared against a group of non-chunked polices.
- Further, we found an 8-digit chunking policy (0000-00-00) that noticeably outperformed shorter 7-digit PINs in long-term memorability with statistical significance.
- Differences in short-term memorability of system-generated PINs for different PIN lengths and chunking policies were found not to be statistically significant in most cases. Even in the cases where the differences were found to be statistically significant, the observed differences were relatively small ($< 3\%$).

Our findings suggest that, while chunking randomly generated PINs may not be universally effective in improving memorability, such chunking does have the potential to improve memorability in certain contexts. More studies are needed to understand the contexts in which chunking can help improve memorability of system generated PINs.

The next section discusses related work on PIN security and chunking techniques. Section 3 explains our hypotheses, methodology, and empirical study. Section 4 discusses the memorability results and participants' thoughts on the policies. Section 5 discusses our hypotheses in terms of the results. We discuss limitations, future directions and conclude in Section 6.

2. RELATED WORK

Over the years, many user authentication technologies have been designed and deployed on security-critical systems. Some of the popular technologies include passwords, PINs, digital certificates, physical tokens, one-time passwords, transaction profile scripts, and biometric identification. Among those, "what you know" forms of authentication, generally passwords or PINs, are still the dominant technology, mainly because of their familiarity and low implementation and deployment costs[16, 6]. However, to ensure good memorability, many users choose passwords or PINs that are easy to remember; such passwords/PINs can be guessed easily and are vulnerable to brute-force and dictionary attacks. Bonneau et al. [18] studied the difficulty of guessing human-selected 4-digit PINs, concluding that many users use their birth dates or other memorable dates as PINs, and an effective attack would involve brute-forcing PINs using dates.

To help users choose stronger PINs, PIN selection policies may be enforced. Such policies would specify, for instance, the combinations of numbers that can be used when a user is creating a PIN, or specify a list of PINs that cannot used. Even with those policies in place, however, users could still choose PINs that are weak but different from prohibited set of PINs, skewing the PIN distribution to the next set of popular, easy to remember PINs that are allowed by the policy.

[2] The study was approved by the Institutional Review Boards of both University of Illinois and Oregon State University.

Hence, the effectiveness and usability of those policies must be carefully evaluated before they are used. Kim et al. [20] analyzed the effectiveness of a few PIN selection policies, and found that a blacklist policy (e.g., forbidding the top 200 most popularly used PINs) can help users choose more secure 4- and 6-digit PINs that also have good memorability. Their study, however, looks at human-selected PINs.

As for passwords, there have also been active debates about the effectiveness of password selection policies. Several studies have already been carried out on understanding the relationship between password selection policies and the resulting passwords. Some of those studies were based on theoretical estimates [8, 26, 25]; some were based on small-scale laboratory studies [23, 7, 33, 13, 32]; and some were based on large-scale studies [29, 21]. Vu et al. [32] conducted a laboratory study that demonstrated that passwords chosen under strong selection policies are generally harder to compromise with automated password-cracking tools, but that they are also harder to generate and remember, affecting the overall usability. Kuo et al. [22] showed that automated tools were less effective against mnemonic passwords than against control passwords. Simulations performed by Shay et al. [26, 25] have shown that stringent-password selection policies can lead users to write down their passwords and thereby jeopardize their confidentiality. Shay et al. [29] examined users' behaviors and practices related to password creation under a new, more strict policy. Users were annoyed by the transition to a stricter password policy, but felt more secure under the new policy. Some users struggled to comply with the new policy, taking longer to create passwords and finding it harder to remember them. In a recent study, Komanduri et al. [21] tried to understand the relationship between password selection policies and the resulting passwords, and recommended some policies (e.g., a 16-character minimum with no additional requirements) that result in strong passwords without unduly burdening users. Shay et al. [27] particularly investigated the password selection policies for long passwords. Inglesant and Sasse [17] have shown that many users, despite knowing that repeated use of the same passwords is a bad security practice, rarely change their passwords.

Despite all those efforts in helping users select better passwords or PINs, advanced attackers have been effective in finding ways to crack them. Attackers combine dictionary attacks (with massive databases of dictionary words as well as commonly used passwords and PINs) with brute-forcing attacks, permutation attacks, rule-based attacks, or fingerprint attacks to crack just about anything that users create and remember [24, 15]. Hashcat [2] is a commonly used password-cracking tool that supports a combination of all of those attacks. With 4-digit PINs, it is even easier to perform those attacks, since the search space is much smaller. An alternative way to guarantee security is to use system-generated passwords or PINs, relying on a computer to generate a random password or PIN for you. System-generated PINs, although widely used by, for example, banks and the Department of Defense, have memorability issues [5]. To overcome such memorability weaknesses, we study the effects of using number-chunking techniques on system-generated PINs. In the past, several studies have shown the effectiveness of using chunking techniques as a memory tool for human brains. The hypothesis is based on the well-known process of chunking, in which primitive stimuli are grouped into larger conceptual groups such as letters into words [14]. Druzal et al. [11] claim that the use of chunking in working memory might be helpful for identification of faces. Thornton et al. [30] suggest that chunking is an effective mechanism for improving social working memory. Carstens et al. [9] show that human errors associated with password-based authentication can be significantly reduced through the use of passwords that are composed of data *meaningful* to the user.

Likewise, previous studies often looked at associating meaningful information with chunks (e.g., the first chunk of three digits in a phone number represents the area code). Our work extends those studies on chunking, but also incorporates other elements in that we apply number-chunking techniques to randomly generated PINs that are not associated with any meaningful information, evaluating the effects on both short-term and long-term memorability. To the best of our knowledge, we are the first to analyze the effects of chunking techniques specifically for PINs and to study so many different variations of chunking combinations (see Table 1) and PIN lengths. We are also the first to study the memorability of system-generated PINs, including the 4- and 6-digit PINs that many banks are currently using. Furthermore, previous studies have often been based on small-scale lab studies, with small numbers of participants; we have conducted a much larger-scale study using Mechanical Turk, recruiting a total of 9,114 participants (see Section 3.5).

3. METHODOLOGY

This section defines the key research questions and the hypotheses, provides an overview of the conducted user study, and explains the participant recruitment methodology.

3.1 Hypotheses

This work was motivated by research questions such as, how usable and memorable are system-generated 6-digit PINs compared to 4-digit PINs? Should banks also consider using 7- or 8-digit PINs? Can chunking techniques help improve the memorability of longer length system-generated PINs, and if so, how significant is the improvement?

Based on these research questions and our intuition, we defined the following three hypotheses.

1. The memorability of system-generated 6-digit PINs is worse than that of 4-digit PINs.
2. The memorability of system-generated 6-digit PINs is better than those of 7 and 8-digit PINs.
3. The memorability of longer (6-, 7- and 8-digit) system-generated PINs improves with chunking.

The user study and experiments were designed with the above hypotheses in mind. In Section 5, we discuss how the study results match up to these hypotheses.

3.2 PIN chunking policies

This section describes the 12 PIN chunking polices that we investigated (see Table 1), and explains why we chose these polices. Each chunking policy defines *the PIN length, how the numbers are chunked, and how the chunks are arranged.*

PIN lengths were defined first. Since banks (and other industrial entities) already use 4- and 6-digit PINs, we included them. Then, to test hypothesis 2 (see Section 3.1)

Table 1: PIN chunking policies. Each policy defines the PIN length, the number of digits each chunk contains, and the arrangement of chunks.

Policy	*Format*	*Example*
4	0000	8854
6	000000	480271
6:2-4	00–0000	48–0271
6:4-2	0000–00	4802–71
7	0000000	1685149
7:3-4	000–0000	168–5149
7:4-3	0000–000	1685–149
8	00000000	75357600
8:4-4	0000–0000	7535–7600
8:2-2-4	00–00–0000	75–35–7600
8:2-4-2	00–0000–00	75–3576–00
8:4-2-2	0000–00–00	7535–76–00

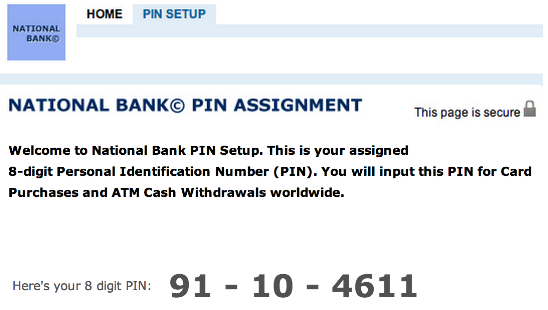
we included PIN lengths of 7 and 8 digits. The psychological literature on memorability suggests that most people can remember up to 4 digits without a problem over the short term [10]. Hence, we decided to include at least one chunk with 4 digits in all of the policies. We also decided that the smallest chunk in a policy should consist of at least two digits, because if we allow one-digit chunks, we could end up with policies with too many chunks in them. Based on those rules and intuition, PIN-chunking policies were designed. For instance, policy `8:2-2-4` says that a PIN would consist of 8 digits, where those 8 digits would be presented in three smaller chunks of 2, 2, and 4 digits, i.e., in the format of 00–00–0000.

3.3 User study design

Given the large number of PIN-chunking policies that we wanted to evaluate through empirical quantitative experiments, we chose to employ Amazon Mechanical Turk. To make the study as realistic as possible, we employed role-playing by simulating an online PIN setup page for a made-up bank, and informing each participant that he or she would use the PIN for card purchases (see Figure 1). Each participant was assigned a specific chunking policy (picked uniformly at random) and given a different system-generated PIN to remember. The given PIN was presented to the participant in the format defined by the chunking policy.

Our user study was designed following the dual memory model by Atkinson-Shiffrin [3] that postulates memories initially reside in a "short-term" memory for a limited time (20 to 30 seconds) while they are strengthening their association in the "long-term" memory. While later memory models (*e.g.*[4]) expand on the multi-store model proposed by Atkinson-Shiffrin they all agree that short-term memory (or working memory) has limited capacity and older items are wiped as new items enter. Further, rehearsing or recalling items while they are in the short-term memory causes the items to stay longer in the short-term memory while at the same time strengthening their association in the long-term memory.

Keeping the dual memory model in mind, our data collection involved two parts. The first part was meant to ensure that a PIN enters the long-term memory, and measure the short-term memorability[3] of that PIN. This part consisted of three steps. First, each participant was asked to complete three training tasks (rehearsing), to help them remember the assigned PIN (associate with long-term memory). Then, each participant was asked several questions related to cognition and memory strength, wiping out their short-term memory during the process. Third, each participant completed a short-term memorability test by entering the assigned PIN (see Figure 2). Two days (48 hours) after completing the first part, each participant that passed the short-term memorability test received an email inviting them to the second part of the study, in which we measured the long-term memorability. Two days was picked to study long-term memorability following the lead of other work in the community (e.g., [31, 19, 12, 28]). In the second part, each participant was simply asked to enter their PIN again.



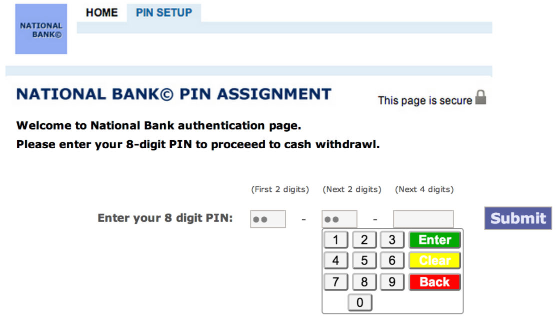
NATIONAL BANK© PIN ASSIGNMENT

This page is secure

Welcome to National Bank PIN Setup. This is your assigned 8-digit Personal Identification Number (PIN). You will input this PIN for Card Purchases and ATM Cash Withdrawals worldwide.

Here's your 8 digit PIN: **91 - 10 - 4611**

Figure 1: User study screenshot: Assigned PIN.



NATIONAL BANK© PIN ASSIGNMENT

This page is secure

Welcome to National Bank authentication page.
Please enter your 8-digit PIN to proceeed to cash withdrawl.

Enter your 8 digit PIN: ●● - ●● -

Figure 2: User study screenshot: PIN entry.

Details of the different tasks and the order in which the participants were asked to complete them are as follows:

Study Part 1: Short-term Memorability

1. **PIN assignment:** Each participant was given a system-generated PIN, which was presented in the format defined by the randomly assigned chunking policy (see Figure 1). Participants were asked to remember their PINs.

2. **Remembrance training:** Each participant was asked to enter the correct PIN three times to help with memorization. If incorrect PINs were entered three times

[3] The terms short-term memory and long-term memory should not be confused with short-term memorability and long-term memorability. Both short-memorability and long-term memorability refer to recalling from long-term memory, but over different lengths of time.

consecutively, the correct PIN was revealed again so that the participant would have another chance to memorize it. The training session ended only when the participant entered the correct PIN three times. Using a universal keypad (seen in Figure 2), participants entered their PIN in the chunking format defined by the chunking policy picked at random for them.

3. **Cognition and memory strength questions:** Each participant was asked 18 cognition questions (e.g., "I prefer complex to simple problems") and 6 memory strength questions (see Table 2). Answering all of those questions was expected to take about 2 minutes. Those questions were asked to help us better understand the participants' characteristics, and to clear participants' short-term memory.

4. **Demographics and survey questions:** Each participant was asked five demographic questions and five survey questions (see Table 3) about the assigned PIN. The survey questions asked about their thoughts on the memorability and usability of the assigned PIN and the chunking policy, taking about a minute to complete.

5. **Enter PIN (short-term test):** Each participant was asked to enter the assigned PIN (see Figure 2) and was given three chances to enter it correctly. To simulate an existing banking scenario, we asked participants to enter the correct PIN to proceed with cash withdrawal.

Study Part 2: Long-term Memorability

6. **Enter PIN after two days (long-term test):** Two days (48 hours) after participating in the first part of the study, participants who passed the short-term memorability study received an email asking them to complete the second part of the study. The participants were asked to enter their PIN again and were given three chances to enter the correct one.

7. **PIN survey questions:** To understand how participants feel about long-term usability of the assigned PIN, the same five survey questions about the assigned PIN (Table 3) were asked again after completing the test.

Table 2: Memory strength questions

#	*Question*
MQ1	I have a difficult time remembering numerical information.
MQ2	I frequently get passwords and numbers mixed up in my head.
MQ3	I have a good memory for things I have done in the past week.
MQ4	I easily lose my train of thought.
MQ5	I have a good memory for phone numbers that I have dialed in the past.
MQ6	I frequently remember details of past events that other people have forgotten.

Table 3: PIN survey questions

#	*Question*
SQ1	How difficult was it for you to remember the assigned PIN?
SQ2	Did you use an external storage (e.g., a sheet of paper or a text file) to write down the assigned PIN?
SQ3	Did you use any special technique (e.g., keypad patterns, assigning images to numbers, converting numbers to words) to help you remember the assigned PIN?
SQ4	If you answered "Yes" to Q3, what was the special technique that you used?
SQ5	Do you currently use a PIN that is equal to or longer than 6 digits?

To minimize the chances that the participants would write down their PINs after the first part of the study, we did not disclose exactly what the participants would have to do in the second part and how they would be rewarded. We simply informed the participants that they might be invited to the second part of the study in two days. However, we informed those who returned to complete the second part that they could earn an extra bonus by getting the PIN right, providing incentives for them to try their best to recall the correct PIN.

3.4 User data collected

Throughout the 7 different stages of the user study (see above), we recorded the following information:

- **Assigned PIN and chunking policy.** We recorded the chunking policy and the PIN each participant was assigned.
- **Number of attempts made in entering PIN.** We recorded the number of attempts a participant made to enter the correct PIN in all of the training sessions and short-term and long-term memory tests.
- **Time taken to enter PIN.** Likewise, we measured the time it took each participant to enter a PIN for every attempt made, in all of the training sessions and short-term and long-term memory tests. Timing began when the participant first accessd the login screen and ended when the participant either entered the correct password or tried and failed all three attempts, capturing both successful and unsuccessful login attempts.
- **Memorability results.** We recorded the results of the memorability tests (i.e., whether a correct PIN was entered) for every attempt made, in all of the training sessions and short-term and long-term memory tests.
- **Survey answers.** We recorded participants' answers to the cognition questions, memory strength questions, and PIN survey questions (see Tables 2 and 3).

3.5 Mechanical Turk

Given the large number of PIN-chunking policies that we wanted to evaluate through empirical quantitative experiments, we chose to employ Amazon Mechanical Turk [1]. Every participant who completed the first part was rewarded with $0.50. Those who came back for the second part were rewarded with an additional $0.25 and another $0.25 if they entered their PINs correctly. The intention of this extra bonus was to provide an incentive for participants to try their best to recall the correct PIN. As can be seen from the high short-term memorability scores in Table 5, participants did not need an extra monetary incentive to recall the PIN in the first part.

3.6 Statistical tests

We first performed the chi-square test on the proportion of successful logins and external storage usage to check whether proportions across all chunking policies are equal or not ($p < 0.05$). If chi-square test results indicated that not all proportions are equal, we performed Fisher's exact test (FET) to check whether a proportion in one chunking policy is significantly greater than that of another chunking policy ($p < 0.05$). All comparisons were corrected for multiple-testing using False Discovery Rate (FDR) estimation when appropriate.

As for authentication time, the Shapiro-Wilk's test was first used to show that the collected data is not normally distributed. To check whether all the policies have equal medians for authentication time, we performed the Kruskal-Wallis test ($p < 0.05$), showing that not all medians are equal. We then used the unpaired Mann-Whitney (MW) U test ($p < 0.05$) to measure the statistical confidence in the authentication time differences between chunking policies. All comparisons were corrected for multiple-testing using False Discovery Rate (FDR) estimation when appropriate.

Table 4: The demographics of the participants

Gender	
Male	5,180 (56.84%)
Female	3,855 (42.30%)
No answer	79 (0.86%)
Age group	
18–29	5,256 (57.67%)
30–39	2,285 (25.07%)
40–49	842 (9.24%)
50–59	488 (5.35%)
60 and over	168 (1.84%)
No answer	75 (0.83%)
Education	
Less than high school	63 (0.69%)
High school	2,825 (31%)
University	5,066 (55.58%)
Masters	768 (8.43%)
Doctoral	104 (1.14%)
Professional	177 (1.94%)
No answer	111 (1.22%)
Ethnicity	
African American	552 (6.06%)
Asian	769 (8.44%)
Caucasian	6,936 (76.10%)
Hispanic	503 (5.52%)
Other	198 (2.17%)
No answer	156 (1.71%)

4. RESULTS

This section presents the key results obtained from the user study, including the memorability results and the participants' responses regarding the difficulty in remembering their assigned PINs.

4.1 Demographics

As mentioned in Section 3.5, participants were recruited using Mechanical Turk. During the study period, a total of 9,114 participants completed the first part of the study, and of those 6,208 participants came back to complete the second part. A majority of the participants were Caucasian (76.10%), and more than half were in the age group of 18–29 (57.67%). 56.84% were male and 55.58% had a university degree. The details of the demographics are presented in Table 4.

4.2 Writing down PINs

In response to our survey question (see SQ2 in Table 3) participants reported using external storage (i.e., having their PIN written down) to store their PINs. The percentage of participants who reported using external storage to remember their PIN for short-term and long-term memorability tests is shown in column five of Tables 5 and 7 respectively.

The number of participants using some form of external storage during short-term memorability test steadily increased with the size of the PIN. In particular, in our study 5% of participants who were assigned a 4-digit PIN reported using external storage. For 6-, 7-, 8-digit PINs the number of users that reported using external storage ranged from $7-8\%$, $10-11\%$ and $11-14\%$ respectively. The chi-square test results showed that not all external storage usage proportions are equal ($\chi^2(11) = 65.34$, $p < 0.0001$). Hence, we used FET to identify differences across the policies that are statistically significant.

The observed increase in the percentage of users using external storage when compared to those with 4-digit PINs was found to be statistically significant for all policies with larger PINs ($p < 0.05$, pairwise corrected FET) except for `6` and `6:4-2`. The observed increase in the percentage of users using external storage when compared to those with 6-digit PINs (both chunked and non-chunked) was found to be statistically significant for all 8-digit PIN policies ($p < 0.05$, pairwise corrected FET) except for policy `6:2-4` vs. `8`, `6:4-2` vs. `8`, and `6:2-4` vs. `8:4-2-2`. Comparing 6-digit PINs with 7-digit PINs, the increase in external storage was found to be significant only for the case of policy `6` vs. `7:3-4` ($p < 0.05$, pairwise corrected FET).

Similar to what was observed during short-term memorability test, the number of participants who reported using external storage to remember the PIN increased with size of the PIN length in long-term memorability test, except for a slight dip when going from 4-digit to 6-digit (see column 5 in Table 7). Again, the chi-square test results showed that not all external storage usage proportions are equal ($\chi^2(11)$ $= 33.18$, $p < 0.0005$).

The observed increase in the percentage of users using external storage when compared to those with 4-digit PINs was found to be statistically significant for all policies with 8-digit PINs (all $p < 0.05$, pairwise corrected FET) except for policy `8:4-2-2`. The observed increase in the percentage of users using external storage when compared to policy `6` was found to be statistically significant for all 7- and 8-digit PIN policies (all $p < 0.05$, pairwise corrected FET).

These observations are consistent with previous findings in literature that users tend to write down passwords when

Table 5: Short term memorability and average time taken to authenticate. Column '% correct PIN' represents the percentage of participants who entered the correct PIN in the short-term test *not counting* those who reported to have their PIN written down on paper or electronically to remember it (see 3.3). Column '% Ext. storage' represents the percentage of participants who reported using external storage. Column 'Time' is the median time taken to authenticate (considering both successful and unsuccessful results) and is measured in seconds. Column 'No. of attempts' is the average number of attempts for a successful login.

Policy	*# Participants*	*# Failed*	*% correct PIN*	*% Ext. storage*	*Time (s)*	*# attempt*	*σ*
4	722	5	99%	5%	9.1	1.0	0.2
6	714	19	97%	7%	11.1	1.1	0.3
6:2-4	717	12	98%	8%	12.9	1.1	0.3
6:4-2	709	14	98%	8%	12.8	1.1	0.3
7	678	25	96%	11%	13.3	1.1	0.4
7:3-4	658	11	98%	11%	13.8	1.1	0.3
7:4-3	669	19	97%	10%	14.1	1.1	0.4
8	682	27	96%	11%	14.7	1.2	0.5
8:4-4	670	16	98%	13%	15.5	1.1	0.4
8:2-2-4	644	16	98%	14%	16.6	1.1	0.4
8:2-4-2	647	14	98%	14%	17.4	1.1	0.4
8:4-2-2	667	12	98%	11%	16.4	1.1	0.4

they are required to remember what they perceive as complex passwords (e.g., [26, 25]. Since this paper focuses on the memorability (and not security) of system-generated PINs, we did *not* include participants who reported to have their PIN written down (on paper or electronically) in all of the following analyses.

4.3 Memorability of individual policies

We first present the short-term and long-term memorability results for individual chunking policies, comparing memorability of each policy against all other policies.

4.3.1 Short term

As shown in Table 5, in our study all of the PIN policies scored high in short-term memorability, ranging between 96% for non-chunked 8-digit PIN to 99% for 4-digit PIN. In our sample, as shown in Figure 3, chunked PINs had the same or slightly better memorability score (i.e., showed higher percentage) than non-chunked PINs of the same PIN length. The chi-square test results showed that not all short-term memorability scores are equal ($\chi^2(11) = 25.91$, $p < 0.01$). However, only the differences in short-term memorability between 4-digit and 7-digit (99% vs. 97%) and between 4-digit and 8-digit (99% vs. 96%) were found to be statistically significant (all $p < 0.005$, pairwise corrected FET). A summary of all of the *statistically significant* short-term memorability differences is presented in Table 6.

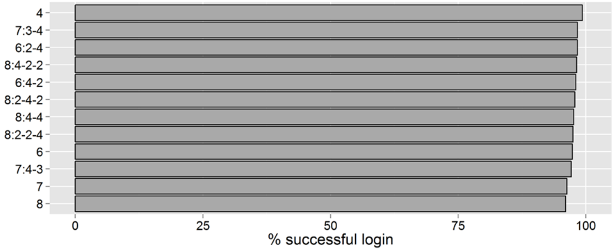


Figure 3: Sorted short-term memorability scores

Table 6: Statistically significant short-term memorability rate comparisons. 'p-value' is generated from pairwise corrected FET.

Superior policy/group	**Inferior policy/group**	**p-value**
Individual policy comparisons		
4 (99%)	7 (97%)	< 0.005
4 (99%)	8 (96%)	< 0.005
Policy group comparisons		
Chunk (98%)	No-Chunk (97%)	< 0.005
4 (99%)	7-Chunk (98%)	< 0.05
4 (99%)	8-Chunk (98%)	< 0.05
6-Chunk (98%)	7 (96%)	< 0.05
6-Chunk (98%)	8 (96%)	< 0.05
8-Chunk (98%)	8 (96%)	< 0.05
4-All (99%)	6-All (98%)	< 0.05
4-All (99%)	7-All (97%)	< 0.005
4-All (99%)	8-All (97%)	< 0.005

4.3.2 Long term

In contrast to the results for short-term memorability, we observed a significant decrease (up to 28%) in long-term memorability as we moved from 4-digit PINs to larger length PINs (see Table 7). Specifically, recall success rate for 4-digit PINs was at 74% while the rate for larger PIN lengths (including chunked PINs) varied from 46% to 57%. As expected, the chi-square test results showed that not all long-term memorability scores are equal across all of the policies ($\chi^2(11) = 79.31$, $p < 0.0001$).

The observed decrease in long-term memorability of larger length PINs when compared with that of 4-digit PINs was found to be statistically significant for all policies (chunked and non-chunked) with larger PINs ($p < 0.0001$, pairwise corrected FET). As expected, 4-digit PINs significantly outperformed larger length PINs in terms of memorability. Interestingly, however, the observed differences in memorability between non-chunked 6 (55%), 7(45%), and 8(50%) PINs were not found to be statistically significant. A summary of all of the *statistically significant* long-term memorability differences is presented in Table 8.

As seen in Figure 4, the recall success rates for all chunked

Table 7: Long term memorability and median time taken to authenticate.Column '% correct PIN' represents the percentage of participants who entered the correct PIN in the short-term test *not counting* those who reported to have their PIN written down on paper or electronically to remember it (see 3.3). Column '% Ext. storage' represents the percentage of participants who reported using external storage. Column 'Time' is the median time taken to authenticate (considering both successful and unsuccessful results) and is measured in seconds. Column 'No. of attempts' is the average number of attempts for a successful login.

Policy	*# Partici-pants*	*# Failed*	*% correct PIN*	*% Ext. storage*	*Time (s)*	*# at-tempt*	*σ*
4	517	137	74%	6%	22.6	1.7	1.0
6	506	228	55%	5%	35.5	2.0	1.0
6:2-4	494	212	57%	8%	41.7	1.9	1.0
6:4-2	502	217	57%	8%	40.7	1.9	1.0
7	462	248	46%	10%	47.1	2.1	1.0
7:3-4	461	233	49%	10%	46.8	2.1	1.1
7:4-3	461	215	53%	10%	44.2	2.0	1.0
8	454	227	50%	11%	49.6	2.1	1.1
8:4-4	456	218	52%	10%	51.7	2.0	1.0
8:2-2-4	440	201	54%	12%	50.4	2.0	1.0
8:2-4-2	436	203	53%	12%	53.0	2.0	1.0
8:4-2-2	456	203	55%	10%	48.8	2.0	1.0

Table 8: Statistically significant long-term memorability rate comparisons.

Superior policy/group	Inferior policy/group	p-value
Individual policy comparisons		
4 (74%)	6 (55%)	< 0.0001
4 (74%)	6:2-4 (57%)	< 0.0001
4 (74%)	6:4-2 (57%)	< 0.0001
4 (74%)	7 (46%)	< 0.0001
4 (74%)	7:3-4 (49%)	< 0.0001
4 (74%)	7:4-3 (53%)	< 0.0001
4 (74%)	8 (50%)	< 0.0001
4 (74%)	8:4-4 (52%)	< 0.0001
4 (74%)	8:2-2-4 (54%)	< 0.0001
4 (74%)	8:2-4-2 (53%)	< 0.0001
4 (74%)	8:4-2-2 (55%)	< 0.0001
6:2-4 (57%)	7 (46%)	< 0.01
6:4-2 (57%)	7 (46%)	< 0.01
8:4-2-2 (55%)	7 (46%)	< 0.05
Policy group comparisons		
Chunk (54%)	No-Chunk (51%)	< 0.05
4 (74%)	6-Chunk (57%)	< 0.05
4 (74%)	7-Chunk (51%)	< 0.0001
4 (74%)	8-Chunk (54%)	< 0.0001
6-Chunk (57%)	7 (46%)	< 0.001
6-Chunk (57%)	7-Chunk (51%)	< 0.05
8-Chunk (54%)	7 (46%)	< 0.05
4-All (74%)	6-All (56%)	< 0.0001
4-All (74%)	7-All (50%)	< 0.0001
4-All (74%)	8-All (53%)	< 0.001
6-All (56%)	7-All (50%)	< 0.001

PINs were observed to be better than that of their corresponding non-chunked PINs. Surprisingly, none of the observed increases in memorability, when using chunked PINs, compared to their non-chunked peer policies, were found to be statistically significant. Interestingly, however, chunked 6-digit PIN policies (6:2-4 and 6:4-2) both outperformed (by 11 %) non-chunked 7-digit PINs when there was no statistically significant difference found in long-term memorability of non-chunked 6- and 7-digit PINs ($p < 0.01$, pairwise corrected FET). Further, policy 8:4-2-2 outperformed policy 7 that is shorter in length (55% vs. 46%) with statistical significance ($p < 0.05$, pairwise corrected FET).

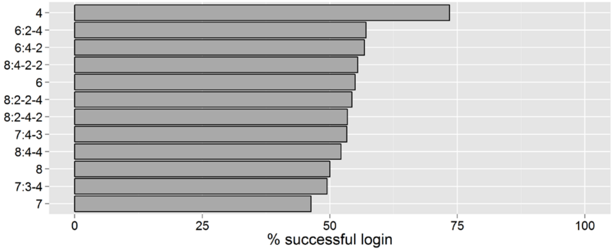


Figure 4: Sorted long-term memorability scores

4.4 Memorability of chunking and non-chunking policy groups

Although we did not find any statistically significant improvement in memorability when chunking policies were compared against their non-chunked peers, Figures 3 and 4 do indicate that chunking policies might potentially outperform their non-chunked peers. To further probe effectiveness of chunking techniques, we bundled the chunking policies together and compared them as a group against the group of non-chunking policies. This section presents two such analyses.

4.4.1 Chunking policy group vs. non-chunking policy group

First, we divided the policies into a non-chunking group (No-Chunk) that consists of policies 6 and 7, and a chunking group (Chunk) that consists of policies 6:2-4, 6:4-2, 7:3-4, and 7:4-3. We excluded 4- and 8-digit PIN policies from this analysis because there is no 4-digit chunking policy, and there are twice as many 8-digit chunking policies as there are for 6-digit and 70digit which could skew the results.

As shown in Tables 9 and 10, the differences in the memorability scores between the two groups were about 1% in the short-term and about 4% in the long-term, which were statistically significant differences (all $p < 0.05$). Even though

Table 9: Short term memorability of the chunking policy group (Chunk) and the non-chunking policy group (No-chunk) for 6- and 7-digit PINs.

Policy	*# Participants*	*# Failed*	*% correct PIN*
No-chunk	2,074	71	97%
Chunk	5,381	114	98%

Table 10: Long term memorability of the chunking policy group (Chunk) and non-chunking policy group (No-chunk) for 6- and 7-digit PINs.

Policy	*# Participants*	*# Failed*	*% correct PIN*
No-Chunk	1,422	703	51%
Chunk	3,706	1,702	54%

none of the 6- and 7-digit chunking policies individually showed statistically meaningful improvement when compared to their non-chunked peer, when grouped together, they showed statistically significant difference against the non-chunked 6 and 7 policies.

4.4.2 Grouping chunking policies with the PIN length

Table 11: Short term memorability of chunking policies grouped by the PIN length.

Policy	*# Participants*	*# Failed*	*% correct PIN*
4	722	5	99%
6	714	19	97%
6-Chunk	1,426	26	98%
7	678	25	96%
7-Chunk	1,327	30	98%
8	682	27	96%
8-Chunk	2,628	58	98%

Table 12: Long term memorability of chunking policies grouped by the PIN length.

Policy	*# Participants*	*# Failed*	*% correct PIN*
4	517	137	74%
6	506	228	55%
6-Chunk	996	429	57%
7	462	248	46%
7-Chunk	922	448	51%
8	454	227	50%
8-Chunk	1,788	825	54%

In the second analysis, we grouped just the chunking policies together by their PIN length (i.e., three chunking groups of length 6 as 6-Chunk, 7 as 7-Chunk, and 8 as 8-Chunk) and compared them against their non-chunked peers as well as other chunking policy groups.

Table 11 shows the short-term memorability of those chunking policy groups. As shown in Table 6, only 8-Chunk showed statistically significant 2% improvement over its non-chunked peer policy 8 ($p < 0.05$, pairwise corrected FET). Group 6-Chunk showed statistically significant superiority over both 7 and 8 (all $p < 0.05$, pairwise corrected FET).

Long-term memorability of those chunking policy groups are shown in Table 12. In contrast to the short-term results, even policy 8-Chunk failed to show statistically significant improvement over 8. Policy 6-Chunk failed to show statistically significant superiority over 8 in the long-term, but still showed statistically significant improvement over 7 ($p < 0.001$, pairwise corrected FET). Group 6-Chunk also showed significant superiority over 7-Chunk in the long-term ($p < 0.05$, pairwise corrected FET). Similarly, policy 8-Chunk showed statistically significant superiority over 7 ($p < 0.05$, pairwise corrected FET). This can be explained by the best-performing individual 8-digit chunking policy 8:4-2-2 that outperformed policy 7 (see Table 8).

4.5 Memorability of PIN length groups

To further analyze the memorability differences between PINs of different lengths, we grouped all the policies of the same PIN length together, creating four groups of 4-All, 6-All, 7-All, and 8-All.

Table 13 shows the short-term results. As expected, group 4-All outperformed all other groups with statistical significance (all $p < 0.05$, pairwise corrected FET). Statistically significant differences are captured in Tables 6 and 8. Long-term results, presented in Table 14, were more interesting. Group 6-All at 56% showed statistically significant superiority over group 7-All (all $p < 0.001$, pairwise corrected FET), which showed the lowest memorability score of 50%. Groups 8-All (53%) and 7-All (50%), however, did not show statistically significant difference against each other. As expected, group 4-All outperformed all other groups again in the long-term with statistical significance (all $p < 0.001$, pairwise corrected FET).

Table 13: Short term memorability of four PIN length groups.

Policy	*# Participants*	*# Failed*	*% correct PIN*
4-All	761	5	99%
6-All	2,323	45	98%
7-All	2,245	55	97%
8-All	3,785	85	97%

Table 14: Long term memorability of four PIN length groups.

Policy	*# Participants*	*# Failed*	*% correct PIN*
4-All	548	137	74%
6-All	1,611	657	56%
7-All	1,534	696	50%
8-All	2,515	1052	53%

4.6 Time taken to authenticate

Tables 5 and 7 show median time taken to authenticate for the short-term test and the long-term test, respectively. Both successful and unsuccessful authentications were considered. Kruskal-Wallis test results showed that not all medians across the policies are equal, respectively, for short-term memorability ($\chi^2(11) = 1204.72$, $p < 0.0001$) and for long-term memorability ($\chi^2(11) = 360.79$, $p < 0.0001$). The Mann-Whitney U test was then used (since the time data was not normally distributed) to identify statistically significant differences in authentication times.

In the short-term memorability tests, policy 4 was the clear winner, outperforming all other policies with a median of 9.1 seconds authentication time (all $p < 0.01$, pairwise corrected MW U test). Policy `6`, at 11.1 seconds, was the next best policy, and outperformed all other policies that required 6 or more digits to be entered (all $p < 0.01$, pairwise corrected MW U test). Similarly, policies `6:2-4` and `6:4-2` outperformed all 8-digit policies (all $p < 0.01$, pairwise corrected MW U test). Those Policies (`6:2-4` and `6:4-2`) also outperformed policies `7:3-4` and `7:4-3`, respectively (all $p < 0.01$, pairwise corrected MW U test). All 7-digit policies outperformed all chunked 8-digit policies with statistical significance ($p < 0.05$, pairwise corrected MW U test).

In the long-term memorability tests, policy `4` was the winner again with respect to authentication time (see Table 7). The observed median authentication time for policy `4` was 22.6 seconds, and its advantage over other policies was found to be statistically significant ($p < 0.01$, pairwise corrected MW U test). Interestingly, policy `6` did not fare as well as it did during short-term memorability tests. We recorded a median authentication time of 35.5 seconds, a significant increase relative to policy `4`. However, policy `6` was still lower than for all other policies we tested except for `6:4-2` (all $p < 0.05$, pairwise corrected MW U test).

4.7 Number of authentication attempts

Tables 5 and 7 also show the average numbers of authentication attempts made in the short-term and long-term tests, respectively. In the short-term test, the average number of attempts was around 1.1, with a standard deviation around 0.2-0.5. There was very little difference in the average values among all the policies, indicating that most participants entered the correct PIN on their first attempt.

In the long-term test, the average value rises to around 1.9 to 2.1, except for policy `4`, which averaged 1.7 attempts. This shows that the majority of participants made about two attempts in the long-term test. This explains the significant increase in average authentication time—we measured both successful and unsuccessful attempts—observed during long-term memorability test relative to the short-term memorability test.

4.8 User perception of recall difficulty

We compiled participants' responses to `SQ1` in Table 3, to gauge user perception of "recall difficulty" of PINs across different policies. The results for user perception of short-term and long-term recall difficulty are shown in Figures 5 and 6, respectively.

Not surprisingly, in the short-term test, the shorter the PIN length, the greater was the percentage of participants who felt that the PIN is easy to remember. The chi-square test results showed that not all recall difficulty proportions are equal ($\chi^2(44) = 604.60$, $p < 0.0001$). 90% of the participants felt 4-digit PINs are easy to remember, compared to only 61% who felt the same way about 8-digit PINs ($p < 0.01$, pairwise corrected FET). Policy `7:4-3` outperformed its non-chunked policy `7` with statistical significance ($p < 0.05$, pairwise corrected FET).

Similar trends in user perception were observed for the long-term test with one exception. In the long-term test, the chi-square test results also showed that not all recall difficulty proportions are equal ($\chi^2(44) = 257.53$, $p < 0.0001$). Specifically, the shorter the PIN length, the greater was the percentage of participants who felt that the PIN is easy to remember. However, all chunking policies with 8-digits except `8:4-4` did better than 7-digit PINs (although not found to be statistically significant) in the sample. In contrast to the short-term results, however, none of the chunked policies showed statistically significant improvement over their peer non-chunked policies.

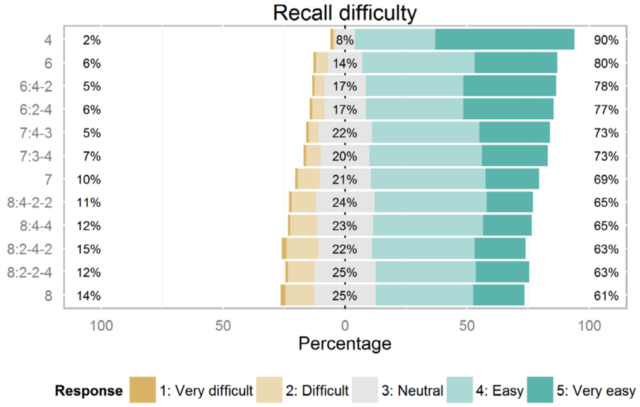


Figure 5: Results for short-term recall difficulty

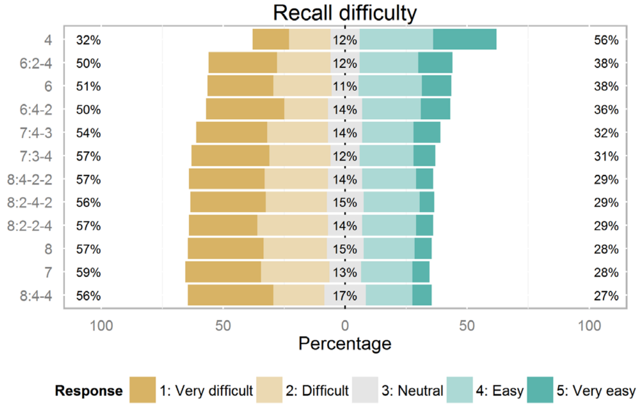


Figure 6: Results for long-term recall difficulty

4.9 Special remembrance techniques

Questions `SQ3` and `SQ4` from Table 3 asked the participants about any special techniques that they used to help them remember their PINs. Tables 15 and 16 show the relationship between the reported use of special techniques and the memorability scores for the short-term test and the long-term test, respectively. Special techniques offered no significant advantage or disadvantage in the short-term test, but in the long-term test, those who used special remembrance techniques clearly performed better than those who did not: the difference between the participants who correctly recalled their PINs was 36% ($p < 0.0001$, pairwise corrected FET). About 29% of the total number of participants reported using a special remembrance technique.

Among the 1,715 participants who used a special technique in the long-term test, 614 (36%) mentioned the use of 'keypad patterns', 166 (9.7%) mentioned the use of one of 'chunking', 'grouping', and 'splitting' technique, and 44 (2.6%) mentioned that they 'converted numbers to words' using letter associations on the keypad. Interestingly, from the 166 participants who used some form of chunking technique, 90 (54%) were given non-chunked policies (6, 7, 8), indicating that those participants decided for themselves to use chunking.

Table 15: Use of special techniques: Short-term memorability.

Group	*# Participants*	*# Failed*	*% Successful login*
Used	2,391	41	98%
Not used	5,636	137	98%
No ans	150	12	—

Table 16: Use of special techniques: Long-term memorability.

Group	*# Participants*	*# Failed*	*% Successful login*
Used	1,715	321	81%
Not used	3,545	1,933	45%
No ans	385	288	—

4.10 Ownership of 6-digit or longer PINs

By asking SQ5 from Table 3, we found that 28% of the participants own 6-digit or longer PINs in real life. An interesting observation is that those who own 6-digit or longer PINs performed better in the long-term memorability tests than those who do not own one (see Tables 17 and 18). The long-term memorability score difference between the two groups was statistically significant: 60% versus 56% ($p < 0.01$, pairwise corrected FET). This result indicates that memorability could be improved over time with training.

Table 17: Ownership of 6-digit or longer PINs: Short-term memorability.

Group	*# Participants*	*# Failed*	*% Successful login*
Owns 6-digit PIN	2,269	47	98%
No 6-digit PIN	5,747	131	98%
No ans	161	12	—

Table 18: Ownership of 6-digit or longer PINs: Long-term memorability.

Group	*# Participants*	*# Failed*	*% Successful login*
Owns 6-digit PIN	1,479	592	60%
No 6-digit PIN	3,773	1,660	56%
No ans	393	290	—

5. DISCUSSION

Our discussion of results is organized into several topics, according to the hypotheses we set up in Section 3.1. We also offer recommendations for PIN policies in organizations.

5.1 6-digit versus 4-digit PINs

We hypothesized that "the memorability of system-generated 6-digit PINs is worse than 4-digit PINs." As apparent in Tables 5 and 7, while policy 4 clearly outperformed 6 in both short-term and long-term memorability; only the result for long-term memorability showed statistical significance. Since long-term memorability is the what is desired, our findings *accept* the first hypothesis. Further, the memorability score gap was considerable in the long-term test, in which 6-digit PINs scored 19 points lower than 4-digit PINs (almost 26% drop). 6-digit PINs also showed longer authentication times with statistical significance. Banks should consider all of those memorability and usability trade-offs when moving from 4- to 6-digit system-generated PINs.

5.2 Should banks consider using 7 and 8-digit PINs?

Our second hypothesis stated that "the memorability of system-generated 6-digit PINs is better than that of 7- and 8-digit PINs." Our results show that between policies 6, 7, and 8, there is no statistically significant difference in memorability, not providing enough evidence to accept the second hypothesis (see Tables 6 and 8).

As for authentication time, 6 outperformed both 7 and 8 in the short-term test, but only outperformed 7 in the long-term test. This indicates that 6 loses its shorter authentication time advantage over 8 over time. Looking at those results, there is no reason for banks to rule out 7- or 8-digit system-generated PINs if they are considering increasing the PIN length.

Our PIN length group analysis (see Section 4.5), which grouped all policies of the same PIN length together, showed that there is no statistically significant difference between groups 6-All and 8-All and between groups 7-All and 8-All, but showed statistically significant inferiority of 7-All against 6-All. Hence, if enhancing PIN security is a primary concern for a bank, length 8 should also be considered and carefully evaluated.

5.3 Can chunking techniques improve PIN memorability?

Our third hypothesis predicted that "the memorability of longer (6-, 7- and 8-digit) system-generated PINs improves with chunking." While we observed improvements in both short-term and long-term memorability when using chunked PINs (see Tables 5 and 7), our analysis did not show any statistically significant differences between the chunked and their peer non-chunked policies. Hence, we do not have sufficient evidence to accept the third hypothesis (see Tables 6 and 8).

Surprisingly, policies 8:4-2-2 showed statistically significant superiority of 9% in long-term memorability over 7. This was the only case where a policy of a longer PIN length outperformed a policy of a shorter PIN length with statistical significance. Similarly, When we grouped chunking policies of the same PIN length together (see Section 4.4.2) and compared them against other non-chunked and grouped chunking policies, group 8-Chunk (54%) showed statistically significant superiority of long-term memorability over 7 (46%). Further, while no statistically significant difference was found among long-term memorability of 6-, 7- and 8-digit policies, policies 6:2-4 and 6:4-2 did show statistically significant difference (57% vs. 46%) compared to policy 7, and policy 6-Chunk outperformed both policy 7-Chunk and 7-digit PINs with statistical significance (57% vs 51%, and 57% vs. 46%; see Table 8).

Those mixed findings lead us to believe that, while chunking of system-generated random PINs may not be equally effective under all circumstances, they do show promise in certain cases and warrant a more focused study.

5.4 Policy recommendations

The findings of our study lead us to make the following recommendations for system-generated PINs.

- If a PIN length increase (from traditional 4-digit) is being considered, lengths 6 and 8 should all be considered.

- If 7- or 8- digit PIN lengths are being considered, chunking techniques such as 8:4-2-2 should be considered as chunking techniques seem to have some impact overall, and that policy in particular, can outperform shorter 7-digit PINs. However, the usability of the selected chunking policy should be studied more extensively (e.g., through a qualitative study) before deployment.

6. CONCLUSIONS AND FUTURE DIRECTIONS

We studied the memorability of system-generated PINs through a large-scale online user study, focusing on the effects of increasing the PIN length and applying number-chunking techniques that were traditionally applied to semantically meaningful chunks. Our results, not surprisingly, suggest that traditional 4-digit PINs have the best short-term and long-term memorability. While the memorability advantage of 4-digit PINs was small in the short-term, long-term memorability exhibited a significant drop when larger PIN lengths (6-, 7- and 8-digit) were used. What is interesting is that among 6-, 7-, and 8-digit PINs, we found no statistically significant difference in long-term memorability.

With regards to the effectiveness of chunking, we found that the number-chunking techniques used with larger PIN lengths did not provide a statistically significant improvement in memorability over their corresponding non-chunked PINs. However, chunked PINs did show significant improvements in some cases such as 8-digit chunking policy (0000-00-00) which exhibited statistically significant superiority in memorability against a non-chunked 7-digit policy (that is shorter in length). Further study is needed to understand this intriguing observation.

Our study used a 48 hour interval to study long-term memorability. It would be interesting to study long-term memorability of system-generated PINs using longer PIN recall intervals and compare findings as users may not necessarily use their PINS within 48 hours of assignment or use them every 48 hours. Similarly, it would be interesting to study how long-term memorability changes with multiple PIN recall sessions—especially where each recall session is also used as a remembrance training opportunity. Further, it would be interesting to study the impact on memorability when semantics are associated with chunks either through the use of mnemonics or training.

Our near term future work is to analyze the data collected to study correlations between PIN memorability and self-identified demographic and memory strength characteristics of participants.

7. ACKNOWLEDGEMENTS

This work was supported in part by the National Research Foundation of Korea (No. 2014R1A1A1003707), the ITRC (IITP-2015-H8501-15-1008), the NIPA (NIPA-2014-H0301-14-1010), the Information Trust Institute at University of Illinois, and the School of EECS at Oregon State University.

Authors would like to thank Andrew Patrick for shepherding the paper, and all the anonymous reviewers for their valuable feedback. Authors would also like to thank David Nicol of Information Trust Institute for supporting the initial study, and Ji Won Yoon for his help with the statistical analysis.

8. REFERENCES

[1] Amazon Mechanical Turk. https://www.mturk.com/mturk/welcome, 2014.

[2] hashcat advanced password recovery. http://hashcat.net/oclhashcat/, 2014.

[3] R. Atkinson and R. Shiffrin. Human memory: A proposed system and its control processes. volume 2 of *Psychology of Learning and Motivation*, pages 89 – 195. Academic Press, 1968.

[4] A. D. Baddeley and G. Hitch. Working memory. volume 8 of *The psychology of learning and motivation: Advances in research and theory*, pages 47–89. Academic Press, New York, NY, USA, 1974.

[5] M. Bishop. Password management. In *Compcon Spring '91. Digest of Papers*, pages 167–169, Feb 1991.

[6] J. Bonneau, C. Herley, P. van Oorschot, and F. Stajano. The quest to replace passwords: A framework for comparative evaluation of web authentication schemes. In *Security and Privacy (SP), 2012 IEEE Symposium on*, pages 553–567, May 2012.

[7] A. S. Brown, E. Bracken, S. Zoccoli, and K. Douglas. Generating and remembering passwords. *Applied Cognitive Psychology*, 18(6):641–651, 2004.

[8] W. E. Burr, D. F. Dodson, and W. T. Polk. Electronic authentication guideline. Technical report, 2006.

[9] D. S. Carstens and L. C. Malone. Applying Chunking Theory in Organizational Password Guidelines. *Journal of Information, Information Technology, and Organizations*, 2006.

[10] N. Cowan. The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, 24(1):87–114, 2001.

[11] T. J. Druzgal and M. D'esposito. Dissecting contributions of prefrontal cortex and fusiform face area to face working memory. *Journal of Cognitive Neuroscience*, 15(6):771–784, Aug. 2003.

[12] S. Fahl, M. Harbach, Y. Acar, and M. Smith. On the ecological validity of a password study. In *Proceedings of the Ninth Symposium on Usable Privacy and Security*, SOUPS '13, pages 13:1–13:13, New York, NY, USA, 2013. ACM.

[13] S. Gaw and E. W. Felten. Password management strategies for online accounts. In *Proceedings of the second symposium on Usable privacy and security*, SOUPS '06, pages 44–55, New York, NY, USA, 2006. ACM.

[14] F. Gobet, P. C. R. Lane, S. Croker, P. C. H. Cheng, G. Jones, I. Oliver, and J. M. Pine. Chunking mechanisms in human learning. *Trends in Cognitive Sciences*, 5(6), June 2001.

[15] D. Goodin. Anatomy of a hack: How crackers ransack passwords like "qeadzcwrsfxv1331". `http://arstechnica.com/security/2013/05/how-crackers-make-minced-meat-out-of-your-passwords/`, May 2013.

[16] C. Herley and P. van Oorschot. A research agenda acknowledging the persistence of passwords. *Security Privacy, IEEE*, 10(1):28–36, Jan 2012.

[17] P. G. Inglesant and M. A. Sasse. The true cost of unusable password policies: password use in the wild. In *Proceedings of the 28th international conference on Human factors in computing systems*, CHI '10, pages 383–392, New York, NY, USA, 2010. ACM.

[18] S. P. Joseph Bonneau and R. Anderson. A birthday present every eleven wallets? The security of customer-chosen banking PINs. In *FC' 12: The 16 th International Conference on Financial Cryptography and Data Security*, 2012.

[19] P. Kelley, S. Komanduri, M. Mazurek, R. Shay, T. Vidas, L. Bauer, N. Christin, L. Cranor, and J. Lopez. Guess again (and again and again): Measuring password strength by simulating password-cracking algorithms. In *Security and Privacy (SP), 2012 IEEE Symposium on*, pages 523–537, May 2012.

[20] H. Kim and J. H. Huh. PIN selection policies: Are they really effective? *Computers & Security*, 31(4):484–496, 2012.

[21] S. Komanduri, R. Shay, P. G. Kelley, M. L. Mazurek, L. Bauer, N. Christin, L. F. Cranor, and S. Egelman. Of passwords and people: measuring the effect of password-composition policies. In *Proceedings of the 2011 annual conference on Human factors in computing systems*, CHI '11, pages 2595–2604, New York, NY, USA, 2011. ACM.

[22] C. Kuo, S. Romanosky, and L. F. Cranor. Human selection of mnemonic phrase-based passwords. In *Proceedings of the second symposium on Usable privacy and security*, SOUPS '06, pages 67–78, New York, NY, USA, 2006. ACM.

[23] M. A. Sasse, S. Brostoff, and D. Weirich. Transforming the 'weakest link' – a human/computer interaction approach to usable and effective security. *BT Technology Journal*, 19:122–131, July 2001.

[24] B. Schneier. Choosing Secure Passwords. `https://www.schneier.com/blog/archives/2014/03/choosing_secure_1.html`, March 2014.

[25] R. Shay and E. Bertino. A comprehensive simulation tool for the analysis of password policies. *International Journal of Information Security*, 8:275–289, August 2009.

[26] R. Shay, A. Bhargav-Spantzel, and E. Bertino. Password policy simulation and analysis. In *Proceedings of the 2007 ACM workshop on Digital identity management*, DIM '07, pages 1–10, New York, NY, USA, 2007. ACM.

[27] R. Shay, S. Komanduri, A. L. Durity, P. S. Huh, M. L. Mazurek, S. M. Segreti, B. Ur, L. Bauer, N. Christin, and L. F. Cranor. Can long passwords be secure and usable? In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '14, pages 2927–2936, 2014.

[28] R. Shay, S. Komanduri, A. L. Durity, P. S. Huh, M. L. Mazurek, S. M. Segreti, B. Ur, L. Bauer, N. Christin, and L. F. Cranor. Can long passwords be secure and usable? In *Proceedings of the 32Nd Annual ACM Conference on Human Factors in Computing Systems*, CHI '14, pages 2927–2936, New York, NY, USA, 2014. ACM.

[29] R. Shay, S. Komanduri, P. G. Kelley, P. G. Leon, M. L. Mazurek, L. Bauer, N. Christin, and L. F. Cranor. Encountering stronger password requirements: user attitudes and behaviors. In *Proceedings of the Sixth Symposium on Usable Privacy and Security*, SOUPS '10, pages 2:1–2:20, New York, NY, USA, 2010. ACM.

[30] M. A. Thornton and A. R. A. Conway. Working memory for social information: Chunking or domain-specific buffer? *NeuroImage*, 70:233–239, 2013.

[31] B. Ur, P. G. Kelley, S. Komanduri, J. Lee, M. Maass, M. L. Mazurek, T. Passaro, R. Shay, T. Vidas, L. Bauer, N. Christin, and L. F. Cranor. How Does Your Password Measure Up? The Effect of Strength Meters on Password Creation. In *Proceedings of the 21st USENIX Conference on Security Symposium*, Security '12, 2012.

[32] K.-P. L. Vu, R. W. Proctor, A. Bhargav-Spantzel, B.-L. B. Tai, J. Cook, and E. E. Schultz. Improving password security and memorability to protect personal and organizational information. *International Journal of Human-Computer Studies*, 65(8):744–757, 2007.

[33] J. Yan, A. Blackwell, R. Anderson, and A. Grant. Password memorability and security: empirical results. *Security Privacy, IEEE*, 2(5):25–31, 2004.

Evaluating the Effectiveness of Using Hints for Autobiographical Authentication: A Field Study

Yusuf Albayram
Department of Computer Science & Engineering
University of Connecticut, Storrs, CT, USA
yusuf.albayram@uconn.edu

Mohammad Maifi Hasan Khan
Department of Computer Science & Engineering
University of Connecticut, Storrs, CT, USA
maifi.khan@engr.uconn.edu

ABSTRACT

To address the limitations of static challenge question based authentication mechanism, recently smartphone-based autobiographical authentication mechanisms are being explored where challenge questions are generated using users' day-to-day activities captured by smartphones dynamically. However, users' poor recall rate in such systems is still a significant problem that negatively affects the usability of such systems. To address this challenge, this paper investigates the possibility of using hints that may help users to recall recent day-to-day events more easily and explores various design alternatives for generating hints. Specifically, in this paper, we generate challenge questions and hints for three different kinds of autobiographical data (e.g., call logs, SMS logs, and location logs), and evaluate the effect of different question types and hint types on user performance by conducting a real-life study with 24 users over a 30 day period. To test whether hints are useful/harmful for adversaries' response accuracy, we simulate various kinds of adversaries (e.g., naive and knowledgeable) by recruiting volunteers in pairs (e.g., close friends, significant others). In our study, we observed that, for legitimate users, hint was effective for all different question types. Interestingly, we found that hint has negative effect on strong adversarial users and no significant effect on performance for naive adversarial users.

1. INTRODUCTION

Several recent research efforts have investigated the idea of autobiographical authentication leveraging smartphone usage data (e.g., call log, SMS log) and highlighted the advantages of these systems over static challenge question based approaches due to the dynamicity offered by such smartphone-based solutions [4, 13]. However, as recall of information relies solely on users' memory in such systems, which are often imperfect and unreliable [19], error rates while answering autobiographical questions generated based on day-to-day activity logs (e.g., call log, SMS log) are often high [13], negatively affecting the usability of such systems [4]. Interestingly, while providing cues/hints may act as a "memory trigger" to jog one's memory and can be an effective way to improve the recall rate and reduce the error rate, only a small number of prior efforts looked into the possibility of using hints to facilitate recall and mostly focus on static password based systems [21, 32]. To address this void and complement prior efforts, in this paper, we focus on generating hints for autobiographical authentication systems where hints are generated along with the challenge questions dynamically. Specifically, for hint and challenge question generation, the presented system leverages events that are related to *episodic memory* [11] which refers to memories related to everyday experiences (e.g., call logs) instead of events that are related to *semantic memory* which refers to memories related to facts and general knowledge (e.g., important events such as graduation) [24]. This is due to the fact that, in case of autobiographical authentication, events that are related to *episodic memory* is more suitable and relevant for hint and challenge question generation as these are the types of memories that are formed every day and have a short shelf life (i.e., they are forgotten within days or weeks), offering the required dynamicity compared to relatively less dynamic semantic memory.

Based on these observations, in our study, we investigate hint generation algorithms that consider frequency of events along with historical patterns of a user to generate customized hints that may help a user to recall near past events without revealing the actual information. Specifically, we consider three different kinds of day-to-day events for generating challenge questions and hints (e.g., phone call send and receive events, SMS message send and receive events, location-visit events). As hints may reveal secrets and help adversaries to learn about a user if not designed carefully, over multiple iterations, we finalized the design of hints for specific event types that present information in a non-revealing way. To study the effect of hints on users' performance, we recruited 24 users and conducted a real-life study for over a month. To test whether hints are useful/harmful for adversaries' accuracies, we simulate two different kinds of adversaries (e.g., naive vs. knowledgeable) by recruiting volunteers in pairs (e.g., close friends). Over the course of the study, each user is periodically presented with three sets of challenge questions. The first set is generated based on users' own data. The second set is generated based on a randomly selected user's data. Finally, each user is presented with a third set of questions which is generated based on user's friend's data.

Symposium on Usable Privacy and Security (SOUPS) 2015, July 22–24, 2015, Ottawa, Canada.

In our study, we observed that, for legitimate users, hints were effective for all different question types. We also found that hints had no significant effect on response correctness of naive adversaries and had negative effect on strong adversaries. Finally, users' exit interview data suggests that legitimate users found hints to be helpful and indeed improve the usability of such systems. To summarize, this paper makes the following key contributions:

1. Investigate the effect of hints on legitimate users' response correctness for different types of autobiographical challenge questions.

2. Simulate different kinds of adversarial users (e.g., naive vs. knowledgeable) by recruiting participants in pairs (e.g., close friends), and investigate the effect of hints on adversarial users' response correctness for different types of autobiographical challenge questions.

3. Finally, investigate the effect of hints on usability of such systems through performance change and qualitative feedback collected using an exit survey.

The rest of the paper is organized as follows. Section 2 presents prior work that are related to our current study and discusses the basic concepts underlying our analysis. Section 3 explains the study design and describes how the autobiographical questions and hints are generated. Key findings along with detailed analysis are presented in Section 4. Limitations of our study along with possible future directions are discussed in Section 5. Finally, Section 6 concludes the paper.

2. RELATED WORK

To facilitate resetting of passwords or provide an extra layer of security for authentication, various fallback authentication mechanisms are being investigated. One of the most widely used approaches is known as knowledge-based authentication (KBA) (e.g., static challenge question based technique) [7–9, 35]. KBA can be further divided into two categories, namely, static KBA and dynamic KBA. In static KBA, the questions are often predetermined (e.g., "What is the name of your first pet?"), and are often susceptible to various forms of vulnerabilities such as easy predictability and poor recall rate [6, 22, 29–31, 33–35, 41]. Furthermore, static questions are becoming weaker due to improved information retrieval techniques and increase in online content [31]. For instance, by mining online sources (e.g., social networking sites, public records), an attacker often can obtain the details about one's personal information to answer many of the challenge questions commonly used for backup authentication.

To address the limitations of static KBA schemes, dynamic KBA schemes are being investigated recently where challenge questions are generated on the fly based on user's recent activities such as browsing history [5], Facebook activity [12], calendar events [27], user's email history [26], electronic personal history [28], or financial activity (e.g., several major American credit bureaus authenticate users by generating questions based on past financial transactions).

More recently, Gupta et al. [16] investigated the memorability of users' smartphone usage behavior (e.g., emails, calendar events, calls) and attempted to leverage that to authenticate users. One of the main limitations of this work is that the challenge questions are generated based on a user's routine (e.g., who do you call the most?) rather than day-to-day activities which are more dynamic. Similarly, another work used email activities to generate challenge questions (e.g., who sent you the most emails?) [40]. Das et al. [13], Albayram et al. [4], and Hang et al. [18] presented authentication frameworks that exploit smartphone usage data (e.g., phone call history, location traces, app usage) to generate dynamic challenge questions.

While these recent efforts on smartphone based autobiographical authentication mechanisms report encouraging results, they rely solely on users' memory for recall which is often imperfect and thereby unreliable [19], causing poor recall rate [4, 13]. Although providing cues or hints can be an effective way to jog one's memory and improve recall rate, none of these works has investigated the possibility of using hints to facilitate recall. Interestingly, a number of prior efforts have looked into this possibility in different contexts. For example, Wagenaar [38] studied the recall performance of a person's daily life events, and pointed out the effectiveness of providing cues on memory retention. He also highlighted that providing different forms of cues (e.g., who, when, where) increase the chance of recalling an event. Vemuri et al. [37] noted that one needs only a trigger rather than original content in order to remember.

There is a limited number of works in the literature that looked into the possibility of using hints in authentication settings. For example, Hertzum [21] proposed using minimal-feedback hints where users select certain characters of their passwords that will be revealed during password entry in order to jog users' memory. He conducted a user study with 14 users and found that, while hints aid users' to recall their passwords, the selected passwords were weak. Similar findings were found by Lu et al. [25]. Renaud et al. [32] investigated the use of some abstract images (e.g., Cueblot) as password cues. However, they found that the presence of abstract images did not have a positive effect on users' performance in password-based authentication.

Based on prior studies, we identify that, in case of autobiographical based systems, events related to *episodic memory* is more suitable for question and hint generation as these are the types of memories that are formed every day and have a short shelf life (i.e., they are forgotten within days or weeks). Specifically, Conway [10] conducted a study in which participants were asked to list as many specific memories as they can remember from yesterday, from two days back, from three days back, and so on. They found that users can recall a good number of events that are one day old compared to events that are older. Beyond a 3-day retention interval, memories appear to be much more concerned with routines and schema than with specific episodic memories. Further, Kristo et al. [24] examined several factors that may influence recall rate of recent autobiographical events using an Internet-based diary study. They found that the content and the time of the events were remembered better compared to the details of the events. Also, among the time elements, time of the day was remembered better. Events that occur less frequently were also remembered better compared to events that occur frequently.

Motivated by these prior efforts, in this work, we looked into the challenge of generating effective hints for smartphone-based dynamic authentication mechanism by identifying events that are more likely to be remembered by a user. While our work is inspired by prior efforts, our work differs from prior work in several aspects. First, to the best of our knowledge, we are the first to investigate the challenge of generating hints dynamically for smartphone based autobiographical authentication systems. Second, we conducted a real-life study that investigates the strengths and weaknesses of providing hints for different categories of questions and users (e.g., legitimate, naive adversary, strong adversary). Finally, we evaluate the effect of hints on usability aspect of such systems through an interview style exit survey. The details of our work is presented in the following sections.

3. METHODOLOGY

In this section, we describe the smartphone application that was developed for collecting and analyzing autobiographical data along with the algorithms for question and hint generation. We then present the design of the study. The details are below.

3.1 Autobiographical Data Collection Application

We developed an android application for devices running Android 2.3 or higher to collect and analyze autobiographical user data. The application collects the communication history and the location traces of a user while running in the background. It then generates challenge questions and hints using the collected data. Table 1 lists the details of the data that are collected in our study.

Data	Details of collected data
Call	Type (outgoing, incoming), Duration, Name of the person, Time
SMS	Type (sent, received), Receiver/Sender Name, Length of SMS message, Time
Location	Latitude, Longitude, Time, Accuracy (i.e., the expected error bound)

Table 1: Details of the collected data.

In order to obtain the location information with minimal energy overhead, we utilize the latest Google Fused Location API [1] along with Android's activity recognition API [2]. Specifically, Android's activity recognition API provides an easy way to detect if a user is moving or not (e.g., walking, biking, or in a vehicle). The application software leverages the Android's activity recognition API to decide whether to track location or not. For example, when a user is not moving, the app does not track location at all. However, if a user is walking, biking, or in a moving vehicle, the app starts logging location data. Once the data items are collected, the question generation component generates challenge questions as follows.

3.2 Autobiographical Question Generation Component

Using the aforementioned data items, the application generates 5 different types of questions as listed in Table 2. Details about each type of question are below.

Questions Based on Communication Activity

Communication questions are generated based on a user's recent communication history (e.g., SMS history that includes both sent and received messages, and call history that includes both incoming and outgoing phone calls). This category of questions asks a user to recall the name of the person he/she called or SMS messaged, or the name of the person who called him/her or SMS messaged him/her at a certain time. Examples of communication questions are shown in Figure 1(a) and Figure 1(b). For this type of question, a user is asked to enter the answer (i.e., person name) into a textbox. To enhance the usability, we utilize the "auto-complete" feature which suggests possible entries as a user types in the textbox. This is especially helpful as "auto-complete" feature reduces potential errors due to possible misspellings.

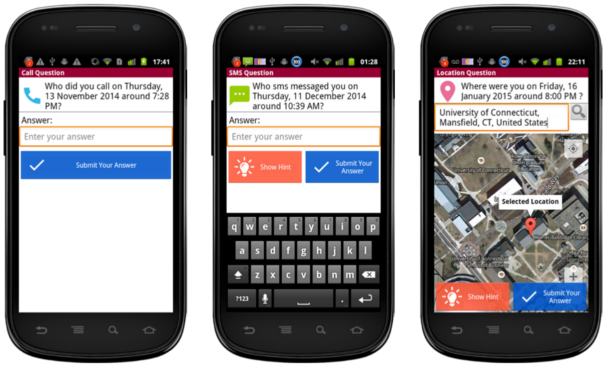
Question Type	Question
Phone call (Incoming)	Who called you at <time>?
Phone call (Outgoing)	Who did you call at <time>?
SMS (Received)	Who sent you the SMS message at <time>?
SMS (Sent)	Who did you send the SMS message at <time>?
Location	Where were you at <time>?

Table 2: List of the question types.

Questions Based on Location Information

Location questions are generated based on a user's recent location traces tracked by the application. The collected location data is composed of a sequence of coordinates with latitude, longitude, and the relevant temporal information (i.e., time stamp). To avoid considering each geographical coordinate as a unique physical location, in our work, we use a clustering algorithm that groups geographical coordinates based on their distance in order to infer user's locations. However, as a user may visit new places over time and we do not know a priori the total number of places a user may visit, we chose to employ a density-based clustering approach that can incrementally adapts the number of clusters (i.e., the number of distinct physical locations). Specifically, we use the DBSCAN (Density-Based Spatial Clustering of Applications with Noise) algorithm [14] that is based on the notion of density reachability. Briefly, the DBSCAN algorithm requires two parameters: ϵ distance threshold (e.g., 75 meters) and min_{pts} minimum number of points within a cluster. The algorithm starts by assigning a random point in a cluster and expands it with neighborhoods of at least min_{pts} points that are within a distance ϵ from it. In our work, to calculate the distance between two geographical coordinates, we use Haversine Distance, though the algorithm can work with any distance function. Based on the distance of examined coordinates and ϵ distance threshold, the DBSCAN algorithm either creates new clusters or expands/updates the existing clusters. As new coordinates arrive, new coordinates are first examined to determine whether they can be assigned to any of the existing clusters. If not, the new coordinates are given as input to the DBSCAN algorithm to regenerate the clusters including the new locations. The output of this algorithm is a set of clusters that is used to generate questions. Please note that this algorithm only runs whenever the application needs to generate location questions for a user.

For location-based questions, a user is presented an interactive map and is asked to select the location that he/she had visited during a certain time window of a specific day. The interactive map was implemented leveraging Google Maps Android API [3] where the initial zoom level was set to 1 to make the most of the world visible. The rationale behind this choice is to avoid influencing users to select locations



Who did you call on Thursday, 13 November 2014 around 7:28 PM?

Who sms messaged you on Thursday, 11 December 2014 around 10:39 AM?

Where were you on Friday, 16 January 2015 around 8:00 PM ?

University of Connecticut, Mansfield, CT, United States

(a) Call (b) SMS (c) Location

Figure 1: Screenshots showing different types of questions.

from a certain geographic area, which may reduce the overall security of the system [36]. In order to select a location on the map, the minimum required zoom level is set to 16, which gives reasonable details and higher security since an adversary has to guess a location at a finer resolution. A user needs to long press on the map to pin his/her location and set a marker at the selected location (e.g., like the one in Figure 1(c)). Instead of zoom in/out manually, user may also use a search box to zoom-in on the right area/location very quickly.

Algorithm for Generating Challenge Questions
As a user often makes/receives a large number of phone calls and/or sends/receives a large number of SMS messages, and/or visits many different places in a given day, it is non-trivial to pick the specific instance of an event that may be used to generate the question. Ideally, the system should pick an event that is easy for a legitimate user to recall but hard for an adversary to guess.

To address this challenge, we develop an algorithm that gives preference to rare events compared to more predictable events. Intuitively, if a person rarely receives a phone call from person X, it is more likely that he/she will remember that event. To implement the algorithm, we represent a user's history H as a sequence of certain type of events (e.g., phone call). In H, each event is represented as a triplet of the form $e_i = (a_i, d_i, t_i)$, where a_i represents an activity (e.g., making a phone call), d_i is the duration of the event, and t_i is the time-stamp of the event. Assuming that n activities were recorded for a user in a given time frame, the history for that time frame will be represented as a time ordered sequence of triplets, and will be denoted as $H = \{e_1, ..., e_n\}$. Subsequently, we convert the history H into a Time Window-Event Matrix as shown in Table 3 by splitting each day into a set of m time windows $W = \{w_0, ..., w_m\}$ of fixed size (e.g., 1 hour). Next, each event is assigned to a specific time window W_i based on the event's time-stamp t_i.

Once events are assigned in specific time windows, the system computes an "interestingness" weight for each event based on statistical measure of randomness, and attempts to pick events for generating questions by giving preference to more infrequent events in a user's schedule. To identify the infrequent events for a given *Time Window-Event Matrix* (e.g., as shown in Table 3), the algorithm analyzes daily and weekly activity patterns of a user and calculates the weight for an event as follows.

1. First, the algorithm calculates $P(e_i)$ which denotes the probability of an event e_i. For example, probability of calling John in the last 30 days based on call log data.

2. Next, calculate $P(e_i|w_m)$ which denotes the probability of event e_i for a specific time interval w_m. For example, the probability of calling John between 10:00 am and 11:00 am in the last 30 days. This probability is calculated to identify daily patterns.

3. Next, calculate $P(e_i|w_m, dow_k)$ where dow_k denotes the "day of the week" from the set $DOW = \{dow_1, ..., dow_7\}$ where $dow_1 = Monday, \ldots, dow_7 = Sunday$. $P(e_i|w_m, dow_k)$ denotes the probability of an event e_i for a time interval w_m on day dow_k of the week. For example, the probability of calling John between 10:00 am and 11:00 am on Mondays in the last 30 days. This probability is calculated to identify weekly patterns.

4. Finally, to give priority to long lasting events which are more likely to be remembered by a user easily, the algorithm calculates T_e^i which denotes the sum of the duration of event e_i in the history H and subsequently, multiply with d_i (duration of the event). For example, multiply a recent phone call duration made to John with the sum of the duration of phone calls that made to John in the last 30 days based on call log data. The main intuition behind this multiplication is that we want to give priority to the latest events that lasted longer compared to other events of the same type.

5. Based on the above probabilities, we compute the *weight* of an event as follows:

$$Weight = \frac{P(e_i)\, P(e_i|w_m)\, P(e_i|w_m, dow_k)}{T_e^i \times d_i}$$

   Once weight for individual events are calculated, the algorithm sorts all events based on *weight* and pick according to that order whenever the system needs to generate challenge questions for a particular data type. Please note that the lower the weight, the higher the chance of that event to be selected by the algorithm.

Due to the above scheme, higher weight questions that are relatively easy to guess because of "regularity" are filtered out and the preference is given to more infrequent events which are more likely to be harder to guess but easier to recall by legitimate users.

Please note that the above scheme can be applied for any data types such as call log, SMS log, and location log. However, necessary changes may need to be made based on data types. For instance, for SMS log, there is no duration for SMS messages, and thus duration needs to be ignored or may be replaced with the length of SMS messages. For location log, to avoid considering each geographical coordinate as a unique physical location, geographical coordinates need to be clustered first.

3.3 Hint Generation

Since human memory is fallible, when it comes to autobiographical authentication where the challenge questions

Window \Day	Nov 27	Nov 28	...	Dec 27
$00:00-00:59$	$-$	$-$	...	$\{Receivedcallfrom-Jeff,55sec\}$
$01:00-01:59$	$-$	$-$	...	$-$
$\vdots$	$\vdots$	$\vdots$	$\vdots\ \vdots\ \vdots$	$\vdots$
$14:00-14:59$	$\{Called-Alice,55sec\},\{Called-Bob,32sec\}$	$\{Called-Bob,89sec\}$	...	$\{Called-Bob,17sec\}$
$15:00-15:59$	$\{Called-John,300sec\}$	$\{Receivedcallfrom-Jeff,42sec\}$	...	$-$
$16:00-16:59$	$\{Called-John,14sec\}$	$\{Called-Bob,20sec\}$	...	$\{Called-Bob,89sec\}$
$17:00-17:59$	$\{Called-Bob,27sec\}$	$-$	...	$-$
$\vdots$	$\vdots$	$\vdots$	$\vdots\ \vdots\ \vdots$	$\vdots$
$23:00-23:59$	$-$	$\{Receivedcallfrom-Bob,14sec\}$	...	$\{Called-Mike,14sec\}$

Table 3: Time Window-Event Matrix of a user's phone call log history. Numbers right next to a person's name indicates the duration of the phone calls in seconds.

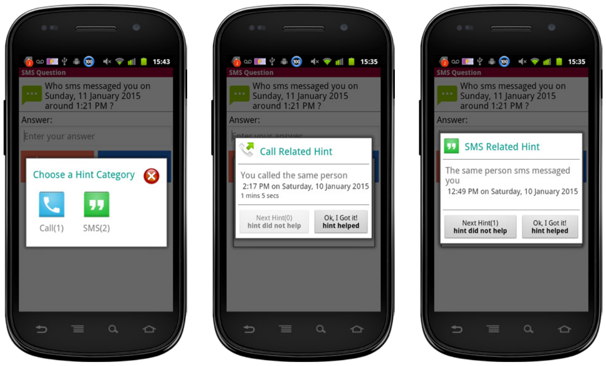
are generated using users' everyday interactions with their smartphones, a user may not always remember who he/she called or texted at a specific time. However, providing some auxiliary information about a certain event as hints may help users to recall that particular event. For example, a user may not remember whom he texted at 3 pm today, as the user may have texted more than one person around that time, making it a bit difficult to guess which person the question is referring to. However, if the user is provided some auxiliary information such as "the same person texted at 11:47pm yesterday", the user is more likely to remember the answer. Please note that such hints do not reveal any privacy sensitive information other than the fact that he texted someone at 11:47 pm yesterday.
While hints might be helpful for users, generating hints is a nontrivial problem and ideally should satisfy the following properties.

1. **Efficacy:** A hint should be useful for legitimate users.

2. **Ambiguity:** A hint should not be useful for anyone else other than the legitimate user.

3. **Privacy:** A hint should not leak/reveal user's privacy sensitive information (i.e., preserve the privacy of the legitimate user).

To ensure the above properties, in this work, we designed different kinds of hints as follows.
For communication questions (i.e., Call and SMS), hints are generated based on recent communication events that involves the same receiver/sender that the challenge question is asking about. Intuitively, knowing that the user talked (or messaged) to the same person a few hours or days earlier can jog the user's memory. For example, if the question is "Who did you call at 2 pm on Wednesday?" and the correct answer is "John", information regarding phone call(s) made to "John" or received from "John" within the last few days (e.g., 2 days) can be used as hints (e.g., sample hint: you called the same person on Wednesday at 11:25 am). Similarly, information regarding SMS message(s) received from "John" or sent to "John" may be used as hints as well (e.g., sample hint: you sent a SMS message to the same person on Wednesday at 1 pm). For location questions, hints are generated based on recent historical location information (e.g., sample hint: you visited the same place on Monday at 11 am).
Furthermore, in the absence of recent historical information relevant to the event, we generate hints in a negative format.


SMS Question

Who sms messaged you on Sunday, 11 January 2015 around 1:21 PM ?

Answer:

Enter your answer

Choose a Hint Category

Call Related Hint

You called the same person 2:17 PM on Saturday, 10 January 2015 1 mins 5 secs

SMS Related Hint

The same person sms messaged you 12:49 PM on Saturday, 10 January 2015

15:43
15:35
15:35
Call(1)
SMS(2)
Next Hint(0) hint did not help
Ok, I Got it! hint helped
Next Hint(1) hint did not help
Ok, I Got it! hint helped

Figure 2: Screenshots showing a SMS question along with the provided hints.

For example, we may generate hints such as "You did not call this person within the last 2 days". All different kinds of hints that may be generated for different types of questions are listed in Table 4.
In case there are multiple hints in different categories, users are provided with an option to choose from any of the available hint categories they want (e.g., Call, SMS). Finally, once a user uses a hint during the session, the user is asked whether the viewed hint was helpful or not. A sample question along with the provided hints are shown in Figure 2.

3.4 User Score Calculation

As users may make different types of mistakes for different question types while answering the challenge questions, we need to calculate the score differently for different question types. For example, in case of communication questions (i.e., Call and SMS), a user can either pick the answer from a list of suggestions that are populated using an "auto-complete" functionality, or a user may type in his/her answer instead of selecting from the list of names suggested by the autocomplete feature. However, while typing, a user may make spelling mistakes. For example, a common first name "Adrianna" may be spelled as "Adrianne" or "Adrienne". Thus, instead of scoring the answer based on an exact match (where the correct answer and the user's answer are matched 100%), in the current implementation, there is an error tolerance to accommodate typing errors (e.g., 85% similarity score between two strings). Specifically, if the Jaro-Winkler distance between the entered text and the correct answer is greater than 85% similarity score, the answer is considered to be cor-

Question Type	Hint Type and Hints
Call (e.g., Who did you call on <time>?)	1) Call/Outgoing/Positive: You called the same person on <time>, and the duration of the phone call was <duration>.
	2) Call/Incoming/Positive: The same person called you on <time>, and the duration of the phone call was <duration>.
	3) Call/Outgoing/Negative: You did not call this person within the last <♯ of days>.
	4) Call/Incoming/Negative: This person did not call you within the last <♯ of days>.
	5) SMS/Incoming/Positive: The same person SMS messaged you on <time>.
	6) SMS/Outgoing/Positive: You SMS messaged the same person on <time>.
	7) SMS/Outgoing/Negative: You did not SMS messaged this person within the last <♯ of days>.
	8) SMS/Incoming/Negative: This person did not SMS message you within the last <♯ of days>.
SMS (e.g., Who did you SMS message on <time>?)	1) SMS/Outgoing/Positive: You SMS messaged the same person on <time>.
	2) SMS/Incoming/Positive: The same person SMS messaged you on <time>.
	3) SMS/Outgoing/Negative: You did not SMS messaged this person within the last <♯ of days>.
	4) SMS/Incoming/Negative: This person did not SMS message you within the last <♯ of days>.
	5) Call/Outgoing/Positive: You called the same person on <time>, and the duration of the phone call was <duration>.
	6) Call/Incoming/Positive: The same person called you on <time>, and the duration of the phone call was <duration>.
	7) Call/Outgoing/Negative: You did not call this person within the last <♯ of days>.
	8) Call/Incoming/Negative: This person did not call you within the last <♯ of days>.
Location (e.g., Where were you on <time>?)	1) Location/Positive: You were in the same location on <time>, and you stayed there <duration>.
	2) Location/Negative: You did not visit this location within the last <♯ of days>.

Table 4: List of possible hints for different types of questions.

rect. Otherwise, the score is set to 0. Please note that the Jaro-Winkler distance metric is best suited for comparing short strings such as names [39].

In case of location questions, as users may not place the marker on exactly the same location coordinates estimated and identified by the system, there is an error tolerance (e.g., 75 meters great circle distance) in our system, which is calculated based on the Haversine distance formula[1] between the selected coordinates and the estimated location. If the distance between the selected geographical location and the estimated location is greater than 75 meters, the answer is considered to be incorrect and the score is set to 0.

3.5 Study Design

To evaluate our system, we recruited 24 participants from the college campus through the university email list server. To simulate strong adversaries, we recruited participants in pairs (e.g., close friends, significant others). The social relationships between the pairs of participants are shown in Table 5.

Over the course of the experiment, each participant was presented with three sets of questions multiple times each week. The first set of question was generated based on participant's own data. For example, a participant would receive a phone call question in the following format: "Who did you call at 11:25am on Wednesday?". The second set of question was generated based on participant's pair's (e.g., close friend or couple) data. In this case, the role of a strong adversary is played by the pair of each participant. For example, the participant would receive a phone call question about his/her partner in the following format: "Who did your partner call at 4:20pm on Friday?". The third set of question was generated based on a randomly selected participant's data whose identity was not revealed to the participant who answered the question. In this case, participants played the role of a naive adversary. For example, the participant would receive a phone call question about a stranger in the following format: "Who did a stranger call at 2:51pm on Monday?".

In order to evaluate the effectiveness of using hints for autobiographical authentication, we devised a within-subject user study with two conditions. In order to have consistent comparisons between questions with hint(s) and with no hint, a question was asked twice. In the first condition, a question was presented without hint and subsequently, in the second condition, the same question was presented with hint. In all cases, participants were not given any feedback regarding his/her performance throughout the study.

Each participant was compensated with a $25 Amazon gift card for two weeks of participation. The study was approved by the University's Institutional Review Board (IRB).

Pair#	Relationship	Closeness Rate	Live Together
Pair-1	Friends	4	No
Pair-2	Friends	5	No
Pair-3	Friends	4	No
Pair-4	Friends	5	No
Pair-5	Friends	4	No
Pair-6	Friends	4	No
Pair-7	Friends (Roommates)	5	Yes
Pair-8	Friends (Roommates)	5	Yes
Pair-9	Boyfriend/Girlfriend	5	No
Pair-10	Boyfriend/Girlfriend	5	No
Pair-11	Boyfriend/Girlfriend	5	No
Pair-12	Boyfriend/Girlfriend	5	No

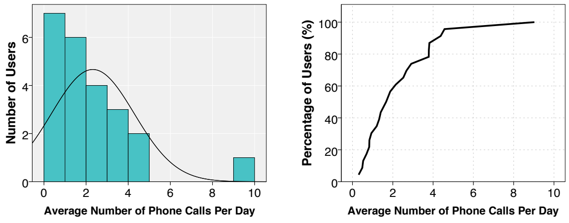
Table 5: The social relationships between the pairs of participants and their ratings on how well they know each other on a Likert-scale of 1 (Very little) to 5 (Pretty well). The last column shows whether participants live together or not.

4. FINDINGS

During a period of 30 days, from 24 participants (12 paired participants), we collected a total of 3296 question-answer responses where the questions were presented with no hints and 3296 question-answer responses where the questions were presented with hints. Out of 3296 questions, participants used hints in 832 questions. One of the participants withdrew from the study after two weeks of participation. All participants (10 female, 14 male) were undergraduate students from a broad range of degree programs. The age of participants ranged from 18 to 23 years with an average age of 19.33 years (Median=19 years with SD=1.28).

Figure 3 shows the statistics for phone call data including the number of phone calls made and received. As shown in the histogram in Figure 3(a), the plot appears to be right-skewed as most participants make 1 - 4 phone calls per day. Figure 3(b) shows the cumulative distribution of participants with respect to the average number of phone calls per day. 80% of

[1] `http://en.wikipedia.org/wiki/Haversine_formula`



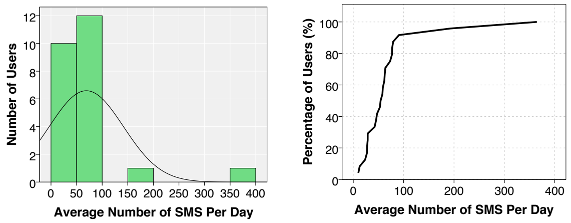
(a) (b)

Figure 3: (a) Histogram of average number of phone calls per day across all users, (b) Cumulative distribution of users with respect to the average number of phone calls per day.

the participants made less than 4 phone calls per day during the study period (i.e., 1 month). Figure 4 shows the statistics for SMS data including the number of sent and received text messages. The histogram for the SMS data appears to be right-skewed as most participants (91%) sent and received 50-100 SMS messages per day. One of the participants received and sent 150-200 SMS messages per day and another participant received and sent 350-400 SMS messages per day. Figure 5 shows the number of distinct locations visited by users during the study period. The histogram appears to be centered and closer to normal as shown in Figure 5(a). The cumulative distribution of participants with respect to the average number of distinct locations per day is shown in 5(b). Most participants visited 2-6 unique locations per day.

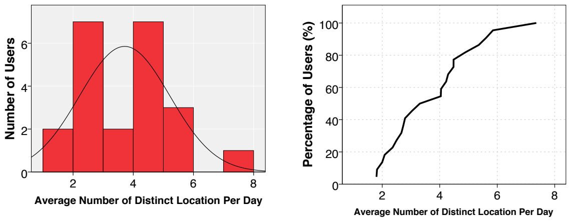
Intuitively, as users send/receive a large number of SMS messages per day, we expected that SMS related questions will be the hardest to answer, and the call and location-based questions will be comparatively easier to answer. Also, we expected hints to improve performance of legitimate users while having minimal/no effect on adversarial users' accuracy.

To explore the effects of different factors (e.g., question type, hint vs. no hint) on response correctness for different categories of users (e.g., legitimate vs. adversarial users), we used a Mixed-effect logistic regression model [17] to analyze the data. The Mixed-effect logistic regression model contains fixed effects and random effects. As we have repeated measurements from the same individual, a *user* was included as a random-effect variable to account for multiple measurements (i.e., multiple responses) within users. Thus, each user has his/her baseline likelihood for answering a question correctly. All other independent variables were included as fixed-effect variables. The coefficients listed in Table 6 show the relationships between the dependent variable (i.e. response correctness) and independent variables (e.g., question type, time taken to answer). The categorical variables are designated using their baselines. The coefficients marked with the "*" represent the variables that have statistically significant (p <0.05) effect on response correctness. The log odds was used as a measure of association between the dependent variable and independent variables and their influencing factors. The coefficients listed in Table 6 represent the change in response correctness when the coefficient



(a) (b)

Figure 4: (a) Histogram of average number of SMS messages per day across all users, (b) Cumulative distribution of users with respect to the average number of SMS messages per day.



(a) (b)

Figure 5: (a) Histogram of average number of distinct locations per day across all users, (b) Cumulative distribution of users with respect to the average number of distinct locations per day.

Features	Coefficients			Baseline
	Legitimate	Strong Adversary	Naive Adversary	
Age	0.0297317	-0.3907673	-0.0404524	
Gender	0.1075227	-0.1329869	0.6497308	Female
Time to answer (seconds)	0.0014976	0.0066606	-0.0030984	
Question Type: Call	1.379083 *	0.7397752 *	1.354868	Question Type: SMS
Question Type: Location	0.6789546 *	1.503114 *	2.878599 *	Question Type: SMS
Hint Used	0.2568961 *	-0.8211419 *	-0.5674617	Hint Not Used
Confidence	0.6430592 *	0.5549432 *	0.112277	

Table 6: Coefficients for the Mixed-Effect Logistic Regression model for the user study. The coefficients show whether listed features had a statistically significant effect on response correctness. Significant features are designated by a * next to their coefficients.

is increased by one-unit while controlling all other numerical variables at their mean values and categorical variables at their baseline. In addition, a positive coefficient implies that an increase in the independent variable affect response correctness positively. A negative coefficient suggests the opposite. We discuss the details regarding our findings below.

• **Effect of Question Type on Accuracy Score.**
In our evaluation, we condition a question type on its baseline as shown in Table 6. This indicates that the coefficient for one question type (e.g., Location questions) significantly differs from the baseline (i.e., SMS message questions). The response correctness of phone calls and location questions significantly differ from the SMS message questions for legitimate and strong adversarial users. For naive adversarial users, only location questions appeared to significantly differ from the SMS message questions. This may be due to the fact that naive adversarial users had a higher chance of guessing where a random person was at any given time as they knew that the other participants were from the same campus/locality, which is less likely to be the case in real-life. Hence, in real-life, the success rate for naive adversary is more likely to be lower for location-based questions. However, it was very difficult for a naive adversary to guess who the person texted or called at any given time.
We found that phone call questions were answered more often than SMS and location based questions. When it comes to guessing, location based questions were guessed more easily than questions about communications (i.e., phone call and SMS message).

• **Effect of Hints on Accuracy Score.**
As one of the main purposes of this paper is to evaluate the effectiveness of using hints for autobiographical authentication, we devised a within-subject user study with two conditions. In order to have consistent comparisons between questions with hint(s) and with no hint, a question was asked twice. In the first condition, a question was presented without hint and subsequently, in the second condition, the same question was presented with hint. In all cases, no feedback was given to participants to avoid biasing them.
In our evaluation, we compared different users' accuracy to evaluate whether hints help users to recall the answer. We found a significant difference between questions with hint and without hint in terms of response correctness. More interestingly, when legitimate users used hints, their response correctness improved significantly, whereas when strong adversarial users used hints, their response correctness reduced significantly, and hints had no significant effect on response correctness for naive adversarial users. While such negative effect on strong adversarial users could be due to increased ambiguities caused by hints, further investigation focusing on this particular aspect of our finding is needed to identify the underlying reasons behind such effect.
Furthermore, in our study, we found that strong adversaries used hints for 35 call, 100 SMS, and 73 location questions out of 260 call, 476 SMS, and 440 location questions respectively. Usage of hints reduced response accuracy for 84.6% adversarial users for call questions, 94% adversarial users for SMS questions, and 87.5% adversarial users for location questions. In contrast, in case of legitimate users, it improved response accuracy for 84.2% users for call question, 91% users for SMS questions, and 77.5% users for location questions. Figure 6 shows the effect of hints on overall average response correctness across three different question types (i.e., Call, SMS, and Location) for three different types of users (i.e., Legitimate, Strong, and Naive adversarial users).

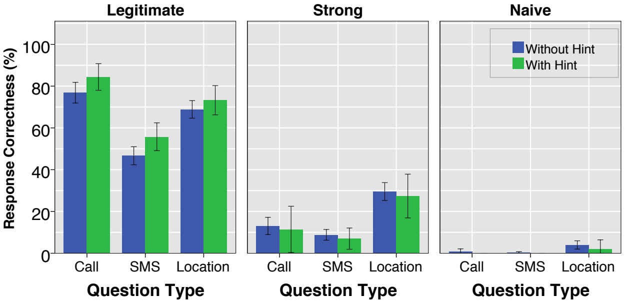


Figure 6: Comparison of response correctness rate of hint and non-hint conditions across three different question types (i.e., Call, SMS and Location) for three different user types (i.e., Legitimate, Strong and Naive adversarial users). 95% confidence intervals are included.

• **Effect of Hints on Users' Confidence.**
To understand the effect of hints on users' level of confidence, which may passively indicate whether users find hint to be useful or not, after answering a question, users (i.e., both legitimate and adversarial users) were asked to rate their level of confidence in their answer on a 5-point Likert scale where 1 means "Not confident at all" and 5 means "Very confident". To analyze the effect of hints on user's level of confidence, we used Wilcoxon signed-rank test to evaluate the difference in confidence level for their answers. The Wilcoxon test is similar to t-tests, except it does not make any assumption regarding the distributions of the compared samples, which is appropriate for our analysis. Our analysis shows that the use of hints significantly increases users' confidence on their answer for questions about communication (i.e., Phone Call and SMS message). Specifically, for phone call and SMS message based questions, we found that legitimate users were significantly more confident in the correctness of their answers when hints are used with ($Z = -5.986$, $p < 0.01$) and ($Z = -3.313$, $p = 0.01$) respectively. However, in the case of location based questions, legitimate users were not statistically more confident when they used hints ($Z = -0.741$, $p = 0.459$). Figure 7 shows the effect of hints on legitimate users' level of confidence for 3 different question types (i.e., Call, SMS, and Location) when they used hints.

• **Relation Between "Time taken to Answer" and Accuracy Score.**
"Time taken to Answer" indicates the amount of time that was taken by a user to answer a question. The effect of time on accuracy score was insignificant for legitimate and adversarial users. Furthermore, we observed that adversarial users (i.e., strong and naive adversaries) took less time on average to answer the questions compared to legitimate users. Also, when users used hints, the amount of time to answer the questions were longer as expected. Table 7 summarizes the time taken by legitimate users to answer different types of questions. Specifically, legitimate users took on average 15.50 seconds with a median of 11 seconds to answer phone call questions with no hint compared to 22.45 seconds with

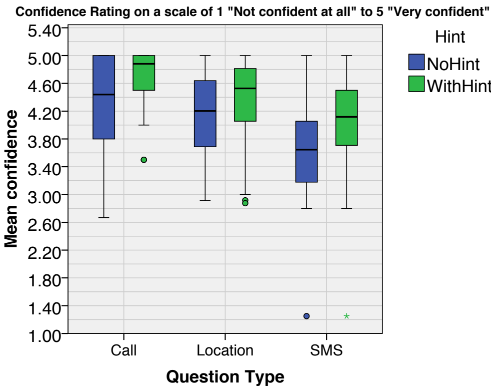


Figure 7: Impact of Hints on Legitimate User's Confidence rating while answering questions with hint and without hint.

a median of 18 seconds with hints, 18.15 seconds with a median of 12 seconds to answer SMS based questions with no hint compared to 29.34 seconds with a median of 25 seconds with hints, and 28.67 seconds with a median of 12 seconds to answer location based questions with no hint compared to 30.21 seconds with a median of 25 seconds with hints. The mean time taken to answer phone call and SMS based questions varied significantly when users used hints. However, the differences were not significant when users used hints for location based questions. Part of the reason may be due to the fact that the number of hints generated for location based questions were less compared to call and SMS based questions. Hence, users did not have to spend significant amount of time to go over the hints.

	Without Hint			**With Hint**		
	Mean	Median	SD	Mean	Median	SD
Call	15.50	11	13.24	22.45	18	15.18
SMS	18.15	12	18.03	29.34	25	17.90
Location	28.67	23	19.12	30.21	25	17.57

Table 7: Time taken for legitimate users to answer different types of questions.

- **Effect of Age and Gender on Accuracy Score.**

While due to the limited size of the study, the effect of age and gender on response correctness cannot be claimed with high confidence, in our study, gender and age had no significant effect in predicting response correctness. Part of the reason may be because participants were undergraduate students with similar age (between 18-23) and we also had an almost equally balanced male and female population. A large-scale study is needed to verify the effect of age and gender on response accuracy.

4.1 Accuracy of Model-based Authentication

As each individual user is different and may perform differently, we attempt to account for this variations by building models for each user based on individual response patterns, and subsequently leverage that model to identify legitimate users. In our work, we first present a simple threshold based scheme, and next compare that against the performance of a more sophisticated Bayesian classifier based model which is inspired based on prior work [13].

4.1.1 Classification Performance Evaluation Metrics

ROC (Receiver operating characteristics) plots are commonly used for evaluating classification performance, in which TPR (true positives rate) on Y-axis is plotted as a function of FPR (false positives rate) on X-axis, and shows the tradeoff between TPR and FPR for all possible thresholds (i.e., cut-off points) [15]. In this context, the true positive rate (TPR) corresponds to the success rate of legitimate users, while the false positive rate (FPR) denotes the success rate of adversaries. We also use the area under the ROC curve (AUC) to measure the performance of the test. Note that the performance of the test can be quantified with a single value by calculating AUC value [20] which is an important indicator of the classification performance. AUC = 0.5 represents a test performed at chance for binary classification (i.e., the model performs no better than a coin flip), while AUC = 1 means a perfect test where all legitimate users succeed and all adversaries failed to enter the system. Hence, the larger the AUC value, the better the model/test. Please note that, while evaluating performance of both threshold and Bayesian classifier based model, we use AUC value for different attack scenarios and vary the number of questions using the data collected from our field study where users are presented hints. The details are below.

4.1.2 Classification Accuracy of Threshold Based Scheme

As a single question may not be enough for reliably authenticating a user, we assume that multiple questions may be asked in a single session. Hence, in this scheme, we calculate the score of a user by taking average accuracy over multiple challenge questions in a session.

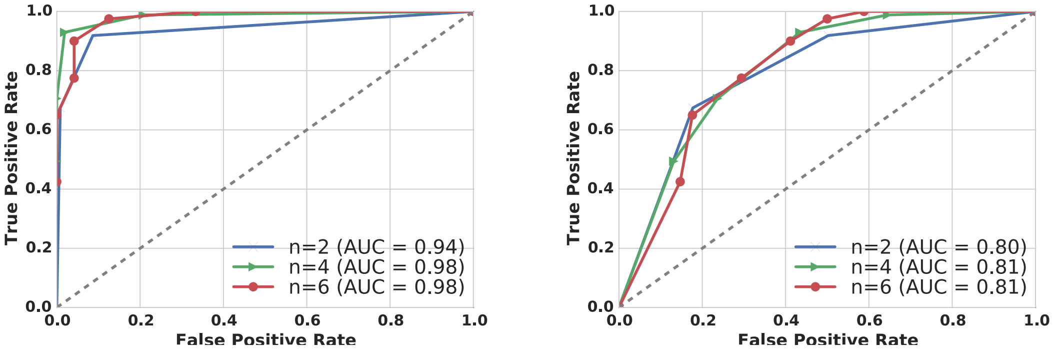
We generate ROC curves for the threshold-based scheme to show how the number of questions would affect TPR and FPR in identifying users for two different attack scenarios. In particular, in the first scenario, we assume the existence of only strong adversaries in the system (i.e., all attackers are strong adversaries), while in the second scenario, we assume the existence of only naive adversaries in the system (i.e., all attackers are naive adversaries).

The three curves in each plot in Figure 8 are generated for different number of questions (n). For brevity, we only show the ROC curves for $n = 2$, $n = 4$, and $n = 6$. From these figures, it can be seen that the AUC values are 0.94 for $n = 2$, 0.98 for $n = 4$, 0.98 for $n = 6$ when modeled against naive adversaries. In contrast, the AUC values are 0.80 for $n = 2$, 0.81 for $n = 4$, and 0.81 for $n = 6$ when modeled against strong adversaries.

Although the performance is better when modeled against naive adversary compared to strong adversary (Figure 8(a)), the performance of threshold based scheme against strong adversary is not impressive, even when the number of questions in a session increases. This motivates us to explore the Bayesian classifier based scheme which is explained next.

4.1.3 Classification Accuracy of Bayesian Classifier Based Scheme

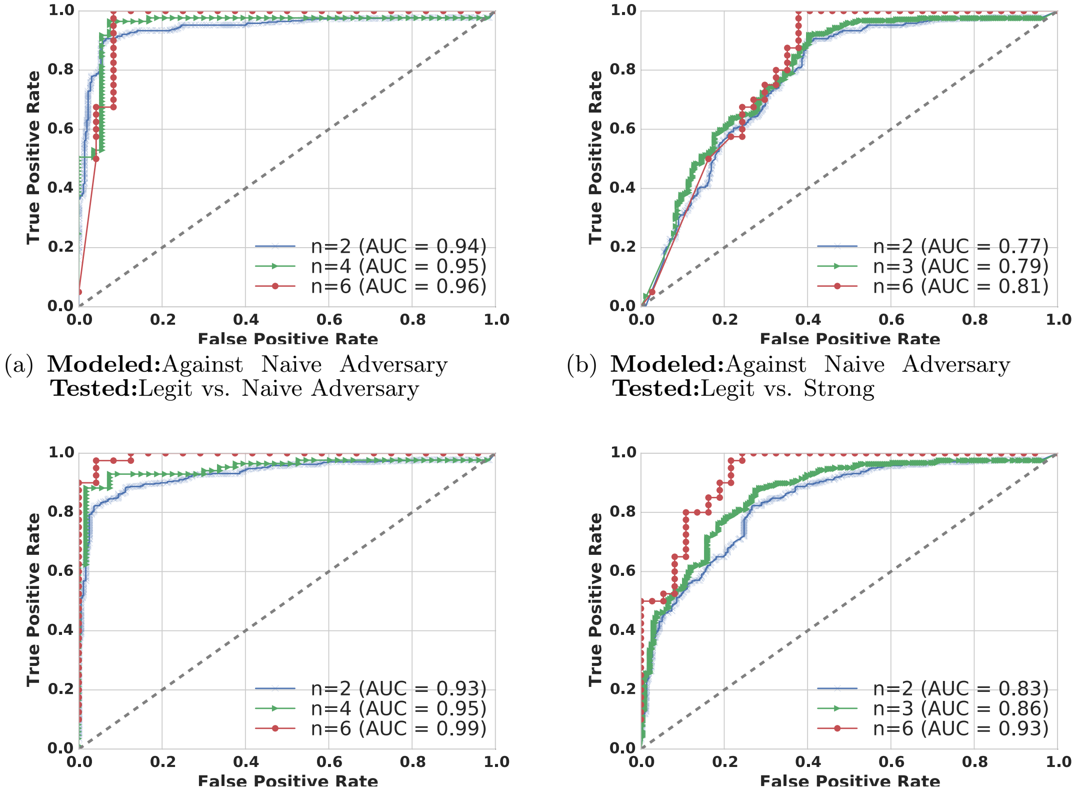
As different user's performance vary significantly, instead of relying solely on user's accuracy score (e.g., threshold-based scheme), one possible alternative is to learn a user's response pattern and subsequently leverage the response patterns along with accuracy score to authenticate a user.



(a) Against Naive Adversary

(b) Against Strong Adversary

Figure 8: Receiver Operating Characteristic (ROC) curves for threshold based scheme when modeled against strong and naive adversary for different number of questions ($n = 2$, $n = 4$ and $n = 6$).



(a) **Modeled:**Against Naive Adversary **Tested:**Legit vs. Naive Adversary

(b) **Modeled:**Against Naive Adversary **Tested:**Legit vs. Strong

(c) **Modeled:**Against Strong Adversary **Tested:**Legit vs. Naive

(d) **Modeled:**Against Strong Adversary **Tested:**Legit vs. Strong

Figure 9: Receiver Operating Characteristic (ROC) curves for Bayesian classifier based scheme when modeled against strong and naive adversary for different number of questions (e.g., $n = 2$, $n = 4$ and $n = 6$). ROC curves are plotted for different user types (legitimate user, strong adversary, and naive adversary) and attack scenarios.

For example, a user who usually answers call and location questions correctly but SMS questions incorrectly is more likely to answer call and location questions correctly and SMS questions incorrectly in future attempts (i.e., repeat a similar pattern). Using this scheme, even if an adversary can somehow observe and learn a user's daily activities and answers all the questions correctly, the adversary will require to closely imitate the response errors and behavior of a legitimate user to gain access to the system. Based on this observation, in our work, we use the Bayesian classifier based model from Das et.al paper [13] to authenticate users reliably. Please note that, in [13], authors applied the model only with two features which are categorical variables (i.e., question type and answer selection method (e.g., multiple choice vs. open ended)). We extended this prior approach and applied this model using different features such as question type, hint used, amount of time to answer, and user's level of confidence regarding the correctness of their answers. In our study, question type and hint used are categorical variables, and the amount of time to answer and user's confidence regrading their answers are continuous features. Here, if the response feature is a categorical variable such as questions type, we compute its probability by using its contingency table. If the response feature is a continuous variable such as the amount of time taken to answer a question, we compute its probability using probability density function assuming that the continues variable is distributed according to Gaussian distribution [23] as follows:

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2} \qquad (1)$$

where μ is the mean of the samples and σ is the variance of the samples. Given the μ and σ, equation 1 returns the probability of a particular sample belonging to the particular distribution.

Briefly, the Bayesian classifier based model [13] takes in a sequence of responses along with an "adversary model" as input (for simplicity, we pick a specific adversary type (e.g., strong, naive adversary)), and outputs a confidence rating ranging from 0 to 100 which represents how confident the system is in identifying the user as a legitimate user based on the user's response pattern.

Model Evaluation

To explore how the confidence ratings of the Bayesian classifier based model vary for different types of users and for different attack scenarios, we used the data collected from our field study where three different user types (i.e., legitimate, strong, and naive adversarial users) were simulated in each session. We then split each of these three different users' responses into n folds where n denotes the number of sessions. Subsequently, we use data from $(n-1)$ sessions to train the model and use the remaining session for testing. We repeat the process n times where each time we use a different session for testing.

To test the system, we try two different attack scenarios (against strong adversaries and against naive adversaries) as follows.

In the first case, we construct two models: one model represents the legitimate user which is trained using the legitimate user's data excluding the present session's data. The second model represents the strong adversary which is trained using the corresponding legitimate user's strong adversary data excluding the present session's data. For testing, we use the corresponding legitimate user's naive adversary data excluding the present session's data. Once the models are constructed, we calculate confidence rating for this scenario by using the remaining session for testing for three different user types.

Likewise, in the second case, we construct two models one using the legitimate user's data and the other using the corresponding legitimate user's naive adversary data. For testing, the corresponding legitimate user's strong adversary data is used. As before, the present session's data is excluded from the training data. We vary the number of challenge questions that we consider in each session. We also utilized combinations of different response features and we observed different confidence ratings for different response features after answering n questions aggregated across all sessions. For brevity, we only show plots for $n = 2$, $n = 4$, and $n = 6$ where we obtain the highest classification accuracy by using two features–question type and hint used.

Figure 9 shows the ROC curves for the Bayesian classifier-based scheme for two different attack scenarios. Specifically, Figure 9(a) and Figure 9(b) show the ROC curves when modeled against naive adversary (i.e., two models are trained–one for the legitimate user and one for the naive adversarial user) and tested using legitimate, naive, and strong adversarial user's data respectively. Similarly, Figure 9(c) and Figure 9(d) show the ROC curves for naive and strong adversarial users when modeled against strong adversary. As before, performance of each scenario/test is provided using the AUC value for $n = 2$, $n = 4$, and $n = 6$.

From these figures, for all attack scenarios, we observe that, as users answer more questions, regardless of the modeled adversary, the confidence estimate increases along with the AUC values. In other words, the system becomes more confident in identifying the actual user as the number of questions answered in a session increases. Also, we see that the classification performance varied greatly depending on the modeled adversary. For example, the highest classification performance was obtained against naive adversary. The AUC values are 0.96 and 0.99 for $n = 6$ when modeling against naive and strong adversary respectively. On the other hand, the classifier offers a much more conservative classification performance when tested against strong adversary. Specifically, AUC values are 0.93 and 0.81 after answering 6 questions when tested against strong adversary and modeled against naive and strong adversary respectively. Intuitively, as a strong adversary has significant knowledge regarding a user's schedule (e.g., girlfriend), strong adversarial users are more likely to gain access to the system by answering questions more accurately compared to naive adversarial users.

Please note that, while the Bayesian classifier based model achieves high accuracy, the model requires training data for both legitimate user and his/her adversarial users, which may not be available in real-life. Investigating alternative models that can be trained using a group of adversarial users' data that do not include any specific adversary, which is more likely to be available, is one of our future work.

4.2 Usability Aspect Regarding Autobiographical Authentication and Usage of Hints: A User's Perspective

To understand the impact of providing hints on usability of such systems, at the end of the study, the participants were asked to complete an exit survey for an additional $10 Amazon Gift card. We asked the participants to answer several questions to understand their perception regarding *Autobiographical Authentication* and the effectiveness of using hints. We used a five-point Likert-scale where 1 indicates strong disagreement and 5 indicates strong agreement with the given statement. Table 8 summarizes the survey results. As the survey responses are ordinal data, it is appropriate to employ median and mode rather than mean and standard deviation. Also, in case where multiple modes exist, the smallest value is shown in Table 8.

As per the survey data, most of the participants found phone

	Call		SMS		Location	
Question	**Mode**	**Median**	**Mode**	**Median**	**Mode**	**Median**
Q_1: It was easy for me to recall	4	4	4	3.5	5	4
Q_2: It was easy for my close friends to guess	3	3	3	3	3	3
Q_3: It was easy for me to guess my close friends' questions	3	2.5	1	2	3	3
Q_4: It was easy for a stranger to guess	1	1	1	1	1	1
Q_5: It was easy for me to guess stranger's questions	1	1	1	1	1	1
Q_6: I found hints to be useful while answering	4	3.5	4	3	4	3.5
Q_7: I found hints to be useful while guessing my close friend's questions	1	2	1	1.5	1	2
Q_8: I found hints to be useful while guessing a stranger's questions	1	1	1	1	1	1
Q_9: I do not think hints generated based on my data can leak my privacy sensitive information if shown to my close friend	4	4	5	4	4	4
Q_{10}: I do not think hints generated based on my data can leak my privacy sensitive information if shown to a stranger	5	5	5	5	5	4

Table 8: User Feedback on Autobiographical Authentication scheme and effectiveness of using hints. A five-point Likert-scale was used on a scale of 1 (strong disagreement) to 5 (strong agreement) with the given statement.

call and location based questions to be easier to recall (mode 4, median 4 and mode 5, median 4 respectively) compared to SMS based questions (mode 4, median 3.5 in Q_1). When it comes to guessability, most of the participants disagreed that guessing the answers of their close friend's question would be easy. They reported that questions about communications (i.e., Call and SMS) would be harder to guess compared to location based questions (with mode 1 and median 2 for SMS based questions, mode 3 and median 2.5 for call questions, mode 3 and median 3 for location questions in Q_3). A majority of the participants strongly agreed that a stranger would not be able to guess their questions and vice versa (mean and median values are all 1 for these cases). When it comes to the effectiveness of using hints, the majority of participants found hints to be useful while answering communication and location based questions (for call and location based questions mode 4, median 3.5, for SMS based questions mode 4, median 3 in Q_6). Moreover, a majority of the participants disagreed that hints would be helpful while guessing (mean and median values are all less than 2 for these cases). Finally, most of the participants did not think that hints generated based on their data can leak their privacy sensitive information if shown to their close friends (response mean and median values are all greater than or equal to 4 in Q_9), or if shown to a stranger (response mean and median values are 5, except for location question with a mode 5 and median 4 in Q_{10}).

5. DISCUSSION

In this paper, we investigated the strengths and weaknesses of providing hints for different categories of autobiographical questions and users (e.g., legitimate, strong, and naive adversarial users). We also presented a Bayesian classifier based model to account for differences in individual user's response pattern and subsequently leveraged that to identify legitimate users with high accuracy while reducing the success rate of adversaries. Based on our findings, we strongly believe that the proposed authentication framework can significantly improve the overall system security by adding another layer of easy-to-use authentication mechanism. For instance, the presented scheme can be used in addition to password based mechanism to prevent attacks launched from remote locations using stolen passwords, or can be used to facilitate resetting of passwords.

While we found that providing hints improved legitimate users' response accuracy while having no significant effect on naive adversarial users' performance, interestingly, hints appear to have negative effect on response correctness for strong adversarial users. While such negative effect could be due to increased ambiguities caused by hints, further investigation focusing on this particular aspect of our finding is needed to identify the underlying reasons behind such effect.

Finally, we would like to point out that, in our study, all of the participants ($n = 24$) were undergraduate students with similar age (range is 18-23). Due to the limited size of the study and skewed age distributions of the participants, the effect of age and gender on response correctness cannot be claimed with high confidence, and needs further investigation. Also, the effect of smartphone usage behavior (e.g., low vs. heavy smartphone users) on response accuracy cannot be verified due to limited size of the study, and needs further investigation.

6. CONCLUSION

This paper investigates the possibility of using hints to improve users' recall rate for autobiographical based authentication systems. To evaluate the effect of hints on legitimate and adversarial users' performance, a real-life user study is conducted with 24 users over a period of 30 days. The findings suggest that hint indeed has a significant positive effect on legitimate users' response correctness while negative effect on strong adversarial users' response correctness, and no significant effect on response correctness for naive adversarial users. Finally, based on exit survey data, it was noted that users feel positively about using hints while answering challenge questions.

7. ACKNOWLEDGMENTS

This work is supported by the National Science Foundation under Grant No. CNS-1251962. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the funding agency. We thank the anonymous reviewers for their valuable suggestions and comments.

8. REFERENCES

[1] Android fused location provider api. https://developers.google.com/android/reference/com/google/android/gms/location/FusedLocationProviderApi (Accessed: 06/11/2015).

[2] Android's activity recognition api: Recognizing the user's current activity. https://developers.google.com/android/reference/com/google/android/gms/location/ActivityRecognitionApi (Accessed: 06/11/2015).

[3] Google maps android api. https://developers.google.com/maps/documentation/android/ (Accessed: 06/11/2015).

[4] Y. Albayram, M. M. H. Khan, A. Bamis, S. Kentros, N. Nguyen, and R. Jiang. A location-based authentication system leveraging smartphones. In *Mobile Data Management (MDM), 2014 IEEE 15th International Conference on*, volume 1, pages 83–88. IEEE, 2014.

[5] F. Asgharpour and M. Jakobsson. Adaptive challenge questions algorithm in password reset/recovery. *First International Workshop on Security for Spontaneous Interaction (IWIISI '07), Innsbruck, Austria, (2007)*, 7:6 pages, 2007.

[6] J. Bonneau, E. Bursztein, I. Caron, R. Jackson, and M. Williamson. Secrets, lies, and account recovery: Lessons from the use of personal knowledge questions at google. In *Proceedings of the 24th International Conference on World Wide Web*, WWW '15, pages 141–150, Republic and Canton of Geneva, Switzerland, 2015. International World Wide Web Conferences Steering Committee.

[7] Y. Chen and D. Liginlal. Bayesian networks for knowledge-based authentication. *IEEE Trans. on Knowl. and Data Eng.*, 19(5):695–710, May 2007.

[8] Y. Chen and D. Liginlal. A maximum entropy approach to feature selection in knowledge-based authentication. *Decision support systems*, 46(1):388–398, 2008.

[9] S. Chokhani. Knowledge based authentication (kba) metrics. In *KBA Symposium-Knowledge Based Authentication: Is It Quantifiable*, 2004.

[10] M. A. Conway. Episodic memories. *Neuropsychologia*, 47(11):2305–2313, 2009.

[11] M. A. Conway, D. C. Rubin, A. Collins, S. Gathercole, M. Conway, and P. Morris. The structure of autobiographical memory. *Theories of memory*, pages 103–137, 1993.

[12] S. K. Dandapat, S. Pradhan, B. Mitra, R. Roy Choudhury, and N. Ganguly. Activpass: Your daily activity is your password. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, CHI '15, pages 2325–2334, New York, NY, USA, 2015. ACM.

[13] S. Das, E. Hayashi, and J. I. Hong. Exploring capturable everyday memory for autobiographical authentication. In *Proceedings of the 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, UbiComp '13, pages 211–220, New York, NY, USA, 2013. ACM.

[14] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu. A density-based algorithm for discovering clusters in large spatial databases with noise. In *KDD*, volume 96, pages 226–231, 1996.

[15] T. Fawcett. An introduction to roc analysis. *Pattern recognition letters*, 27(8):861–874, 2006.

[16] P. Gupta, T. K. Wee, N. Ramasubbu, D. Lo, D. Gao, and R. K. Balan. Human: Creating memorable fingerprints of mobile users. In *Pervasive Computing and Communications Workshops (PERCOM Workshops), 2012 IEEE International Conference on*, pages 479–482. IEEE, 2012.

[17] L. C. Hamilton. *Statistics with STATA: Version 12*. Duxbury Press, Boston, MA, USA, 8 edition, 4 2012.

[18] A. Hang, A. De Luca, and H. Hussmann. I know what you did last week! do you?: Dynamic security questions for fallback authentication on smartphones. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, CHI '15, pages 1383–1392. ACM, 2015.

[19] E. Hassan. Recall bias can be a threat to retrospective and prospective research designs. *The Internet Journal of Epidemiology*, 3(2):339–412, 2006.

[20] T. Hastie, R. Tibshirani, J. Friedman, and J. Franklin. The elements of statistical learning: data mining, inference and prediction. *The Mathematical Intelligencer*, 27(2):83–85, 2005.

[21] M. Hertzum. Minimal-feedback hints for remembering passwords. *interactions*, 13(3):38–40, 2006.

[22] M. Jakobsson. *The death of the Internet*. John Wiley & Sons, 2012.

[23] G. H. John and P. Langley. Estimating continuous distributions in bayesian classifiers. In *Proceedings of the Eleventh conference on Uncertainty in artificial intelligence*, pages 338–345. Morgan Kaufmann Publishers Inc., 1995.

[24] G. Kristo, S. M. Janssen, and J. M. Murre. Retention of autobiographical memories: An internet-based diary study. *Memory*, 17(8):816–829, 2009.

[25] B. Lu and M. B. Twidale. Managing multiple passwords and multiple logins: Mifa minimal-feedback hints for remote authentication. In *IFIP INTERACT 2003 Conference*, pages 821–824, 2003.

[26] M. Nishigaki and M. Koike. A user authentication based on personal history-a user authentication system using e-mail history. *The Journal on Systemics, Cybernetics and Informatics*, 5(2):18–23, 2007.

[27] A. Nosseir, R. Connor, and M. Dunlop. Internet authentication based on personal history-a feasibility test. 2005.

[28] A. Nosseir and S. Terzis. A study in authentication via electronic personal history questions. In *ICEIS 2010 - Proceedings of the 12th International Conference on Enterprise Information Systems, Volume 5, HCI, Funchal, Madeira, Portugal, June 8 - 12, 2010*, volume 5, pages 63–70, 2010.

[29] L. O'Gorman, A. Bagga, and J. Bentley. Call center customer verification by query-directed passwords. In *Financial Cryptography*, pages 54–67. Springer, 2004.

[30] J. Podd, J. Bunnell, and R. Henderson. Cost-effective computer security: Cognitive and associative passwords. In *Computer-Human Interaction, 1996. Proceedings., Sixth Australian Conference on*, pages

304–305. IEEE, 1996.

[31] A. Rabkin. Personal knowledge questions for fallback authentication: Security questions in the era of facebook. In *Proceedings of the 4th Symposium on Usable Privacy and Security*, SOUPS '08, pages 13–23, New York, NY, USA, 2008. ACM.

[32] K. Renaud, T. McBryan, and P. Siebert. Password cueing with cue (ink) blots.

[33] S. Schechter, A. B. Brush, and S. Egelman. It's no secret. measuring the security and reliability of authentication via "secret" questions. In *Proceedings of the 2009 30th IEEE Symposium on Security and Privacy*, SP '09, pages 375–390, Washington, DC, USA, 2009. IEEE Computer Society.

[34] S. Schechter, S. Egelman, and R. W. Reeder. It's not what you know, but who you know: A social approach to last-resort authentication. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '09, pages 1983–1992. ACM, 2009.

[35] S. Schechter and R. W. Reeder. 1 + 1 = you: Measuring the comprehensibility of metaphors for configuring backup authentication. In *Proceedings of the 5th Symposium on Usable Privacy and Security*, SOUPS '09, page 9, New York, NY, USA, 2009. ACM.

[36] J. Thorpe, B. MacRae, and A. Salehi-Abari. Usability and security evaluation of geopass: A geographic location-password scheme. In *Proceedings of the Ninth Symposium on Usable Privacy and Security*, SOUPS '13, page 14, New York, NY, USA, 2013. ACM.

[37] S. Vemuri and W. Bender. Next-generation personal memory aids. *BT Technology Journal*, 22(4):125–138, 2004.

[38] W. A. Wagenaar. My memory: A study of autobiographical memory over six years. *Cognitive psychology*, 18(2):225–252, 1986.

[39] W. E. Winkler. String comparator metrics and enhanced decision rules in the fellegi-sunter model of record linkage. 1990.

[40] K. Xu, D. Yao, M. A. Pérez-Quinones, C. Link, and E. Scott Geller. Role-playing game for studying user behaviors in security: A case study on email secrecy. In *Collaborative Computing: Networking, Applications and Worksharing (CollaborateCom), 2014 International Conference on*, CollaborateCom '14, pages 18–26. IEEE, 2014.

[41] M. Zviran and W. J. Haga. User authentication by cognitive passwords: An empirical assessment. In *Information Technology, 1990.'Next Decade in Information Technology', Proceedings of the 5th Jerusalem Conference on (Cat. No. 90TH0326-9)*, JCIT, pages 137–144, Los Alamitos, CA, USA, 1990. IEEE Computer Society Press.

Usability and Security Perceptions of Implicit Authentication: Convenient, Secure, Sometimes Annoying

Hassan Khan, Urs Hengartner, and Daniel Vogel
Cheriton School of Computer Science
University of Waterloo
Waterloo, ON Canada
{h37khan,urs.hengartner,dvogel}@uwaterloo.ca

ABSTRACT

Implicit authentication (IA) uses behavioural biometrics to provide continuous authentication on smartphones. IA has been advocated as more usable when compared to traditional explicit authentication schemes, albeit with some security limitations. Consequently researchers have proposed that IA provides a middle-ground for people who do not use traditional authentication due to its usability limitations or as a second line of defence for users who already use authentication. However, there is a lack of empirical evidence that establishes the usability superiority of IA and its security perceptions. We report on the first extensive two-part study (n = 37) consisting of a controlled lab experiment and a field study to gain insights into usability and security perceptions of IA. Our findings indicate that 91% of participants found IA to be convenient (26% more than the explicit authentication schemes tested) and 81% perceived the provided level of protection to be satisfactory. While this is encouraging, false rejects with IA were a source of annoyance for 35% of the participants and false accepts and detection delay were prime security concerns for 27% and 22% of the participants, respectively. We point out these and other barriers to the adoption of IA and suggest directions to overcome them.

1. INTRODUCTION

Recent studies on locking behaviour of smartphone users have shown that 40% or more users did not use any authentication mechanism on their devices [8, 15, 20]. Furthermore, participants of these studies most commonly cited "inconvenience" as the reason for not using any authentication mechanism on their devices [8, 15]. The traditional explicit authentication (EA) mechanisms (such as PIN, pattern lock, facial and fingerprint recognition) provide all-or-nothing access control. However, smartphone sessions are short and frequent, and PIN entry for every short session is inconvenient for users and unnecessary for some situations [16]. In a field study, Harbach et al. [15] demonstrated that smartphone users considered unlock screens unnecessary in 24% of the situations and they spent up to 9% of the time they use their smartphones to deal with the unlock screens.

To address these usability issues with EA, researchers have proposed implicit authentication[1] (IA). IA employs behavioural biometrics to continuously and transparently recognize and validate the identity of smartphone users. Some of the recent touch behaviour-based IA schemes provide high accuracy rates (≥95%) and have gained traction in technology media news with claims like: *"Identifying someone by the way they tap and swipe on a touchscreen might be the more natural, unobtrusive future of smartphone biometrics"* [22]. Researchers have proposed to use IA as a middle ground for those smartphone users who do not configure any EA on their devices due to EA's usability issues [18, 31] or as a second line of defense in case the EA mechanism on the device is compromised [12, 19, 31].

The focus of the majority of IA research is on improving the accuracy of IA schemes with existing behavioural biometrics and by leveraging new sensors like orientation [3]. The usability evaluation of IA has largely been ignored. Existing IA literature has assumed without empirical evidence that since IA authenticates without requiring explicit input, it is more usable. For instance, when Shi et al. introduced the term IA with a behaviour-based scheme, they postulated that *"this is a meaningful approach, whether used to increase usability or increase security"* [25].

Given that IA does not require explicit input to authenticate the device user for every session, intuitively it seems that IA should reduce the amount of time spent on authentication. However, despite the reasonably high detection accuracy of some IA schemes, these schemes are still subject to false rejects, false accepts and detection delays (these terms are explained in § 2.1), which could introduce new usability issues and affect users' security perceptions. If the IA detection model is unsure about the user's identity, it naturally resorts to an interrupt-authenticate approach in which the current task is pushed to the background and the user has to explicitly authenticate to establish their identity [9, 19]. This interrupt-authenticate approach for false rejects is quite different from consistent EA authentication. It remains unclear how it affects usability in terms of annoyance and task performance in terms of time and error. Similarly, it is not obvious whether the usability-security trade-off offered by IA overcomes the perception of security given the risks of false accepts and delay in detection of an intruder.

Symposium on Usable Privacy and Security (SOUPS) 2015, July 22–24, 2015, Ottawa, Canada.

[1] The terms active authentication, continuous authentication, implicit authentication and transparent authentication have been used in the literature interchangeably.

To find answers to these questions, we conducted an ambitious user study (n=37) to empirically evaluate usability claims and security perceptions of IA. Our study uses a two-part design combining a controlled lab experiment with a field study. In a controlled lab environment, we compare IA with the preferred EA scheme of each participant. Then, a subsequent field study captured real-life experiences with IA false rejects over three days. We use a pseudo-IA scheme for both parts to simulate representative false reject rates in a controlled fashion and avoid potential confounds from idiosyncratic behaviour caused by specific IA schemes. During both parts of the study, we logged measurement data and collected feedback through surveys for a quantitative analysis and we conducted interviews for a qualitative analysis and to gather insights into subjective perceptions of IA.

In terms of usability, we show that IA decreased the overall task completion time without affecting the task error rate. While there was no statistically significant difference in the system usability scale score [4] between IA and EA, IA was favoured by 91% of the participants in terms of convenience (26% more compared to EA). However, annoyance due to false rejects is a potential issue with IA: 35% of the participants found this to be somewhat or very annoying.

In terms of security perceptions, we found that although detection delay and false accepts were serious concerns, 82% of participants were satisfied with the level of protection provided by IA. Consequently, 33% of participants were interested in using it as a primary authentication mechanism, 30% were interested in using it as a secondary authentication mechanism and another 30% wanted to try and test IA more to see if it satisfied their security needs.

Feedback gathered during the interviews further supports these findings and provides insights for how to mitigate the effect of false rejects and how to address deployment issues such as opacity of IA and operating threshold customization. We believe our findings provide important evidence and guidelines for future IA research and deployments.

2. BACKGROUND AND RELATED WORK

In this section we first provide a brief background on IA and its terminology. We then survey the related literature on usability evaluations and security perceptions of IA.

2.1 Background on Implicit Authentication

IA employs behavioural biometrics to continuously and transparently recognize and validate the identity of smartphone users. To provide IA on smartphones, many IA schemes have been proposed by the research community that rely on a diverse set of behavioural biometrics including a user's device usage patterns [25, 26], touchscreen input behaviour [12, 19, 24, 31], keystroke patterns [7, 10, 13], and gait patterns [11, 21]. IA schemes create a normal behavioural profile of the device owners using their behavioural data and it then monitors the real-time device usage to detect anomalous behaviour. The anomaly detection based origins of IA result in the same inherent limitations including: *false rejects (FR):* when an access attempt by a legitimate user is rejected due to variations in the user behaviour or imperfections in the behavioural profile; *false accepts (FA):* when an access attempt by an adversary is granted due to behavioural similarities between the adversary and the legitimate user; *delayed detection:* due to the unavailability of behavioural data, an adversary may be able to use the device for some duration before she fails authentication; and *training delay:* time spent to collect usage data in order to create the behavioural profile of the user.

Due to the security limitations of IA, researchers have proposed to use IA only as a middle-ground for the smartphone owners who do not configure any EA on their devices due to EA's usability issues [18, 31] or as a second line of defense in case the EA on the device is compromised [12, 19, 31]. For both of these deployment scenarios, researchers have suggested to interrupt and then explicitly authenticate the user (using an "administrator password") in case the behavioural profile of the current user does not match that of the device owner [9, 18, 19]. We use the term *interrupt-authenticate* to describe this phenomenon. Some related terms that we use throughout this paper are also described: a *true accept (TA)* is when an access attempt by a legitimate user is granted. A *true reject (TR)* is when an access attempt by an adversary is rejected. Finally, *operating threshold* is used to define the desired values for the negatively correlated FA and FR entries (by increasing the operating threshold, FRs can be decreased at the cost of increased FAs and vice versa).

2.2 Related Work

Usability evaluation of EA on smartphones is a well researched area [2, 15, 28, 29]. However, it is only partly related to our work since some of the evaluation metrics for EA (such as time-to-authenticate and memorability) are not applicable to IA. Another avenue partially related (due to the overlapping limitations) to our work is the usability evaluation of anomaly detection systems. However, to the best of our knowledge, the literature on the usability evaluation of anomaly detection systems only discusses challenges in determining an appropriate operating threshold and does not evaluate the security and usability perceptions due to FAs and FRs [23, 30]. Therefore, in this section, we only discuss the related work in the field of IA.

While there are dozens of papers on IA, the focus of contemporary IA research is on using novel behavioural biometrics for authentication and on improving the accuracy of IA. The usability issues surrounding IA have been ignored except for the work by Clarke et al. [5] and Crawford and Renaud [6]. Clarke et al. developed a prototype on a personal computer for an IA scheme that employed a combination of face, voice and keystroke biometrics to continuously authenticate users. They evaluated their prototype using 27 participants and found that 92% of the participants considered it more secure in comparison to the traditional forms of authentication. The participants were also asked to rate the convenience on a 5-point Likert scale and although the responses were mixed, a slight skew towards the system being convenient existed. While Clarke et al. are the only authors who provide a usability evaluation of the IA scheme that they proposed, their evaluation was limited because: (i) it was not a strictly behavioural biometric-based scheme since they used a combination of physiological (facial recognition) and behavioural biometrics (voice and keystroke data); and (ii) participants evaluated the prototype on a personal computer instead of a mobile device.

More related to our work is the recent work by Crawford and Renaud [6] in which they determine the security perceptions of IA by conducting an in-lab study with 30 participants. They provided a smartphone with a pseudo-IA scheme and asked the participants to perform tasks that required different levels of security. The participants were divided into three groups: (G1) the participants started with a low device confidence level. If a participant wanted to perform a task of medium/high security level, she may increase the device confidence by providing a matching keystroke or voice biometric or by explicitly authenticating; (G2) participants were always successfully implicitly authenticated (0% FR rate); and (G3) participants always failed implicit authentication (100% FR rate). G2 and G3 were used to get the perceptions of distrustful and frustrated participants, respectively. Crawford and Renaud found that 73% of participants felt IA was more secure than EA and 90% indicated that they would consider adopting it. While Crawford and Renaud provide the only in-depth study on the security perceptions of IA, it has some limitations including: (i) no usability evaluation is performed; (ii) annoyance due to FRs is not quantified; and (iii) security perceptions due to FAs and detection delays are not evaluated.

3. GOALS

We divide our goals to investigate IA usability and perceived security into seven questions. Later, we organize our study results around these seven questions.

Our main goal regarding IA usability was to test established usability metrics and commonly accepted usability assumptions, as captured by the following research questions:

U1 Does IA decrease the overall task completion time and the authentication overhead when compared to EA?

U2 Do the interrupt-authenticates in IA increase the error rate of the primary task?

U3 Are fewer but less predictable authentication interrupts of IA less annoying or tolerable as compared to EA or no authentication at all?

U4 Does IA score higher on the system usability scale [4] as compared to EA?

U1 and U2 address standard usability metrics for time and error as they may be affected by the interrupt-authenticate model of IA. These metrics have never been evaluated directly with IA, but they have been used to measure performance impact of similar task interruptions with personal computers [1] and they have been implied as benefits of IA in previous work [25]. By answering U3, we will test levels of annoyance caused by FRs and through U4, we test claims of higher perceived usability for IA compared to the primary authentication baselines of EA [5, 6] and no authentication.

Our main goal for the security perceptions of IA was to explore the following research questions:

S1 Are the security properties of current IA schemes (such as the FA rate) acceptable to users?

S2 Is the perception of IA security better than common current authentication schemes?

S3 Are smartphone users interested in adopting IA?

S1 has never been explored in the IA literature. S2 has been explored in previous studies [5, 6] and we attempt to validate these prior findings. In addition to evaluating the overall perceived level of security, we elicit the perceived level of security against different types of adversaries, different device states and different types of tasks. Finally, answering S3 provides an indication of IA deployment potential from a human-centric perspective since it essentially combines security perceptions and usability.

4. STUDY

We use a two-part study for our evaluations. The first part is a lab-based experiment where each participant performs simulated tasks with IA and with their current authentication scheme. This provides highly controlled, quantitative results. Measuring annoyance and other subjective feedback caused by IA interruptions is more ecologically valid when evaluated with real tasks over a longer time period, so the second part is a three-day field study where participants used IA on their own smartphone. For experimental control, both parts use a pseudo-IA scheme (described below). Our methodology was reviewed and approved by the IRB of our university.

4.1 Participants

The in-lab study was completed by 37 participants and 34 of those same participants completed the field study. Three participants dropped due to technical issues (two participants had device encryption enabled and one participant reported a broken device). We recruited these participants using multiple sources including: (i) an advertisement on Craigslist and Kijiji in November of 2014, under the "other jobs" section; and (ii) on the university-wide mailing list. The title of the advertisement was "Participate in a research study on the efficacy of a novel authentication scheme on smartphones" and it stated that the study was about the evaluation of a novel authentication scheme and adults who owned and used an Android-based smartphone for over six months could participate. Those interested were requested to fill out an online screening survey (provided in Appendix B), which collected information about their age, gender, profession, security preferences, smartphone make and model, amount of time they have used a smartphone, and email address. Participants were paid $35 ($10 for each of two in-lab sessions and $15 for the field study).

Participant demographics, current authentication schemes, and authentication preferences are summarized in Table 1. Current authentication scheme by age group is provided in Figure 1. Overall, our participant pool has good diversity by profession, age, and current authentication scheme. For our research questions, this kind of diversity is important. Similar to the prior studies [8, 15], the top reason our participants gave for not using any authentication scheme was inconvenience. Furthermore, about half of the participants who used some authentication scheme agreed that it was inconvenient or annoying at times. The annoyance was split by current authentication scheme: PIN users were significantly more likely (53% more) to find their authentication scheme inconvenient as compared to the pattern lock users (Fisher's Exact Test, $p = 0.028$).

In terms of current authentication, 14 participants used no authentication, eleven used Android's Pattern Lock, nine used a four-digit PIN and three used other schemes (two participants used a password and one participant used a longer PIN conforming to his company policy). We use the participants' current authentication scheme as an independent between-subject variable, which we refer to as *Use*. Where relevant, we summarize results using groups and subgroups formed by this variable. Specifically, *DontUseAuth* is the

n = 37	
Gender:	56% Female
	43% Male
Occupation:	32% Employed
	30% Grad student
	24% Undergrad student
	13% Unemployed/retired
IT experience:	22% Studied/worked in IT
Current authentication scheme:	38% None
	24% PIN
	30% Pattern lock
	8% Other
Sharing habits:	51% Never
	41% Rarely (once a month)
	5% Occasionally (once a week)
	3% Daily
Top reasons for not using any authentication:	8/14 It's a hassle/takes time
	5/14 Nothing to hide
	3/14 Never thought about it
Top reasons for using authentication:	19/23 Protected if lost/stolen
	18/23 Protected when unattended
	12/23 Someone casually picking it
	10/23 Unwanted disclosures
Protecting against:	18/23 Strangers
	12/23 Coworkers
	8/23 Friends/roommates
	7/23 Spouse/own children
Thoughts on authentication:	13/23 It is inconvenient sometimes
	10/23 It is easy
	3/23 It takes time

Table 1: Demographics and security preferences of the study participants

group of participants who reported that they do not use any authentication and *UseAuth* refers to the group of participants who reported that they use some EA scheme. We further separate *UseAuth* into two common EA schemes: *UsePIN* for the subgroup of *UseAuth* participants who reported using a PIN and *UsePAT* for the subgroup of *UseAuth* participants using a pattern lock.

4.2 Apparatus

We developed two Android apps, Explicit Authentication and Implicit Authentication, that executed on the devices of the participants during both parts of the study. For the lab-based experiment, the apps presented a series of tasks to the participants. The apps contained authentication screens to authenticate the participants using a PIN, Android's Pattern Lock or a six character password. We used the Android Open Source Project's UI and implementation[2] for the pattern lock and used a UI identical to that of Android for PIN and password screens. We simulated authentication on the participants' devices using our apps (explained in S 4.3.3) to accurately measure the time spent on authentication. For the field study, the Implicit Authentication app executed as a background service to simulate FRs. Although IA was simulated, participants were told that all biometric data remained solely on their device.

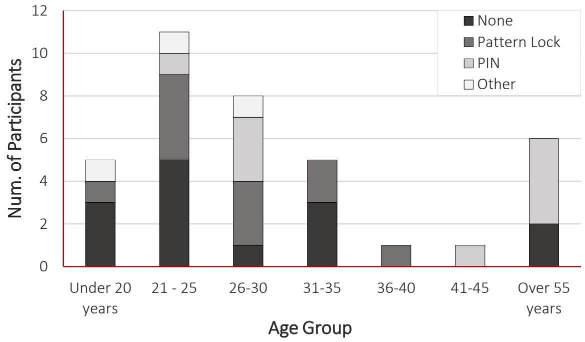


Figure 1: The "Age Group" – "Current Authentication Scheme" distribution for study participants

Deception: In both parts of the study we used a pseudo-IA scheme that ostensibly employed the touch behaviour biometric. The pseudo-IA scheme interrupt-authenticated users by triggering authentication screens during their interaction with the device to simulate FRs. Our pseudo-IA scheme was configurable to simulate different FR rates and detection delays. We used a pseudo-IA scheme because: it was not possible to intercept touch input events for IA without rooting the device due to the constraints imposed by Android [18]; and a pseudo-IA scheme enabled us to strictly control the frequency and the detection delay of FRs. Participants were told that the IA scheme on their device was fully operational and that one of the in-lab sessions served as training for the IA algorithm (details provided in § 4.3). Furthermore, to reduce the chance of participants discovering this deception by testing it with other users, we asked them not to share their devices during the field study arguing that sharing would interfere with the brief re-training phases required by our early-stage IA scheme. There are limitations of using a pseudo-IA and we discuss these limitations in § 7.

Configuration parameters selection: A challenging aspect was the selection of IA configuration parameters. By choosing different parameter values, we can change the behaviour of IA schemes. For instance, there is a negative correlation between FAs and FRs, and increasing detection delay reduces the number of FRs. While there is no recommended operating threshold, we inspected different operating thresholds because: (i) high accuracy input behaviour-based IA schemes have deployment constraints as compared to low accuracy schemes that can be readily deployed [17]; (ii) studies report a high degree of variability for FR rate between users [7]; (iii) variability in FR rate for an individual user may be caused by the type of activity that the user is performing [9]. We chose a representative and realistic FR rate range based on previous work [17]. We evaluated three FR rates: 5%, 10% and 20% that have a corresponding FA rate range between 3%-18% and 5-30 seconds of detection delay for different IA schemes. We discuss operating threshold configurations specific to each experiment in § 4.3 and § 4.4.

4.3 Part 1: Controlled Lab Experiment

The purpose of the controlled lab experiment is to: (i) introduce the participants to IA; (ii) perform an A/B testing of IA with non-IA (*UseAuth* or *DontUseAuth*); (iii) demonstrate an ostensible TR; (iv) elicit initial feedback on usability and security perceptions of IA; (v) collect data to eval-

[2] http://code.google.com/p/android-lockpattern/

uate the performance metrics of U1 and U2; and (vi) test the pseudo-IA scheme on the participants' devices without any EA scheme configured (this was not possible in the field study due to data security and privacy threats without EA).

4.3.1 *Task*

Our two apps presented a series of experiment tasks on the participant's own smartphone. In the experiment, each *task* represented a device usage session. The participant waited for the device to ring (with vibrate) to indicate it was time to perform a task. For the Explicit Authentication app, the participant turned the screen on, performed authentication if required, completed an activity, and turned the screen off. For the Implicit Authentication app, instead of the authentication at the beginning, the participant was interrupted and authenticated in the middle of an activity (frequency and timing of interrupt-authenticates are discussed in § 4.3.2). There was no time limit to complete a task. To simulate longer breaks between real device usage sessions, the participant waited a random time between 8-15 seconds before performing the next task.

We chose a subset of activities from the primary activities proposed by Bailey and Konstan [1]. We chose those activities that were abstract representations of common mobile activities and were diverse in terms of difficulty and cognitive load, enabling us to inspect error rate and interrupt-authenticate overhead. A description of these activities by increasing level of difficulty due to higher mental loads on working memory (based on the rankings of [1]) are provided below (see also screen captures in Appendix A):

- *Input Activity:* entering a sequence of characters. Our activity required participants to enter a nine digit number displayed on the screen into an input field. For the input number, we carefully chose permutations that did not overlap with the local area codes and did not have any consecutive or repeating integers. This activity is representative of common smartphone activities like entering search queries, composing texts, and entering emails.
- *Selection Activity:* selecting multiple items from a list of items. We used a list of words with selection checkboxes arranged in a 12-row x 3-column table. Thirty-six words were randomly chosen from a base set of six words. Participants had to select each word in the table that matched a given target word (taken from the base set). This activity is representative of common smartphone activities such as choosing a number to dial, choosing an app to launch or scanning the results of a search.
- *Reading Activity:* reading and comprehending information. Participants read a 7-10 sentence narrative passage from Wikipedia and then answered two multiple choice questions regarding its content. This activity is representative of common smartphone reading and comprehension activities such as reading emails or web browsing.

4.3.2 *Design*

Participants completed the lab-based experiment in two sessions held on different days. Each session lasted between 45-60 minutes including introduction, pre-survey, experiment tasks, post-survey, and interview. In each session, experimental tasks were completed under one of two within-subject conditions: the *IA* condition when they used IA and *non-IA* when they did not use IA. The order of IA and non-IA sessions was counterbalanced across participants.

Each session had 30 task trials with each task showing one activity. There were ten instances of each of the three activity types. Since the activity types were of varying difficulty level, we did not use simple random sampling to select their order since some orderings could introduce a confounding effect (e.g. the first 10 tasks are all difficult activities). Instead, we constrained the random presentation order by creating five blocks of six tasks where the tasks have two variations of each activity. We then permuted the order of tasks within each block using the first 5 rows of a 6x6 Latin square. This counterbalanced the varying difficulty levels of activity types. The same order of blocks and tasks was used across IA and non-IA sessions across all participants to create an unbiased comparison and to make sessions directly comparable.

For the non-IA session, participants were assigned the same authentication scheme that they used currently on their device (which could be an EA scheme or no authentication). For the interrupt-authenticate caused by a simulated FR in the IA session, *UseAuth* participants used their current authentication scheme while *DontUseAuth* participants were assigned a scheme that they preferred to use. Although we could have assigned a random authentication scheme to *DontUseAuth* participants, this could have introduced negative bias from dislike or inexperience with the assigned scheme. In both sessions, we were not interested in the memorability of the secret. Participants could write down their secret or reset it in between tasks if they wished.

While the input activity naturally generated tapping data, we rendered the reading activity and the selection activity in such a way that participants had to swipe to scroll to see their content. This led them to believe that their interactions were used as a biometric. We also used deception in terms of training by telling the participants who tested IA in the their first session that the data from the first few tasks was used for training. The participants who tested IA in the second session were told that the data from the first session was used for training.

We used a 20% FR rate (six interrupt-authenticates in total, twice for each type of activity) and a detection delay between 5-10 seconds. A lower detection delay (2-4 seconds) was used for the shorter input activity.

4.3.3 *Procedure and Data Collection*

The shortlisted participants were asked to bring their devices to the lab. We started the first session by showing a two minutes video introducing the apps and activities[3]. Participants were introduced to IA using a three minutes video before the IA session, which explained the operations of touch behaviour-based IA and the associated FAs, FRs and operating threshold[4]. A researcher was available during these video demonstrations to answer any questions. After the briefing, participants downloaded and installed the app for the session through Google Play Store. They were then asked to set the current authentication scheme to 'None' and turn on 'airplane mode' on their devices. This eliminated notifications or interruptions during the experimental tasks and enabled our app to control all authentications.

[3] http://youtu.be/qDQm_Oad6Pw
[4] http://youtu.be/HUR2-bxBtI8

After device setup, participants completed the STAI survey [27] to provide us with their current state of anxiety. The participants were then asked to launch the app to configure an authentication secret and then complete the main experiment tasks. After completing all tasks, they provided another measurement on anxiety by completing the STAI survey again. The participants were asked to complete a post-survey (provided in Appendix C) for the non-IA and the IA sessions. This survey consisted of 12 questions regarding usability and security perceptions of the authentication schemes that they tested. The participants used the survey to rate their perceived level of security (overall and for different adversaries, device states, and different tasks) and usability (in terms of convenience, annoyance, time consuming and tiring). Participants who tested any authentication scheme during a session were asked to complete the system usability scale (SUS) survey [4] after the session. The SUS survey was modified (provided in Appendix D) to explicitly inform the participants that it was evaluating the authentication scheme, not the apps. We also changed the word 'system' to 'method' and we dropped the question 'I found the various functions in this system were well integrated' because it was not applicable to our evaluations. Finally, a semi-structured interview (provided in Appendix E.1) of 10-15 minutes gained insight into survey answers. The interviews were recorded and later transcribed.

4.4 Part 2: Field Study

The field study was conducted after the lab study with the same participants. The main purpose of the field study was to gather realistic data on potential annoyance due to FRs. In addition, we wanted to subject participants to different operating thresholds to determine a tolerable one in terms of the frequency of FRs.

4.4.1 Task

The task for the field study was for participants to use their device as usual and handle simulated IA FRs (experienced as interrupt-authenticate screens) as they occurred. After each interruption the participant also provided brief feedback through an in-situ pop-up.

Each FR interrupted the current smartphone app with an authentication screen requiring an explicit authentication. These were the same simulated authentication screens used in the lab experiment. The background service in the Implicit Authentication app monitored two events: the `ACTION_SCREEN_ON` event to keep track of when the users turned on their screens and the `ACTION_USER_PRESENT` event to know when the users were present on their devices after dismissing the lock screen. An interrupt-authenticate was triggered after k `ACTION_USER_PRESENT` events (k was controlled during the study, details are in the Design section).

We were also interested in measuring the annoyance of each FR. After an interrupt-authenticate, we performed experience sampling with a simple in-situ feedback screen (Figure 14d). It asked the participants about their current annoyance on a 5-point Likert scale ("Very annoying" - "Not annoying at all"). The feedback screen also displayed the current operating threshold and the associated security strength of that threshold (in terms of the proportion of strangers that the IA scheme would likely protect against).

4.4.2 Design

We conducted the field study for three days to measure annoyance for different operating thresholds. FRs were simulated after every k `ACTION_USER_PRESENT` events and we randomly chose a value of k for each day to reflect high, medium and low accuracy corresponding to 20%, 10% and 5% FR rates, respectively. Since we were unable to determine when a participant interacted with the touch screen after an `ACTION_USER_PRESENT` event, we simulated a FR by choosing a random delay between 15-30 seconds. If the participants turned off the screen before the delay timeout, they were authenticated in the next session with a reduced delay. The delay value is decreased by five seconds each time down to a minimum delay value of ten seconds to ensure that the participants with short sessions also experienced FRs. We did not simulate a FR for the sessions when the call state of the device was ringing or off-hook.

We allowed participants to adjust the operating threshold if they wished. To ensure that the participants did not set the operating threshold to zero, the participants' adjusted value was only effective for 30 minutes and after that it was reset. This mitigated the possibility of participants killing the background service if interrupts became too irritating (participants felt they had some control) and gathered data to study the potential need and utility for users to control the trade-off between usability and security. The briefing video explained the trade-off when setting different threshold values. We report frequency of adjustment and discuss the need for such a control in § 6.3. We told participants that the IA scheme automatically adjusted the threshold value after brief re-training phases but they could change it depending on their desired level of protection. Participants were informed about this behaviour and that they could dismiss the interrupt-authenticate by pressing the home button but we asked them to avoid doing so except in extreme cases.

4.4.3 Procedure and Data Collection

After the second in-lab study session, participants were briefed about the background service executing on their devices and the interface of the in-situ feedback pop-up. The researcher then performed an ostensible demonstration of a true reject on their device to lead them to believing that the IA scheme was behaving as expected. After the completion of the three day usage period of the pseudo-IA scheme, the participants were instructed to contact us through emails to arrange for an in-person semi-structured interview (provided in Appendix E.2) of 10-15 minutes and to collect the remuneration. The participants were also debriefed about the deception at the end of this interview.

During the field study, we logged the `ACTION_SCREEN_ON` and `ACTION_USER_PRESENT` events. From the in-situ feedback, we logged the level of annoyance of IA and the adjusted value of operating threshold.

5. RESULTS

The quantitative and qualitative results of the controlled lab experiment and field study are presented together organized by the research questions raised in § 3. A discussion is provided after the results for each research question.

For the in-lab study, participants completed all tasks in 25 minutes on average (*median = 23, sd = 4.2*). Dur-

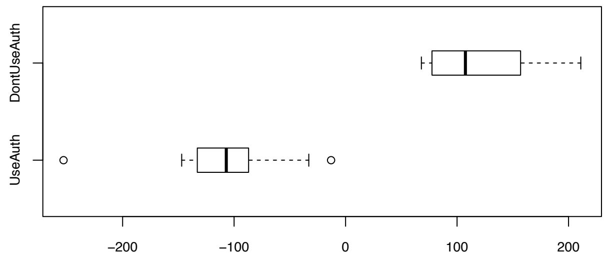


Figure 2: Change in the overall task completion time for the non-IA session as compared to the IA session

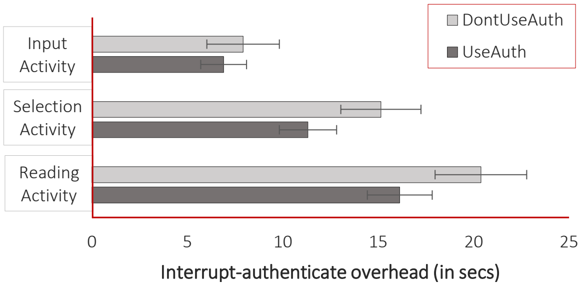


Figure 3: Interrupt-authenticate overhead for different activities (error bars represent ±95% confidence interval)

ing the IA session, participants witnessed 222 FRs in total. For the field study, on average our app logged 104 `ACTION_SCREEN_ON` events (*median = 57, sd = 65*) and 63 `ACTION_USER_PRESENT` events (*median = 42, sd = 28*) per participant per day. In total 10,608 `ACTION_SCREEN_ON` events and 6,420 `ACTION_USER_PRESENT` events for 34 participants were logged across three days. During the field study, participants also provided feedback against 693 FRs (98, 214, and 381 for low, medium and high operating threshold, respectively) and dismissed 42 authentications and feedbacks.

For qualitative analysis of the semi-structured interviews, the researchers coded all participant responses using the grounded theory approach [14] with meetings to achieve consensus. For test statistics, we use a t-test when comparing continuous data between subjects (such as between *UseAuth* and *DontUseAuth* or between *UsePIN* and *UsePAT*). We use a paired t-test when comparing continuous data for the within subjects condition (IA and *Non-IA*). We use a chi-square test for participants' responses to categorical Likert scale questions.

5.1 Usability Evaluation of IA

5.1.1 U1: Effect of IA on overall task completion time and authentication overhead

Overall task completion time is the total time to complete all 30 tasks including EA authentications or IA interrupt-authenticates if present. We calculate the increase or decrease in this time for each individual participant for their IA session compared to their non-IA session. This relative measure compensates for inter-participant differences due to reading level, motor skills, etc. The time differences are aggregated by *UseAuth* and *DontUseAuth* participants (Figure 2). Intuitively, the IA session should take less time as compared to the non-IA session for *UseAuth* participants due to fewer authentications, and more time when compared to the non-IA session for *DontUseAuth* participants. For *UseAuth* participants, the overall task completion time on average decreased by 100 seconds (*median = -103; mean = -101; sd = 59*). A paired t-test between the completion times of the IA and non-IA session for the *UseAuth* participants indicates that they are significantly different (*t = -3.6, p = 0.01*). The overall task completion time for *DontUseAuth* participants increased by 120 seconds on average (*median = 107; mean = 122; sd = 52*) for the IA session. A paired t-test between the completion times of the IA and non-IA session for the *DontUseAuth* participants indicates that they are significantly different (*t = 5.2, p = 0.014*).

We also evaluate the interrupt-authenticate overhead, defined as additional time taken for IA interrupted tasks compared to non-interrupted tasks. For our tasks, interrupt-authenticate overhead is the difference between the average completion times of an activity, with each activity type analysed separately. Figure 3 shows that on average, *DontUseAuth* participants had an interrupt-authenticate overhead of 8, 15, and 20 seconds for the interrupted input, selection, and reading activities, respectively. Similarly, on average, the *UseAuth* participants had an interrupt-authenticate overhead of 7, 11, and 16 seconds for the input, selection, and reading activities, respectively. A t-test for interrupt-authenticate for each activity reveals that the difference is not significant for the input activity between IA interrupted tasks and non-interrupted tasks (*t = 1.6, p = 0.11*), but the difference is significant for the selection (*t = 7.8, p = 0.002*) and the reading activity (*t = 10.5, p = 0.002*).

Discussion: While these results indicate that IA imposes an interrupt-authenticate overhead for the individual interrupted tasks, the total completion time decreased by 7.1% for *UseAuth* participants because they did not authenticate for every task. For *DontUseAuth* participants, we observe 8.8% increase in the total completion time due to interrupt-authenticates. It should be noted that the performance gains (or losses) will be more pronounced when the number of device usage sessions increases. The interrupt-authenticate overhead was primarily due to the unpredicted or sudden "lock-out" and the context switch as pointed out by the participants:

> *"It pops-out very suddenly... in between the tasks at times. I got tensed because I was worried about completing the task without making the pop-up appear... getting pop-up in the middle of task was quite distracting"* (P10)

> *"I generally lost my train of thought when it popped up that authentication"* (P37)

5.1.2 U2: Effect of IA on the task error rate

We classified errors using simple correctness checks built into the apps. An error occurred when: entered numbers mismatched in the input activity; incorrect answers were provided to a question in the reading activity; or a target word was missed or a non-target word was selected in the selection activity. We calculated the error rate separately for the 222 interrupted tasks from the IA session and for the 222 uninterrupted tasks located at the same task index from the EA session (Figure 4).

A t-test indicates that the differences in error rates across uninterrupted and interrupted tasks are not statistically significant for input (*t = -1.0, p = 0.69*), selection (*t = 1.5, p*

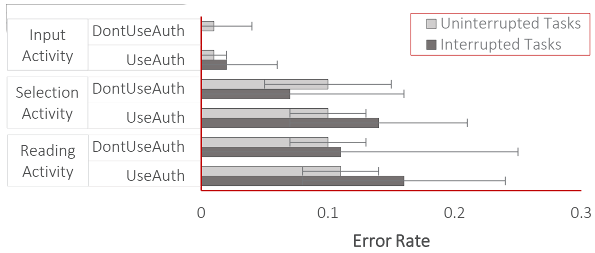


Figure 4: Error rate between interrupted tasks from the IA session and corresponding uninterrupted tasks from the non-IA session (error bars represent ±95% confidence interval)

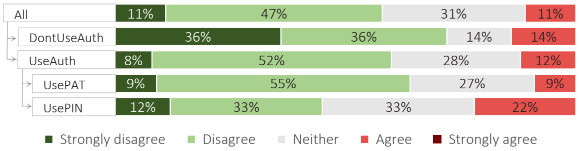


Figure 5: Responses for "Do you agree with the statement 'I think this method is annoying'?"

= *0.32*) and reading activities (*t = -1.8, p = 0.30*). There is no evidence that the interrupt-authenticate model increases the error rate.

Discussion: Our results agree with Bailey and Konstan's [1] findings for personal computers where interruptions did not increase the error rate of interrupted tasks. However, they also found that the expectancy of interruptions caused more errors overall (due to a higher load on the cognitive resources). While our participants complained about the unpredictability of interrupt-authenticates, a paired t-test reveals that there is no significant difference between the error rates of the IA and non-IA session (*t = 0.84, p = 0.4*).

5.1.3 U3: Effect of fewer but less predictable authentication interrupts on annoyance

After the in-lab sessions, participants answered survey questions regarding annoyance. The first question asked if they thought the overall experience of IA was annoying (see Figure 5). Overall, 58% did not say IA was annoying, 11% considered IA as annoying while the rest were neutral. Furthermore, significantly fewer *UseAuth* participants (12% less) thought IA was not annoying compared to *DontUseAuth* participants ($\chi^2(1) = 5.1, p = 0.02$). There is also a significant difference in annoyance for participants based on the type of EA currently used: significantly more *UsePIN* participants (22% more) found IA to be annoying compared to *UsePAT* participants ($\chi^2(1) = 4.09, p = 0.04$). We suspect that IA's lower perceived level of protection by *UsePIN* participants (discussed in § 5.2.1) and consequent low utility is responsible for this.

The second question asked participants how annoying the IA *interrupt-authenticates* were (see Figure 6). Overall, 35% of the participants found them to be annoying (32% somewhat annoying and 3% very annoying), 44% found them to be tolerable and 21% found them to be not annoying. Furthermore, significantly more *UseAuth* participants (28% more) found interruptions to be annoying as compared to *DontUseAuth* participants ($\chi^2(1) = 9.4, p = 0.002$). We did not find evidence for increased anxiety (which could be linked to annoyance) during the lab sessions: a pair-wise t-test does not indicate any changes in anxiety across the participants for the IA and the non-IA session (*t = 10.6, p = 0.82*).

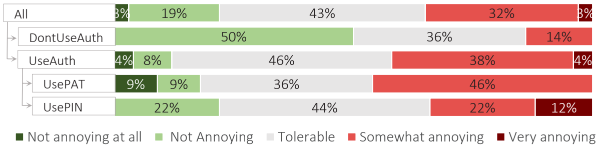


Figure 6: Responses for "How annoying were the interruptions for authentication?"

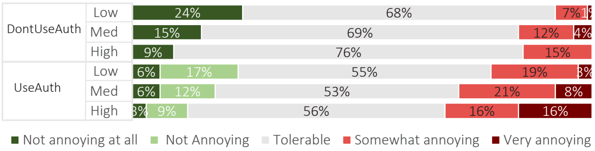


Figure 7: Annoyance against three operating thresholds from the in-situ feedback survey of the field study

During the field study, we subjected participants to three different operating thresholds corresponding to 5%, 10% and 20% FR rate with a goal to determine an *acceptable threshold* (operating thresholds with "tolerable" or better annoyance ratings). The feedback of participants for annoyance across these FR rates is provided in Figure 7. *DontUseAuth* participants found interrupt-authenticates to be more acceptable for different thresholds as compared to *UseAuth* participants. More specifically, for low, medium and high FR rates, interrupt-authenticates for *DontUseAuth* participants were significantly more likely (14%, 13% and 17% more) to be acceptable as compared to *UseAuth* participants ($\chi^2(1) = 11.4, p < 0.001$), ($\chi^2(1) = 8.2, p = 0.004$) and ($\chi^2(1) = 13.2, p < 0.001$), respectively. Figure 7 also illustrates responses in terms of the proportion of interrupt-authenticates that are annoying between low-medium and medium-high thresholds, while differences between medium-high thresholds are negligible. More specifically when *DontUseAuth* participants were subjected to the low threshold, interrupt-authenticates were significantly more acceptable (8% more) as compared to the medium threshold ($\chi^2(1) = 4.7, p = 0.03$). On the other hand, for *DontUseAuth* participants the difference in terms of proportion of acceptable interrupt-authenticates between medium-high threshold was insignificant — 84% vs. 85% ($\chi^2(1) = 0.07, p = 0.78$), respectively. These observations for inter-threshold level correlations across *DontUseAuth* participants were also true for *UseAuth* participants.

Discussion: Although the majority of participants were not annoyed with IA, it is clear that interrupt-authenticates can cause moderate levels of annoyance, more so for users who currently use EA. During the qualitative interviews, we asked the participants for the cause of this annoyance and 10/37 participants indicated the unpredictability of the interrupt-authenticate as the reason:

> *"I think it is a little bit annoying because there is a little stress [when] you don't know what will happen"* (P15)

> *"I am ready to enter a password before I start doing anything whereas for implicit authentication it catches*

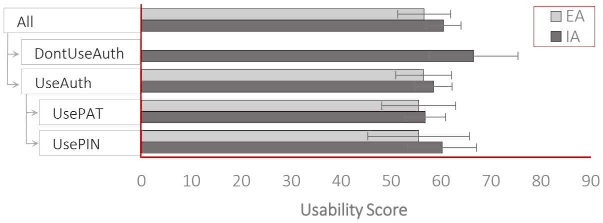


Figure 8: Average system usability scale (SUS) score for IA and non-IA sessions (±95% confidence interval)

me off-guard " (P14)

4/37 participants were more annoyed due to a perceived error at the part of IA:

> *"It is the unpredictability of it... I know that I have to enter my PIN every time and this becomes annoying... it would be frustrating because you don't know what was wrong that you did, with PIN you know because it is something wrong that you entered "* (P14)

> *"It sure was annoying. I use my phone a lot when I am watching TV and at times my device turns off due to inactivity. That's my fault and [it] is understandable but when this interrupts me, I think that the operating system is faulty or something"* (P37)

For the field study, eight *UseAuth* participants had to use IA in addition to their EA scheme despite their preference to replace their current EA scheme with IA. Two of these participants mentioned that the cause of annoyance was "redundant authentications":

> *"I felt like it was a lot of work with two PINs. At times I would confuse which one was which and then had to re-enter it. That made it more annoying"* (P20)

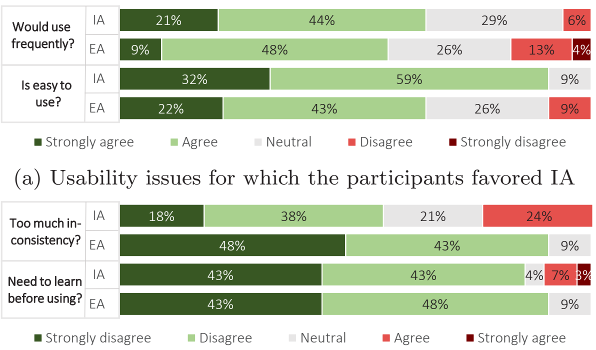
> *"I already use a password and after that it was quite annoying to use a second one. It seemed redundant"* (P32)

5.1.4 *U4: Overall usability of IA*

The SUS scores for IA and EA are provided in Figure 8. Since we used nine questions from SUS, we report the usability scores out of a total of 90. A higher SUS score indicates that a system is more usable. A score is unavailable for the non-IA session of the *DontUseAuth* participants because they did not authenticate in that session. A t-test does not indicate any significant differences for the SUS scores between EA and IA (t = *-2.4*, p = *0.31*). Similarly, a t-test does not indicate any significant differences between *DontUseAuth* and *UseAuth* participants for the IA session (t = *4.7*, p = *0.1*).

Discussion: While IA did not outperform EA on SUS, there are interesting differences for the individual SUS questions between IA and EA. Figure 9a shows that significantly more participants (8% more) indicated that they would like to use IA frequently as compared to EA $(\chi^2(1) = 5.1,\ p = 0.02)$. Also significantly more participants (26%) thought that IA was easier to use compared to EA $(\chi^2(1) > 100,\ p < 0.001)$. Supporting comments for the ease of use:

> *"I like that it really can be easy if your phone is effective at recognizing you and doesn't ask you to enter the password"* (P21)



(a) Usability issues for which the participants favored IA

(b) Usability issues for which the participants favored EA

Figure 9: Responses to the individual SUS questions

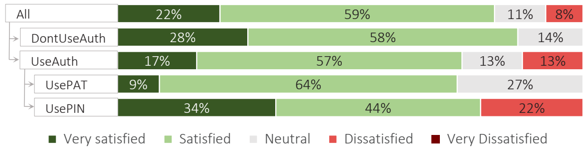


Figure 10: Responses for "How satisfied are you with the overall level of protection that is provided?"

> *"Actually I found it pretty easy to use and it wasn't bothersome"* (P21)

As would be expected, significantly more participants (24% more) thought that IA was more inconsistent than EA $(\chi^2(1) = 64,\ p < 0.001)$ (Figure 9b). Furthermore, significantly more participants (10% more) thought that they need to learn more about IA $(\chi^2(1) = 15,\ p < 0.001)$. In § 5.2.3, we discuss the learnability issue in detail.

5.2 Security Perceptions of IA

5.2.1 *S1: Perceptions of IA security properties*

Participants were made aware of the security properties of IA including FAs, FRs and the detection delay through the briefing video and the lab-based experiment. We then asked participants how satisfied they were with the overall level of protection that was provided by IA (see Figure 10). Overall, 81% of the participants were satisfied (22% very satisfied and 59% were satisfied) with IA, 8% were not satisfied and the rest were undecided. *DontUseAuth* participants were significantly more (12% more) likely to be satisfied with the level of protection that was provided by IA as compared to *UseAuth* participants $(\chi^2(1) = 9.5,\ p = 0.001)$. Two *UsePIN* participants were not satisfied while three *UsePAT* participants were undecided about the overall level of protection that was provided by IA.

We also asked participants regarding their perceived level of protection against different adversaries, device states and tasks (Figure 11). Overall, 12%, 6%, 3% and 15% of the participants were not satisfied with the level of protection that was provided against coworkers, spouse, friends and strangers, respectively. In terms of different device states, 21%, 6% and 3% of the participants were not satisfied with the provided level of protection if their device was lost in a public location, unattended at work and unattended at home, respectively. Finally, 33%, 12% and 9% of the par-

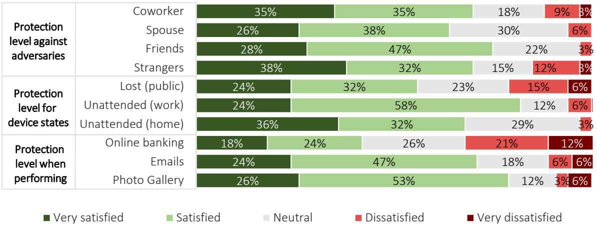


Figure 11: Security perception responses according to different adversaries, device states and tasks

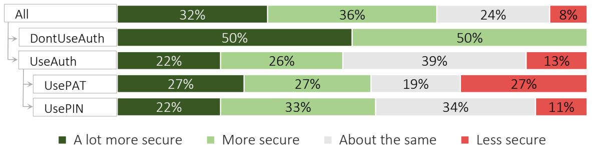


Figure 12: Responses for "How secure is this method as compared to your current authentication method?"

ticipants were not satisfied if IA was protecting their device while there was a banking app, email app and photo gallery app on their device, respectively.

Discussion: Overall, we found that participants were satisfied with the level of security that was provided by IA. However, 18/37 participants showed some concerns regarding FAs, detection delays and possible mimicry attacks in IA. We now shed some light on the concerns based on the participants' comments. The non-zero FA rate was a concern for 8/37 participants:

> *"I'm not sure what will happen when it is lost. It will depend on who picks it up and I may get unlucky if his behaviour is the same as mine"* (P30)

> *"No one can use the phone without entering the PIN but here someone 'can' use it"* (P25)

5.2.2 S2: Perceptions of IA security vs. current method

We asked participants how secure they thought IA was compared to the authentication method that they currently used (Figure 12). All *DontUseAuth* participants perceived IA to be more secure and 87% of *UseAuth* participants thought that IA was at least as secure as their current method or more secure. Only 13% of *UseAuth* participants perceived IA to be less secure due to the security concerns discussed in the previous section.

Discussion: Clarke et al. [5] and Crawford and Renaud [6] found that 92% and 73% of their participants considered IA to be more secure as compared to the traditional authentication schemes, respectively. On the other hand, only 48% of our *UseAuth* participants thought that IA was more secure as compared to their authentication schemes. Our results are not consistent with the previous findings. In the original papers, Clarke et al. [5] and Crawford and Renaud [6] do not mention briefing the participants regarding the IA limitations. We suspect that this difference in results is due to the increased knowledge of our participants about the limitations of IA.

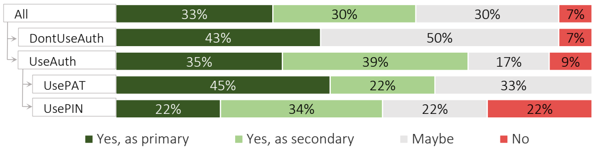


Figure 13: Responses for "Would you use IA?"

5.2.3 S3: Willingness to Adopt IA

We asked participants how willing they were to use IA with four choices: (i) Yes, I would replace my current scheme with IA; (ii) Yes, I would use it in addition to my current authentication scheme; (iii) I may use it; and (iv) No, I will not use it. The purpose of introducing a spectrum of answers was to understand the various types of authentication needs that IA might be able to satisfy. The response of participants is provided in Figure 13. Overall 63% of participants were interested in using IA either as a primary (33%) or a secondary (30%) authentication mechanism. On the other hand, 30% of participants were not sure whether they would use IA and 7% did not want to use IA. A further breakdown across authentication preferences indicates that *DontUseAuth* participants were significantly more likely (37% more) to be unsure about IA adoption as compared to *UseAuth* participants $(\chi^2(1) = 19,\ p < 0.001)$. Interestingly, we found 4/7 participants who were not adequately satisfied with IA's level of protection (S2) were still interested in using it.

Discussion: Although 63% of our participants were willing to use IA as a primary or secondary authentication method, this is less than the 90% willingness to adopt IA that Crawford and Renaud [6] found. This difference is likely due to Crawford and Renaud providing a binary yes-or-no option rather than the spectrum of answers we provide. The interview provides rationale for the participants' choices. 6/14 *DontUseAuth* participants were interested in adopting IA because they thought that it provided convenience and protection which was better than no authentication:

> *"It seemed easier than entering a password and more secure than not using anything"* (P30)

> *" Instead of forcing me to enter the password every time, it offers me not to enter a password which is my current preferred level of security and it provides additional security on top"* (P14)

On the other hand, seven *DontUseAuth* participants who said that they may use IA wanted to test it before making up their minds, for example:

> *"I would give it a shot for a month and if I see that it is getting a lot better I will like using it."* (P12)

> *"I have only used it once so I am not sure. I will have to use it for a longer duration and would like to test it on other people too"* (P7)

One *DontUseAuth* participant who did not choose to use IA did so because she had nothing to protect on her device:

> *"If I had work related data or anything else on my device that needed protection, I would use it. Right now I don't have anything that needs protection"* (P36)

UseAuth participants who chose to replace their EA scheme with IA did so because they felt it was more convenient: saves time (6/8) or has fewer authentications (2/8). They seemed to understand the associated risks but thought the trade-off was reasonable:

> *"Because in past months I never had a non-approved access to my phone and in the seven months I have entered my PIN thousands of times and it will be less annoying to use IA"* (P11)

> *"Even with the disadvantages [perceived low level of protection], I think I like the less number of authentications, given that I carry my phone with me"* (P8)

UseAuth participants who said they would use IA as a secondary authentication did so to have an additional layer of security (4/9) or to test it further (5/9):

> *"I would be interested since I work at the community center and at times I have to leave it at places and then I have to worry about it."* (P6)

> *"It would be beneficial for spouse because they would know your password by asking it or see you type a lot but they won't know your pattern [behaviour]"* (P14)

> *"I think just because of my unfamiliarity with it, passwords, I am accustomed with it, but perhaps the more I use it, the more I will trust it"* (P32)

> *"I am so used to typing something in that I think it will take me a while to feel comfortable that authentication has occurred as oppose to you know when you turn it on and enter it"* (P19)

Finally, *UseAuth* participants who said no to using IA had concerns related to its detection delay (1/2) or they felt that EA was more suitable for them (1/2):

> *"Not the best for adoption because people would start looking at my photo gallery before it would lock them out."* (P16)

> *"I think the current system that I have is enough to deter strangers and for the cases when the phone gets stolen "* (P28)

6. DISCUSSION & DESIGN IMPLICATIONS

Our results suggest barriers for IA adoption and deployment along with associated design implications.

6.1 Mitigating the Effects of Interruptions

Overall, participants ranked IA higher than EA in terms of ease of use and the majority of participants reported that IA and interrupt-authenticates were at least tolerable. However, interrupt-authenticates did increase task completion time and several participants, especially those already using EA, felt the interruptions were annoying. Their comments suggest this was largely due to the unpredictability of interrupt-authenticates and the context-switch. Mitigating the negative effects of these necessary interrupt-authenticates remains a challenge for IA. Two participants suggested authentication without interruption by using the front camera of the device to perform facial-recognition. Similarly, a participant who was interested in using IA as a secondary mechanism to protect from misuse by friends or family members suggested IA took a picture of the perpetrator and email it to her. Two participants also suggested that instead of instantly locking the user out, the IA scheme could display an authentication screen in a smaller window on the screen and allow the user to choose the best moment to context-switch and authenticate:

> *"PINs were really annoying a lot of times. I would forget what I was about to say. I was wondering if there was a mechanism where it could indicate that it was locking on me in three, two, one and show me a small screen on the side to authenticate in parallel."* (P8)

It is important to understand that interruptions serve a purpose beyond authentication. Supporting the findings of Crawford and Renaud [6], some "annoyed" participants indicated that while interrupt-authenticates were annoying, they were necessary to indicate that IA was working and arguably contribute to a perceived sense of security:

> *"I guess it was frustrating that it kicked me out, but I could deal with that... and if it did ask for the PIN just knowing that the phone will be secure, it comes down to that"* (P22)

> *"Yeah the interruptions are annoying and I guess then I have to say to myself, practically it is not good but as soon as I see it, I know that it is protecting me"* (P5)

Balancing the need for interruptions with potential annoyance is a design challenge. The alternate authentication methods discussed above could be one approach, but the visual design of the authentication screen (e.g. choice of colour and language) as well as the timing of the interruption (e.g. postponing for non-sensitive tasks, slow fade in) are critical choices.

6.2 Opaque Deployment of IA

Our IA deployment was hidden as a background service, so participants were essentially unaware of its operation until a FR interrupt-authenticate occurred. While these interruptions currently serve to indicate that IA is working (see above), as IA detection algorithms evolve and FR rates decrease, these interruptions will become very infrequent and thus IA will be more opaque. Furthermore, operating at a relaxed operating threshold also reduces the number of FRs and the consequent visibility of IA. The background operations (opacity) of IA will raise concerns like:

> *"So looking at.. I see there is no lock. Sometimes I felt... the lock exists or not? When PIN is used, I know [it] every time. Now there is no way to discriminate if someone has hacked my phone and removed the lock. PIN is a secure feeling that the phone is safe"* (P31)

> *"The main problem for me is that if I am unaware that what I am doing is authenticated then I don't know if my device is secure. With this [IA] there really is no way of knowing that I have been authenticated when I open an app... How do I know it is not a fluke. Maybe it is no longer running and protecting, how do I know"* (P19)

The concerns of users regarding the background deployment of IA have never been raised in existing literature. Since these issues arise due to their inability to tell whether IA is protecting their device, simple UI changes may be able to address these. For example, an indicator on the status bar can be used to indicate the current status of IA scheme. While such an indicator can keep the users up-to-date and act as a deflector against potential adversaries, it may also notify adversaries and enable them to launch highly focused attacks to gain access to the target data before being locked out. For IA deployment, the design and control of an IA status indicator needs to be studied closely and respective trade-offs need more exploration.

6.3 Operating Threshold Customization

The in-situ feedback screen (Figure 14d) provided participants with an option to adjust the operating threshold during the field study, and we asked them about their experiences with this functionality. 17/34 participants indicated that they found the customization capability to be useful. A common explanation was that they reduced the operating threshold when texting or when at home. 5/34 participants indicated that they always set it to high to get maximum protection (4/5 belonged to *DontUseAuth*). 12/34 participants never adjusted the threshold during the field study and relied on the value chosen by the IA scheme.

These specific results have some limitations since any operating threshold customization in our app was temporary – the threshold was set to a predetermined value each day of the three-day field study. Nevertheless, participant comments indicate that there is a need to explore the various customization options for IA (such as the trade-off between FA and FR; and the detection delay and FR). For example:

> *"I think threshold selection bar would be a useful function. I feel having that is more choice and useful."* (P21)

The threshold customization interface needs to communicate the security and annoyance trade-off for the chosen threshold so users can make informed decisions.

7. LIMITATIONS

Our study has reasonable limitations due to the inclusion of human subjects: the scope is limited to people willing to participate; it contains self reported and subjective views; participants might be inclined to provide favorable responses to the researchers; and the known limited duration of the field study might have made participants more optimistic about their annoyance. Since these are not easily preventable, we focus on limitations specific to our study:

1. We use a pseudo-IA scheme to strictly control FR rates and to circumvent restrictions on Android event data collection. As a result, it was possible for participants to witness unexpected behaviour of IA (for instance, they may get a FR or a TA for what they felt was the exact same sequence of touch input).
2. In the field study, all participants of the *UseAuth* group evaluated IA as a secondary authentication scheme (including those who indicated they would replace their EA scheme with IA). This resulted in multiple authentications during a single session (the system EA first, then an IA interrupt-authenticates), which may have contributed to feelings of annoyance. However, the deployment of IA as a primary authentication scheme in the field was not possible due to security and privacy issues.
3. When a FR occurred in the field study, participants had to authenticate and provide in-situ feedback. Although we designed the feedback pop-up to be simple to complete, it may have increased annoyance. Since participants had the option to dismiss the authentication and the feedback pop-up in extreme situations. Only 6% of the pop-ups were dismissed, but this may have slightly skewed results by underreporting annoyance.
4. In this study, we did not compare IA with biometric-based EA schemes, like fingerprint- or facial-recognition schemes. These alternative biometric-based authentication schemes have the same limitations of EA schemes discussed in § 1 and have usability limitations of their own [2]. A part of our future work is to validate our findings for these alternative EA schemes.

Limitations 1 and 2 are attributed to the pseudo-IA scheme, but this is a reasonable trade-off for the advantages of using pseudo-IA for strict control of FR rates and elimination of confounds from performance idiosyncrasies of specific IA algorithms. Limitation 3 is a commonly accepted trade-off for benefits from gathering in-situ feedback. Limitation 4 must be considered in light of the fact that our study compares IA to no authentication and dominant forms of EA (PIN and pattern): these arguably form lower and upper baselines for usability and security perception.

8. CONCLUSION

Our two-part study on IA usability and security perceptions provides empirical evidence for the "human side" of IA. In terms of performance, the interrupt-authenticate model may impose overhead for individual authentications, but it increases amortized task performance without affecting the error rate. For usability perceptions, there is no significant difference between IA and EA for SUS and 26% more of our participants agreed that IA was more convenient. However, annoyance is a potential issue with IA with 35% of the participants who found interrupt-authenticates annoying. For security perception, detection delay and FAs were issues for 27% and 22% participants respectively, and 11% of our participants were concerned about the feasibility of mimicry attacks. Yet, participants who currently use explicit authentication perceived IA to be more secure, or at least as secure as their current authentication method. Perhaps, most encouraging is that 63% of our participants were interested in adopting IA and a further 30% were interested in trying IA out with possibility of adoption. Based on insights gained from post-study interviews, we propose design implications that may reduce annoyance and increase security perception even more. Overall, our findings provide supporting evidence for earlier work's [25] postulation: IA is indeed a meaningful approach with a reasonable trade-off in terms of usability and security.

9. ACKNOWLEDGMENTS

We thanks the anonymous reviewers and Tao Wang for their valuable comments. We also thank Google and Ontario Research Fund for their support.

10. REFERENCES

[1] B. P. Bailey and J. A. Konstan. On the need for attention-aware systems: Measuring effects of interruption on task performance, error rate, and affective state. *Computers in Human Behavior*, 22(4), 2006.

[2] C. Bhagavatula, B. Ur, K. Iacovino, S. M. Kywe, L. F. Cranor, and M. Savvides. Biometric authentication on iphone and android: usability, perceptions, and influences on adoption. In *Workshop on Usable Security*, 2015.

[3] C. Bo, L. Zhang, X.-Y. Li, Q. Huang, and Y. Wang. Silentsense: silent user identification via touch and movement behavioral biometrics. In *19th Annual International Conference on Mobile Computing & Networking*. ACM, 2013.

[4] J. Brooke. SUS – a quick and dirty usability scale. *Usability Evaluation in Industry*, 189(194), 1996.

[5] N. Clarke, S. Karatzouni, and S. Furnell. Flexible and transparent user authentication for mobile devices. In *Emerging Challenges for Security, Privacy and Trust*. Springer, 2009.

[6] H. Crawford and K. Renaud. Understanding user perceptions of transparent authentication on a mobile device. *Journal of Trust Management*, 1(7), 2014.

[7] B. Draffin, J. Zhu, and J. Zhang. Keysens: Passive user authentication through micro-behavior modeling of soft keyboard interaction. In *Mobile Computing, Applications, and Services*. Springer, 2014.

[8] S. Egelman, S. Jain, R. S. Portnoff, K. Liao, S. Consolvo, and D. Wagner. Are you ready to lock? In *ACM SIGSAC Conference on Computer & Communications Security*. ACM, 2014.

[9] T. Feng, J. Yang, Z. Yan, E. M. Tapia, and W. Shi. Tips: Context-aware implicit user identification using touch screen in uncontrolled environments. In *15th Workshop on Mobile Computing Systems and Applications*. ACM, 2014.

[10] T. Feng, X. Zhao, B. Carbunar, and W. Shi. Continuous mobile authentication using virtual key typing biometrics. In *12th International Conference on Trust, Security and Privacy in Computing and Communications*. IEEE, 2013.

[11] J. Frank, S. Mannor, and D. Precup. Activity and gait recognition with time-delay embeddings. In *AAAI Conference on Artificial Intelligence*, 2010.

[12] M. Frank, R. Biedert, E. Ma, I. Martinovic, and D. Song. Touchalytics: On the applicability of touchscreen input as a behavioral biometric for continuous authentication. *IEEE Transactions on Information Forensics and Security*, 8(1), 2013.

[13] C. Giuffrida, K. Majdanik, M. Conti, and H. Bos. I sensed it was you: Authenticating mobile users with sensor-enhanced keystroke dynamics. In *Detection of Intrusions and Malware, and Vulnerability Assessment*. Springer, 2014.

[14] B. G. Glaser and A. L. Strauss. *The discovery of grounded theory: Strategies for qualitative research*. Transaction Publishers, 2009.

[15] M. Harbach, E. von Zezschwitz, A. Fichtner, A. De Luca, and M. Smith. It's a hard lock life: A field study of smartphone (un) locking behavior and risk perception. In *Symposium on Usable Privacy and Security*, 2014.

[16] E. Hayashi, O. Riva, K. Strauss, A. Brush, and S. Schechter. Goldilocks and the two mobile devices: going beyond all-or-nothing access to a device's applications. In *Symposium on Usable Privacy and Security*. ACM, 2012.

[17] H. Khan, A. Atwater, and U. Hengartner. A comparative evaluation of implicit authentication schemes. In *Recent Advances in Intrusion Detection*. Springer, 2014.

[18] H. Khan, A. Atwater, and U. Hengartner. Itus: an implicit authentication framework for android. In *20th Annual International Conference on Mobile Computing & Networking*. ACM, 2014.

[19] L. Li, X. Zhao, and G. Xue. Unobservable reauthentication for smart phones. In *20th Network and Distributed System Security Symposium*, volume 13, 2013.

[20] Lookout Blog. Sprint and lookout consumer mobile behavior survey. http://blog.lookout.com/blog/2013/10/21/sprint-and-lookout-survey/, Feb. 2015.

[21] M. Muaaz and R. Mayrhofer. An analysis of different approaches to gait recognition using cell phone based accelerometers. In *International Conference on Advances in Mobile Computing & Multimedia*. ACM, 2013.

[22] New Scientist. Touchscreen phones know it's you from taps and swipes. http://www.newscientist.com/article/dn24193-touchscreen-phones-know-its-you-from-taps-and-swipes.html, Feb. 2015.

[23] T. Patil, G. Bhutkar, and N. Tarapore. Usability evaluation using specialized heuristics with qualitative indicators for intrusion detection system. In *Advances in Computing and Information Technology*. Springer, 2012.

[24] M. Shahzad, A. X. Liu, and A. Samuel. Secure unlocking of mobile touch screen devices by simple gestures: you can see it but you can not do it. In *19th Annual International Conference on Mobile Computing & Networking*. ACM, 2013.

[25] E. Shi, Y. Niu, M. Jakobsson, and R. Chow. Implicit authentication through learning user behavior. In *Information Security*. Springer, 2011.

[26] W. Shi, J. Yang, Y. Jiang, F. Yang, and Y. Xiong. Senguard: Passive user identification on smartphones using multiple sensors. In *7th International Conference on Wireless and Mobile Computing, Networking and Communications*. IEEE, 2011.

[27] C. D. Spielberger. Manual for the State-Trait Anxiety Inventory STAI (Form Y). 1983.

[28] S. Trewin, C. Swart, L. Koved, J. Martino, K. Singh, and S. Ben-David. Biometric authentication on a mobile device: a study of user effort, error and task disruption. In *28th Annual Computer Security Applications Conference*. ACM, 2012.

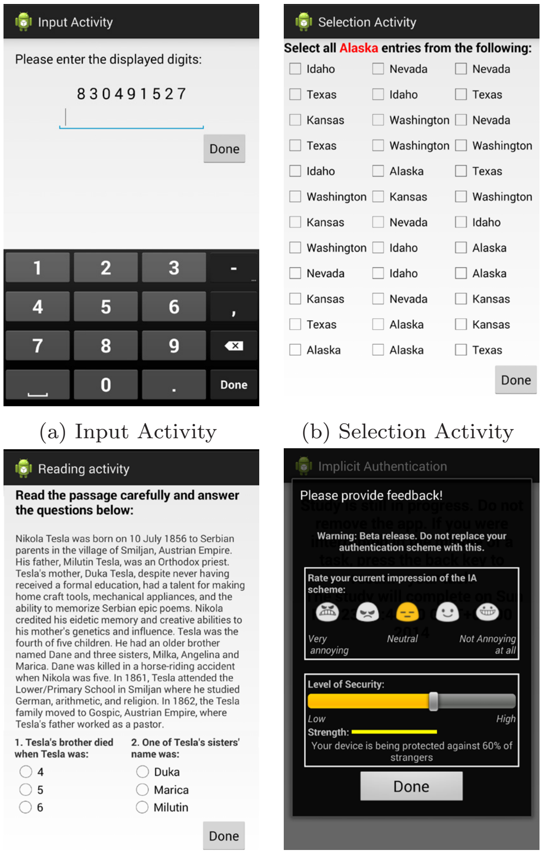
[29] E. Von Zezschwitz, P. Dunphy, and A. De Luca. Patterns in the wild: a field study of the usability of pattern and pin-based authentication on mobile devices. In *15th international conference on Human-computer interaction with mobile devices and services*. ACM, 2013.

[30] R. Werlinger, K. Hawkey, K. Muldner, P. Jaferian, and K. Beznosov. The challenges of using an intrusion detection system: is it worth the effort? In *Symposium on Usable privacy and security*. ACM, 2008.

[31] H. Xu, Y. Zhou, and M. R. Lyu. Towards continuous and passive authentication via touch biometrics: An experimental study on smartphones. In *Symposium On Usable Privacy and Security*, volume 14, 2014.

APPENDIX

A. APPS' ACTIVITY AND FEEDBACK SCREENS



Input Activity

Please enter the displayed digits:

8 3 0 4 9 1 5 2 7

(a) Input Activity

Selection Activity

Select all Alaska entries from the following:

□ Idaho	□ Nevada	□ Nevada
□ Texas	□ Idaho	□ Texas
□ Kansas	□ Washington	□ Nevada
□ Texas	□ Washington	□ Washington
□ Idaho	□ Alaska	□ Texas
□ Washington	□ Kansas	□ Washington
□ Kansas	□ Nevada	□ Idaho
□ Washington	□ Idaho	□ Alaska
□ Nevada	□ Idaho	□ Alaska
□ Kansas	□ Nevada	□ Kansas
□ Texas	□ Alaska	□ Kansas
□ Alaska	□ Alaska	□ Texas

(b) Selection Activity

Reading activity

Read the passage carefully and answer the questions below:

Nikola Tesla was born on 10 July 1856 to Serbian parents in the village of Smiljan, Austrian Empire. His father, Milutin Tesla, was an Orthodox priest. Tesla's mother, Duka Tesla, despite never having received a formal education, had a talent for making home craft tools, mechanical appliances, and the ability to memorize Serbian epic poems. Nikola credited his eidetic memory and creative abilities to his mother's genetics and influence. Tesla was the fourth of five children. He had an older brother named Dane and three sisters, Milka, Angelina and Marica. Dane was killed in a horse-riding accident when Nikola was five. In 1861, Tesla attended the Lower/Primary School in Smiljan where he studied German, arithmetic, and religion. In 1862, the Tesla family moved to Gospic, Austrian Empire, where Tesla's father worked as a pastor.

1. Tesla's brother died when Tesla was:
○ 4
○ 5
○ 6

2. One of Tesla's sisters' name was:
○ Duka
○ Marica
○ Milutin

(c) Reading Activity

Implicit Authentication

Please provide feedback!

Warning: Beta release. Do not replace your authentication scheme with this.

Rate your current impression of the IA scheme:

Very annoying *Neutral* *Not Annoying at all*

Level of Security:

Low *High*

Strength:

Your device is being protected against 60% of strangers

(d) In-situ feedback screen

Figure 14: Apps' screens showing different activities for the lab-based experiment and the feedback screen for the field study

B. PRE-STUDY

For the pre-study screening, in addition to collecting their name, email address, gender, age group, device that they owned (such as Nexus 4, Samsung SIII, LG G2), profession, domain (such as technology, medicine) and how long they have used an Android device, we asked the participants:

1. Which protection mechanism do you use on your smartphone:

   (a) None; (b) PIN (4 digit or more); (c) Password (characters and numbers); (d) Pattern Lock; (e) Face recognition; (f) Fingerprint recognition; (g) Other (please specify)

2. **IF NO AUTHENTICATION** Why do you not use any protection mechanism (choose all that apply):

   (a) It's too much of a hassle / takes time; (b) There is nothing on my phone that I need to hide; (c) No one would care about what's on my phone; (d) In an emergency, others can use my phone; (e) I've never thought about it; (f) Other (please specify)

3. **IF SOME AUTHENTICATION** Which of the following scenarios do you want to protect against (choose all that apply):

   (a) Phone protected if stolen; (b) Phone protected if lost (c) Phone protected if misplaced; (d) Phone protected if left unattended (e) Someone casually picking up the phone; (f) Unwanted disclosure, pranks; (g) Other (please specify)

4. **IF SOME AUTHENTICATION** Which of the following describes you (choose all that apply):

   (a) Unlocking my phone is annoying sometimes; (b) I like the idea that my phone is protected from unauthorized access; (c) It takes too much time; (d) Unlocking my phone is easy

5. **IF SOME AUTHENTICATION** Which of the following attackers do you want to protect from (choose all that apply):

   (a) Coworker; (b) Spouse; (c) Roommate; (d) Own children; (e) Other unwanted individual/stranger; (f) Friends; (g) Other (please specify)

6. Do you sometimes take additional measures to protect your smartphone (choose all that apply):

   (a) None; (b) I leave my phone in a safe place before going somewhere; (c) I conceal my smartphone in my clothes or in a bag (d) I have changed security settings on my device to improve security (such as reduced automatic lock time) (e) I have enabled device encryption on my smartphone (f) Other (please specify)

7. Do you share your smartphone with your friends or family members:

   (a) Never; (b) Rarely (once a month); (c) Occasionally (once a week); (d) Daily

C. POST-SURVEY

The following questions were asked in the post-survey conducted after each in-lab session in which a participant tested an authentication scheme (IA or EA). The questions that were only asked after the IA session are marked [**IA only**].

1. How satisfied are you with the level of protection that is provided against: (*5-point Likert scale "Very satisfied" - "Very dissatisfied"*)

   (a) Coworkers; (b) Spouse; (c) Roommate; (d) Own children; (e) Friends; (f) Strangers

2. How satisfied are you with the level of protection that is provided against the following phone states: (*5-point Likert scale "Very satisfied" - "Very dissatisfied"*)

   (a) Lost at public location; (b) Lost at work; (c) Unattended at work; (d) Unattended at home

3. How satisfied are you with the level of protection that is provided when you are performing following tasks on your Smartphone: (*5-point Likert scale "Very satisfied" - "Very dissatisfied"*)

(a) Online banking; (b) Online shopping; (c) Checking emails; (d) Checking texts; (e) Social networking (FaceBook); (f) Checking photo gallery

4. How satisfied are you with the overall level of protection that is provided? (*5-point Likert scale "Very satisfied" - "Very dissatisfied"*)
5. Do you agree with the statement "I think this method takes a lot of time"? (*5-point Likert scale "Strongly agree" - "Strongly disagree"*)
6. Do you agree with the statement "I think this method is annoying"? (*5-point Likert scale "Strongly agree" - "Strongly disagree"*)
7. Do you agree with the statement "I think this method is tiring"? (*5-point Likert scale "Strongly agree" - "Strongly disagree"*)
8. How annoying were the interruptions for authentication? (*5-point Likert scale "Very annoying" - "Not annoying at all"*)
9. [**IA only**] How annoying were the interruptions for authentication as compared to your current authentication method? (*5-point Likert scale "Very annoying" - "Not annoying at all"*)
10. [**IA only**] How secure this method is as compared to no authentication? (*5-point Likert scale "A lot more secure" - "A lot less secure"*)
11. [**IA only**] How secure this method is as compared to your current authentication method? (*5-point Likert scale "A lot more secure" - "A lot less secure"*)
12. [**IA only**] Would you use this authentication method?
    - Yes, I would replace my current scheme with IA
    - Yes, I would use it in addition to my current authentication scheme
    - I may use it
    - No, I will not use it.

D. SUS SURVEY

The modified SUS form that was completed by participants after each in-lab session in which they tested an authentication scheme (IA or EA). For each question, participants responded on a 5-point Likert scale ("Strongly agree" - "Strongly disagree").

1. I think that I would like to use this method frequently
2. I found this method unnecessarily complex
3. I thought this method was easy to use
4. I think that I would need the support of a technical person to be able to use this method
5. I thought there was too much inconsistency in this method
6. I would imagine that most people would learn to use this method very quickly
7. I found this method very cumbersome to use
8. I felt very confident using the system
9. I needed to learn a lot of things before I could get going with this system

E. SEMI-STRUCTURED INTERVIEWS

Participants were asked the following open-ended questions during the semi-structured interviews:

E.1 Lab-based Experiment

- What did you like about IA?
- What did you dislike about IA?
- Why did you perceive a specific protection level for IA?
- Why do you think IA will provide more/less/same protection as compared to your current scheme?
- Why (or why not) would you use IA?
- **IF NOT SATISFIED WITH IA PROTECTION LEVEL:** Why would you still use IA?
- **IF IA IS ANNOYING:** Why would you still use IA?
- Any particular scenarios where you think IA will be particularly useful/useless?
- **IF INTERRUPT-AUTHENTICATES ARE ANNOYING:** How do you think we can mitigate the annoyance?

E.2 Field Study

- How was your longer term experience of IA?
- Have you changed your opinion about IA? If yes, why?
- How annoying were the interruptions for authentication?
- Which apps were you using on your device when the interruptions were particularly annoying?
- Which apps were you using on your device when then interruptions were not annoying?
- Any particular scenarios where you think IA will be particularly useful/useless?
- Did you use the threshold adjustment bar useful? Why or why not?

Understanding the Inconsistencies between Text Descriptions and the Use of Privacy-sensitive Resources of Mobile Apps

Takuya Watanabe
Waseda University
3-4-1 Okubo Shinuku
Tokyo, Japan
watanabe@nsl.cs.waseda.ac.jp

Mitsuaki Akiyama
NTT Secure Platform Labs
3-9-11 Midoricho Musashino
Tokyo, Japan
akiyama.mitsuaki@lab.ntt.co.jp

Tetsuya Sakai
Waseda University
3-4-1 Okubo Shinuku
Tokyo, Japan
tetsuyasakai@acm.org

Hironori Washizaki
Waseda University
3-4-1 Okubo Shinuku
Tokyo, Japan
washizaki@waseda.jp

Tatsuya Mori
Waseda University
3-4-1 Okubo Shinuku
Tokyo, Japan
mori@nsl.cs.waseda.ac.jp

ABSTRACT

Permission warnings and privacy policy enforcement are widely used to inform mobile app users of privacy threats. These mechanisms disclose information about use of privacy-sensitive resources such as user location or contact list. However, it has been reported that very few users pay attention to these mechanisms during installation. Instead, a user may focus on a more user-friendly source of information: text description, which is written by a developer who has an incentive to attract user attention. When a user searches for an app in a marketplace, his/her query keywords are generally searched on text descriptions of mobile apps. Then, users review the search results, often by reading the text descriptions; i.e., text descriptions are associated with *user expectation*. Given these observations, this paper aims to address the following research question: *What are the primary reasons that text descriptions of mobile apps fail to refer to the use of privacy-sensitive resources*? To answer the research question, we performed empirical large-scale study using a huge volume of apps with our ACODE (Analyzing COde and DEscription) framework, which combines static code analysis and text analysis. We developed light-weight techniques so that we can handle hundred of thousands of distinct text descriptions. We note that our text analysis technique does *not* require manually labeled descriptions; hence, it enables us to conduct a large-scale measurement study without requiring expensive labeling tasks. Our analysis of 200,000 apps and multilingual text descriptions collected from official and third-party Android marketplaces revealed four primary factors that are associated with the inconsistencies between text descriptions and the use of privacy-sensitive resources: (1) existence of app building services/frameworks that tend to add API permissions/code unnecessarily, (2) existence of prolific developers who publish many applications that unnecessarily install permissions and code, (3) existence of secondary functions that tend to be unmentioned, and (4) existence of third-party libraries that access to the privacy-sensitive resources. We believe that these findings will be useful for improving users' awareness of privacy on mobile software distribution platforms.

Symposium on Usable Privacy and Security (SOUPS) 2015, July 22–24, 2015, Ottawa, Canada.

1. INTRODUCTION

Most applications for mobile devices are distributed through mobile software distribution platforms that are usually operated by the mobile operating system vendors, e.g., Google Play, Apple App Store, and Windows Phone Store. Third-party marketplaces also attract mobile device users, offering additional features such as localization. According to a recent report published by Gartner [1], the number of mobile app store downloads in 2014 are expected to exceed 138 billion. Mobile software distribution platforms are the biggest distributors of mobile apps and should play a key role in securing mobile users from threats, such as spyware, malware, and phishing scams.

As many previous studies have reported, privacy threats related to mobile apps are becoming increasingly serious, and need to be addressed [2, 3, 4, 5]. Some mobile apps, which are not necessarily malware, can gather privacy-sensitive information, such as contact list [6] or user location [7]. To protect users from such privacy threats, many of mobile app platforms offer mechanisms such as permission warnings and privacy policies. However, in practice, these information channels have not been fully effective in attracting user attention. For instance, Felt et al. revealed that only 17% of smartphone users paid attention to permissions during installation [4]. The Future of Privacy Forum revealed that only 48% of free apps and 32% of paid apps provide in-app access to a privacy policy [8]. Further more, Chin et al. reported that roughly 70-80% of end users ignored privacy policies during installation process [9].

Let us turn our attention to a promising way of communicating with users about apps and privacy. This information channel is the *text descriptions* provided for each app in a marketplace. The text description is usually written in natural, user-friendly language that is aimed to attract users' attention; it is more easily understood than the typical privacy policy. In addition, when a user searches for an app in a marketplace, s/he create query keywords, which are generally searched on text descriptions. Then, users review the search results, often by reading the text descriptions; i.e., text descriptions can work as a proxy to the user expectations. In fact, text descriptions have a higher presence than permission warnings or privacy policies, and therefore, are a good channel for informing users about how individual apps gather and use privacy-sensitive information.

With these observations in mind, this work aims to address the following research question through the analysis of huge volume of Android applications:

> **RQ**: *What are the primary reasons that text descriptions of mobile apps fail to refer to the use of privacy-sensitive resources*?

The answers to the question will be useful for identifying sources of problems that need to be fixed. To address the research question, we developed a framework called ACODE (Analyzing COde and DEscription), which combines two technical approaches: static code analysis and text analysis. Using the ACODE framework, we aim to identifiy reasons for the absence of the text descriptions for a given privacy-sensitive permission. Unlike the previous studies, which also focused on analyzing the text descriptions of mobile apps [10, 11, 12, 13], our work aims to tackle with a huge volume of applications. To this end, we adopt light-weight approaches, static code analysis and keyword-based text analysis as described below.

Our static code analysis checks whether a given permission is declared. Then, it investigates whether the code includes APIs or content provider URIs [1] that require permission for accessing privacy-sensitive resources. Lastly, it traces function calls to check that the APIs and/or URIs are actually *callable* to distinguish them from apps with dead APIs/URIs that will never be used; e.g., reused code could include chunks of *unused* code, in which privacy-sensitive APIs were used.

Our description analysis leverages techniques developed in the fields of information retrieval (IR) and natural language processing (NLP) to automatically classify apps into two primary categories: apps with text descriptions that refer to privacy-sensitive resources, and apps without such descriptions. Here we present three noteworthy features of our approach. First, since we adopt a simple keyword-based approach, which is language-independent, we expect that it is straightforward to apply our text analysis method to other spoken languages. In fact, our evaluation through the multilingual datasets demonstrated that it worked for both languages, English and Chinese. Second, although our approach is simple, it achieves a high accuracy for nine distinct data sets. The accuracy is comparable to the existing pioneering work, WHYPER [11], which makes use of the state-of-the-art NLP techniques. The reason we developed the ACODE framework instead of using the WHYPER framework was that we intended to extend our analysis to multiple natural languages. The WHYPER framework leverages API documents to infer semantics. As of today, Android API documents are *not* provided in Chinese. Accordingly, we were not able to make use of the WHYPER framework to analyze Chinese text descriptions. Finally, like the WHYPER framework, our text analysis technique does *not* require manually labeled descriptions. Therefore, it enables us to enhance the text analysis of descriptions to any permission APIs without requiring expensive labeling tasks. It also enables us to reduce cost of text analysis significantly. The key idea behind our approach is to leverage the results of code analysis as a useful hint to classify text descriptions.

To the best of our knowledge, only a few previous studies have focused on analyzing the text descriptions of mobile apps [10, 11, 12]. A detailed technical comparison between these studies and ours is given in section 7 (see Table 10 for a quick summary), and here we note that this work is distinguishable from other studies by being an extensive empirical study. The volume of our dataset is several orders of magnitude larger than previous studies. In addition, because we wanted to extract generic findings, we conducted our experiments in such a way as to incorporate differences in the resources accessed, market, and natural language. Our analysis considered access of 11 different resources taken from 4 categories, i.e., personal data, SMS, hardware resources, and system resources (see Table 1). We chose the resources because they are the most commonly abused or the potentially dangerous ones. We collected 100,000 apps from Google Play and a further 100,000 apps from third-party marketplaces. For the natural language analysis, we adopted English and Chinese, because they are the two most widely-spoken languages worldwide [14]. Furthermore, to evaluate the performance of text analysis, we obtained a total of $6,000$ text descriptions from 12 participants. Each description was labeled by three distinct participants.

The key findings we derived through our extensive analysis are as follows:

The primary factors that are associated with the inconsistencies between text descriptions and use of privacy-sensitive resources are broadly classified into the following four categories.:

(1) **App building services/frameworks**: *Apps developed with cloud-based app building services or app building framework, which could unnecessarily install many permissions, are less likely to have descriptions that refer to the installed permissions.*

(2) **Prolific developers**: *There are a few prolific developers who publish a large number of applications that unnecessarily install permissions and code.*

(3) **Secondary functions**: *There are some specific secondary functions that require access to a permission, but tend to be unmentioned; e.g., 2D barcode reader (camera resource), game score sharing (contact list), and map apps that directly turns on GPS (write setting), etc.*

(4) **Third-party libraries**: *There are some third-party libraries that requires access to privacy-sensitive resources; e.g., task information (crash analysis) and location (ad-library, access analysis).*

The main contribution of our work is the derivation of these answers through the extensive analysis of huge volume of datasets. We believe that these findings will be useful for identifying sources of problems that need to be fixed to improve the users' awareness of privacy on mobile software distribution platforms. For instance, as our analysis revealed, there are several HTML5-based app-building framework services that unnecessarily install permissions, which could render the system vulnerable to additional threats of malicious JavaScript injection attacks. Therefore, an app developer should not install unnecessary permissions. However, if a developer used a rogue app-building framework service, he/she may likely not be aware of unnecessary permissions installed. ACODE enables operators of mobile software distribution platforms to pay attentions to these cases, which are invisible otherwise.

The rest of this paper is organized as follows. Section 2 describes our the ACODE framework in detail. In section 3, we show the details of the static code analyzer. Section 4 contains details of the text description classifier. We present our findings in section 5. Section 6 discusses the limitations of ACODE and future research directions. Section 7 summarizes the related work. We conclude

[1] Content providers manage access to data resource with permission using Uniform Source Identifiers (URIs); for instance, `android.provider.ContactsContract.Contacts.CONTENT_URI` is an URI used to get all users registered in the contact list.

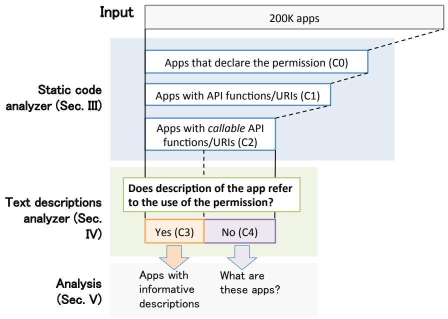

Input
200K apps
Static code analyzer (Sec. III)
Apps that declare the permission (C0)
Apps with API functions/URIs (C1)
Apps with *callable* API functions/URIs (C2)
Text descriptions analyzer (Sec. IV)
Does description of the app refer to the use of the permission?
Yes (C3)
No (C4)
Analysis (Sec. V)
Apps with informative descriptions
What are these apps?

Figure 1: Overview of the ACODE framework.

Table 1: List of permissions used for this work.

Category	Permission	Definition*
Personal data	ACCESS_FINE_ LOCATION	Allows an app to access precise location from location sources such as GPS, cell towers, and Wi-Fi.
	GET_ACCOUNTS	Allows access to the list of accounts in the Accounts Service.
	READ_CONTACTS	Allows an application to read the user's contacts data.
	READ_CALENDAR	Allows an application to read the user's calendar data.
SMS	READ_SMS	Allows an application to read SMS messages.
	SEND_SMS	Allows an application to send SMS messages.
Hardware resources	CAMERA	Required to be able to access the camera device.
	RECORD_AUDIO	Allows an application to record audio.
System resources	GET_TASKS	Allows access to the list of accounts in the Accounts Service . (This constant was deprecated in API level 21)
	KILL_BACKGROUND_ PROCESSES	Allows an application to call killBackgroundProcesses(String).
	WRITE_SETTINGS	Allows an application to read or write the system settings.

*http://developer.android.com/reference/android/Manifest.permission.html

our work in section 8.

2. ACODE FRAMEWORK

In this section, we provide an overview of the ACODE framework. We also connect the components of the ACODE framework to the corresponding sections where we will give their details.

2.1 Goal and overview

Figure 1 is an overview of the ACODE framework. As discussed previously, we used a two-stage filter, employing a static code analyzer and text descriptions analyzer. In the first stage, the first filter extracted apps that declare at least one permission, e.g., location (C0). The second filter extracted apps with code that include corresponding APIs/URIs (C1). The third filter checked whether the APIs/URIs are *callable* from the apps by employing function call analysis (C2). In the second stage, the text classifier determined whether the text descriptions refer to the use of location explicitly or implicitly (C3), or not at all (C4). Note that we are not considering apps that do not declare to use permission, but have descriptions that indicate that permission is needed.

These filtration mechanisms enabled us to quantify the effectiveness of text descriptions as a potential source of information about the use of privacy-sensitive resources. For instance, by counting the fraction of apps that are classified as C3 (see figure 1), we can quantify the fractions of apps with text descriptions that successfully inform users about the use of privacy-sensitive resources for each resource. By examining the sources of apps that are classified as C4, we can answer our research question, **RQ**. The detailed analysis will be shown in section 5.

Figure 2 illustrates the components used in the ACODE framework. For each application, we had an application package file (APK) and a description. APK is a format used to install Android application software. It contains code, a manifest file, resources, assets, and certificates. The text descriptions of apps were collected from mobile software distribution platforms. As shown in the figure, the APKs and text descriptions were input to the static code analyzer and description classifier, respectively.

2.2 Static code analyzer

The goal of the static code analyzer is to extract APK files whose code include *callable* APIs/URIs that are required to use permissions related to a privacy-sensitive resource. For a given permission, first, we extracted apps that declare the use (C1, see section 3.1). Then, we checked whether disassembled code of the app include the APIs/URIs, which require the permission (C2, see section 3.2). If code included at least one API or URI, then, we checked whether it was actually *callable* within the app by investigating the function call graph with some heuristics we developed (C3, see section 3.3). It should be noted that the static code analysis has some limitations that we will discuss in section 6.

2.3 Description classifier

The goal of the description classifier was to classify text descriptions into two categories: those that refer to the use of a resource (C3), and those that do not (C4). In other words, we wanted to determine automatically whether a user can, by reading the text description, know that an app may use a privacy-sensitive resource. To do this, we leveraged several text analysis techniques. We also make use of the results of code analyzer to extract keywords associated with a resource. To extract keywords that are useful in classifying text descriptions, we first present text data preprocessing techniques in section 4.1. Next, in section 4.2, we present the keyword extraction method that leverages techniques used in the field of information retrieval. We also evaluate the accuracy of the description classifier in 4.3.

3. STATIC CODE ANALYSIS

This section describes the static code analysis techniques used in the ACODE framework. The purpose of static code analysis was to extract apps that include *callable* APIs/URIs to use a given permission. Before applying function call analysis, which is a process of checking whether given function is callable, we applied two filtration mechanisms: (1) permission filtration and (2) API/URI filtration. These filtrations are effective in reducing the computation overhead needed for function analysis. We also note that permission filter is useful to prune apps that include callable APIs/URIs, but will not actually use it.

3.1 Permission filtration

First, we applied permission filtration, which simply checks whether an app declares a given permission. According to Zhou et al. [15], permission filtration is quite effective in reducing the overhead of analyzing a huge amount of mobile apps. For each app, we investigated its AndroidManifest.xml file to check whether it declares permissions to access given resources. The process can be easily automated using existing tools such as `aapt` [16]. To further accelerate the data processing, we also leveraged multiprocessing techniques. Table 1 summarizes the 11 different permissions we analyzed in this work. To perform generic analysis, we chose the permissions from 4 categories, personal data, SMS, hardware resources, and system resources. These resources were chosen because they are the most commonly abused or the potentially dangerous ones.

3.2 API/URI filtration

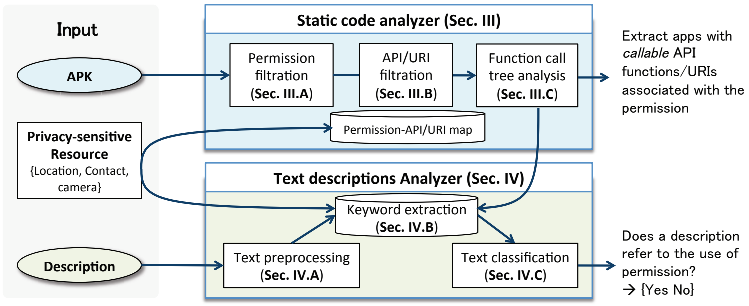

Input
APK
Privacy-sensitive Resource
{Location, Contact, camera}
Description
Static code analyzer (Sec. III)
Permission filtration (Sec. III.A)
API/URI filtration (Sec. III.B)
Function call tree analysis (Sec. III.C)
Permission-API/URI map
Extract apps with *callable* API functions/URIs associated with the permission
Text descriptions Analyzer (Sec. IV)
Keyword extraction (Sec. IV.B)
Text preprocessing (Sec. IV.A)
Text classification (Sec. IV.C)
Does a description refer to the use of permission?
→ {Yes No}

Figure 2: Components of the ACODE framework.

Next, for each sample, we checked whether it includes APIs or content provider URIs that require permissions to access privacy-sensitive resources. For this task, we made use of the API calls for permission mappings extracted by a tool called PScout [17], which was developed by Au et al. [18]. In addition to API-permission mapping, the PScout database also includes URI-permission mapping. To check the existence of APIs or URIs, first, using Android apktool [19], we extracted DEX code from APK files and disassembled them into smali format [20]. Then, we checked whether a set of APIs is included in the code of an APK file.

We note that some apps may require permissions but not include any APIs or URIs that request the permission. This may occur for several reasons. apps. If such possibly overprivileged apps are simply overprivileged due to developer's error, they do not impact our study, because those apps may not need to use APIs or URIs. However, as Felt et al. [3] reported, one of the common developer errors that cause overprivilege is Intent. A sender application can send an Intent to a receiver application, which uses permission API. In such cases, the sender of the Intent does not need to have permissions for the API. We saw many such cases, especially related to camera permissions. In fact, [3] reported that of the apps that unnecessarily request camera permission, 81% send an Intent to open the already installed camera applications (including the default camera) to take a picture. Our observation is in agreement with their finding.

Thus, our API/URI filtration scheme may miss a non-negligible number of apps that actually use the camera through Intent. However, note that our final analysis will be applied to the apps in set $C2$ as shown in figure 1. Therefore, we are confident that the removal of such apps should not affect our analysis, because we do not expect to see significant differences between the descriptions of those apps removed due to the Intent problem and the descriptions of apps included in $C2$.

3.3 Function call analysis

Now, we present the function call analysis of the ACODE framework. For convenience sake, let the term function include method, constructor execution, and field initialization; i.e., we trace not only method calls, but also class initializations. Figure 3 presents a pseudo-code of the algorithm we developed for function call analysis. It checks whether APIs/URIs of a given permission are callable (true) or not (false). The algorithm uses depth-first search to search the function call tree. If it finds a path from the given function to a class of `ORIGIN` (line 4), it concludes that the app has at least one API/URI that is callable, where `ORIGIN` is composed of three classes: `Application`, `App Components`, and `Layout`. `Application` is a class that initiates an Android app. It is called when an app is launched. `App Components` are the essential building blocks that define the overall behavior of an Android app, including `Activities`, `Services`, `Content providers`, and `Broadcast receivers`. While the `Application` and `App Components` classes need to be specified in the manifest file of an app, the `Layout` class does not. It is often used by ad libraries to incorporate ads using XML.

```
 1: INPUT
 2: p : a permission
 3: a : an application (APK)
 4: ORIGIN = [Application, App Components, Layout]
 5: list = getAU(p,a)  # list of APIs/URIs associated with p
 6: done = []            # empty list
 7:
 8: WHILE list is not empty DO
 9:     f = list.pop()
10:     IF f is in done:
11:         skip the function
12:     ENDIF
13:     IF f.parentClass is in ORIGIN:
14:        RETURN True
15:     ENDIF
16:     IF (f.parentClass inherits Android SDK)
17:         AND (f is not init)
18:         AND (f is not a static method):
19:         list.append(f.parentClass.init)
20:     ELSE IF (f is referenced):
21:         list.append(f.refFunctions)
22:     ENDIF
23:     done.append(f)
24: ENDWHILE
25: RETURN False
```

Figure 3: Pseudo-code that checks the callability of APIs of a permission.

`getAU` (Line 5) is a function that returns a list of APIs/URIs for a given permission. As an implementation of getAU, we adopted PScout [17]. `refFunctions` (line 21) is a function that returns a list of functions that reference to the given function or URI. As an implementation of `refFunctions`, we adopted androguard [21], which we modified to handle URIs. If a function of a class, say Foo, implements a function of the Android SDK class whose code is not included in the APK, we cannot trace the path from the function in some cases. To deal with such cases, we made a heuristic to trace the function that calls the init-method of class Foo (lines 16–19). We note that the heuristics can handle several cases such as async tasks, OS message handlers, or callbacks from framework APIs such as `onClick()`. A method is callable if it is overridden in a subclass or an implementation of the Android SDK and an instance of the class is created. Async tasks, the OS message handler, or other callbacks implement their function by overriding the methods of the Android SDK subclass. Therefore, it should be handled by the heuristics.

4. TEXT DESCRIPTION ANALYSIS

This section describes the text description analysis used in the ACODE framework. The aim of this analysis was to classify descriptions into two classes: (1) text descriptions that reference a privacy-sensitive resource, and (2) text descriptions that do not. To this end, we adopted a set of basic techniques used in both IR and NLP fields. As we shall see shortly, our keyword-based approach is quite simple and works accurately for our task. As Pandita et al. [11] reported, a keyword-based approach could result in poor performance if it was designed naively. So, we carefully constructed our keyword extraction processes. As a result, we achieved 87-98% of accuracy for the combinations of 3 resources and two languages. Simple and successful text description classification enabled us to automate the analysis of 200,000 text descriptions.

Section 4.1 describes how we preprocessed the description data so that we can extract keywords that are useful in classifying text descriptions. Section 4.2 presents the keyword extraction method that leverages techniques used in the field of information retrieval. Section 4.3 describes our experiments to compare our description classifier with the WHYPER framework in terms of accuracy.

4.1 Text Data Preprocessing

To analyze natural language text descriptions, we applied several text preprocessing methods. These methods are broadly classified into four tasks; (1) generic text processing, (2) domain-specific stop words removal, (3) feature vector creation, and (4) deduplication. Especially the tasks (2) and (4) are crucial in extracting good keywords that can accurately classify the text descriptions.

4.1.1 Generic text preprocessing

We first apply widely-used generic text preprocessing techniques: word segmentation, stemming, and generic stop words removal. Word segmentation is a process of dividing text into words. This process is required for Chinese but not for English, in which words are already segmented with spaces. We used KyTea [22] for this task. For English, we applied stemming, which is a process of reducing derived words to their stem. It is known to improve the performance of text classification tasks. We used NLTK [23] for this task. Note that the concept of stemming is not applicable to Chinese. Lastly, we applied generic stop words removal, which is a process of removing a group of words that are thought to be useless for classification tasks because they are commonly used in any documentation (e.g., determiners and prepositions). As lists of stop words, we used the data in NLTK [23] for English and the data in imdict [24] for Chinese.

4.1.2 Domain-specific stop words removal

Next, we created domain-specific stop words list so that we can remove terms that are not generic stop words but are commonly used in mobile app descriptions; e.g., “app” or “free”. To this end, we make use of the technique proposed in Ref. [25], which is a term-based sampling approach based on the Kullback-Leibler divergence measure. Since the technique measures how informative a term is, we can remove the least weighted terms as the stop words. Number of sampling trial was set to 10,000. When we changed the threshold of extracting the top-L stop words; i.e., from $L = 20$ to $L = 150$, the following results are not affected at all. In the followings, we use $L = 100$. The extracted domain-specific stop words for English include “app”, “free”, “get”, “feature”, “android”, “like”, etc. Top-100 domain-specific stop words for English and Chinese are listed in Table 11 in the Appendix.

4.1.3 Feature vector creation

Using the preprocessed descriptions, we created a binary feature vector for each text description as follows. Let $\mathbf{W} = \{w_1, w_2, ..., w_m\}$ be a set of entire words after the screening process shown above. A feature of vector of the ith text description is denoted as $\mathbf{x}_i = \{x_i(w_1), x_i(w_2), ..., x_i(w_m)\}$, where $x_i(w_j) = 1$ if w_j is present in the ith text description. If w_j is not present, $x_i(w_j) = 0$.

4.1.4 Deduplication

Because we adopt the keyword extraction approach based on *relevance weights* as shown in the next subsection, the deduplication process plays a crucial role in eliminating the effect of same or similar descriptions generated by a single developer. For instance, if a developer produces thousands of apps with the same text description, which is often the case we observe in our datasets, the words included in the apps may cause unintended biases when computing the relevance weights of terms. To deduplicate the descriptions, we remove the same or similar descriptions by using the cosine similarity measure; i.e., for a given pair of feature vectors $\mathbf{x}_i$ and $\mathbf{x}_j$, the cosine similarity is computed as $s = \cos\left(\mathbf{x}_i \cdot \mathbf{x}_j / |\mathbf{x}_i||\mathbf{x}_j|\right)$, and if s is larger than a threshold, the duplicated description is removed. We note that the value of threshold was not sensitive to the succeeding keyword extraction results if it is set between 0.5 to 0.8.

4.2 Keyword Extraction

To extract keywords, we leverage the idea of *relevance weights*, which measures the relation between the relevant and non-relevant document distributions for a term modulated by its frequency [26]. Relevance weighting was developed in the IR community as a means to produce optimal information retrieval queries. To make use of the relevance weights for our problem, we need to have sets of relevant and non-relevant documents. Since we do not have any labels that indicate whether a document is relevant, i.e., it refers to a permission, or non-relevant, i.e., it does not refer to a permission, we set the following assumption.

Assumption: *For a given permission, descriptions of apps that declare the permission and have callable APIs can be regarded as “pseudo relevant document”, while the descriptions of the remaining apps can be regarded as “pseudo non-relevant document”.*

Note that our research question contradicts with this assumption; i.e., we are interested in the reason why an app with callable API for a permission does not refer to the permission. Nevertheless, our performance analysis using multiple permissions in two spoken languages empirically supports that our approach actually works well in extracting effective keywords.

Under this assumption, we calculate the relevance weights for each word as follows. For a word w_i, the relevance weight (RW) is

$$RW(w_i) = \log \frac{(r_i + 0.5)(N - n_i - R + r_i + 0.5)}{(n_i - r_i + 0.5)(R - r_i + 0.5)},$$

where r_i is the number of relevant documents word w_i occurs in, R is the number of relevant documents, n_i is the number of documents word w_i occurs in, and N is the number of documents, respectively.

Using the entire descriptions with code analysis outputs, we extracted the keywords that have the largest relevance weights. Table 2 presents a subset of extracted keywords for each permission. For space limitation, we present only the Top-3 English keywords. We have listed the top-10 keywords for English and Chinese in Table 12 in the Appendix. In most cases, the keywords look intuitively reasonable. Interestingly, some keywords such as “sms”

Table 2: Extracted top-3 keywords for English descriptions.

Resources	1st	2nd	3rd
Location	gps	location	map
Account	grab	google	youtube
Contact	sms	call	contact
Calendar	calendar	reminder	meeting
SMS (read)	sms	message	incoming
SMS (send)	sms	message	sent
Camera	camera	scan	photo
Audio	recording	voice	record
Get tasks	lock	security	task
Kill background process	task	kill	manager
Write setting	alarm	ring	bluetooth

are found in multiple resources; i.e., contact, SMS (read), and SMS (send). In fact, these resources tend to co-occur. In the following, we will use these keywords to classify descriptions. Once we compiled the keywords, the text classification task is straightforward. If a text description includes one of the extracted keywords for a permission, the description is classified as positive, i.e., it refers to the permission. The problem is how we set the number of keywords to be used. We will study the sensitivity of the threshold in Section 4.3.2.

4.3 Performance Evaluation

To evaluate the accuracy of our scheme, we use manually labeled data sets. We first present the way how we compile the labeled data set. Next, we evaluate the accuracy of our approach, using the labeled data. Finally, to validate the robustness of our approach, we use the external dataset and compare the performance with the existing state-of-the-art solution, the WHYPER framework. In the analysis of accuracy (Section 4.3.3), we use 200,000 apps, which will be described in Section 5.1 as *training sets*; i.e., they are only used for keyword extraction. The *labeled test set* is a subset of those, on which we measure accuracy. We note that in the evaluation, our training set included test set; i.e., we extracted the keywords using the entire text descriptions, which is the training set, and applied the keywords (i.e., classifier) to the labeled descriptions, which is the test set. In general, training classifier using test set is not good because such setting could over-estimate the accuracy of the model. However, the effect should be small because our classifier was based on frequencies of terms and the test set accounted for only 0.6% of entire samples.

4.3.1 Creation of labeled datasets

We created the labeled data sets with the aid of 12 international participants who are from China, Korea, Thailand, and Indonesia. All the participants were university students with different disciplines in science and engineering. 7 were female and 5 were male. 4 were native English speakers, and 8 were native Chinese speakers. None of them had experience of developing Android applications. All the native Chinese speakers were fluent in English (native level). Students who were native speakers of Chinese labeled Chinese descriptions. In summary, six students labeled English descriptions, and the other six labeled Chinese descriptions. Here, we picked up three distinct resources, i.e., location, contact, and camera, out of the 11 resources we considered in this work.

Since a resource is used for various purposes, and referred to by various terms, we wanted to avoid participants focusing too much on a particular keyword, such as “camera”. Instead, we asked participants to identify whether an app will use a camera, rather than whether it mentions a camera. This enabled us to identify several interesting keywords, such as “QR” and “scan”. Also, we note that the question should reflect users’ *awareness* of a resource.

Table 3: Summary of labeled datasets.

English			
	Location	Contact	Camera
# of descriptions	1,000	1,000	1,000
# of labels	3,000	3,000	3,000
Chinese			
	Location	Contact	Camera
# of descriptions	1,000	1,000	1,000
# of labels	3,000	3,000	3,000

Table 4: Statistics of labeled descriptions to be used for performance evaluation.

English			
	Location	Contact	Camera
# of positive descriptions	128	208	276
# of negative descriptions	611	449	289
Chinese			
	Location	Contact	Camera
# of positive descriptions	38	102	157
# of negative descriptions	828	544	583

Before asking participants to label text descriptions, we picked some descriptions from our entire data set. If random sampling were applied to the entire set, there would be a significant imbalance between the two classes. In particular, there would be very few positive samples, i.e., text descriptions that reference a resource. To avoid such an imbalance, we applied the access permission filter shown in section 3.1 so that the sampled text descriptions would include a certain number of positive samples. Although this solution could create some bias toward the positive class, in fact it did not matter, as will be shown later in this paper. From the set of apps that declare access permissions for using resources, we randomly sampled 1,000 text descriptions. In total, we sampled 6,000 descriptions, as shown in table 3.

Having sampled text descriptions, we asked each participant to label 500 text descriptions for each resource (e.g., $500 \times 3 = 1,500$ descriptions in total). A participant labeled text descriptions in either English or Chinese. To increase the quality of labels, each text description was labeled by three distinct, fixed participants. We obtained a total of $18,000$ labels for $6,000$ text descriptions, as shown in table 3.

Finally, we eliminate inconsistent labels to ensure that the quality of labels is high; i.e., we used only the text descriptions upon which all three evaluators agreed. Table 4 summarizes the text descriptions that met this criterion. We used these labeled descriptions for evaluating accuracy of our approach, as described in the next subsection.

4.3.2 Threshold Sensitivity Study

Using the labeled datasets, we empirically studied the relation between threshold and classification accuracy. Here, the definition of the accuracy is the fraction of correctly classified text descriptions, using the top-K keywords. Figure 4 presents how the number of keywords, K is correlated with the classification accuracy. As shown in the graph, across the 6 of labeled datasets, the accuracy is fairly stable around $K = 3$. Also, we notice that $K = 3$ gives the highest accuracy with the minimum variance. As we increase K, the accuracy is degraded; i.e., as K increases, the less relevant the keywords become. In fact, while many of keywords listed in

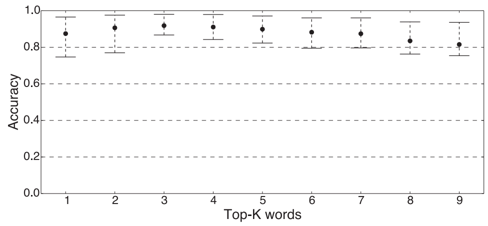


Figure 4: K vs. accuracy. The circles indicate median values and the bars indicate maximum/minimum values, respectively.

Table 5: Accuracy of our approach (K = 3) for the 6 of labeled datasets.

Resource	Lang	TP	TN	FP	FN	ACC	PPV	NPV
Location	EN	118	591	20	10	0.959	0.855	0.983
	CN	23	826	2	15	0.980	0.920	0.982
Contact	EN	177	396	53	31	0.872	0.770	0.927
	CN	64	535	9	38	0.927	0.877	0.934
Camera	EN	206	284	5	74	0.867	0.976	0.802
	CN	98	575	8	59	0.909	0.925	0.907

Table 12 in the Appendix look natural, some lower-ranked keywords such as "gps" for SEND_SMS or "call" for READ_CALENDAR do not really make sense. Given these observations, in the following analysis, we adopt $K = 3$ in classifying the document. We note that the chosen threshold works nicely for the external dataset provided by the authors of WHYPER [11]. We will report the results in Section 4.3.4.

4.3.3 Accuracy of Text Classification

We now evaluate the accuracy of our text classifier. To measure the accuracy, we use several metrics. First, TP, TN, FP, and FN represents number of true positives, number of true negatives, number of false positives, and number of false negatives, respectively. We also use three derivative metrics: accuracy (ACC), Positive predictive values (PPV), and Negative predictive values (NPV), which are defined as

$$\begin{aligned} ACC &= \frac{TP+TN}{TP+TN+FP+FN}, \\ PPV &= \frac{TP}{TP+FP},\ NPV = \frac{TN}{TN+FN}, \end{aligned}$$

respectively. PPV and NPV measure how many of descriptions classified as positive/negative are actually positive/negative. These measures are suitable to our requirements because we aim to derive the answers of our research question by studying the characteristics of classified descriptions. Therefore, we expect that these measures have high values.

Table 5 presents the results of performance evaluation. In both languages, the observed accuracy was good for all categories; e.g., ACCs were 0.87–0.98. Also, in most cases, NPVs were larger than 0.9. Since one of our objectives is to understand the reasons why text descriptions fail to refer to access permissions, the high number of NPVs is helpful, because it indicates that majority of descriptions classified as negative are actually negative. In summary, our scheme was validated to enable automatic classification of text descriptions into the two categories with good accuracy. It works well for both languages, English and Chinese.

4.3.4 Robustness

Table 6: Statistics of the WHYPER datasets.

	Contact	Calendar	Audio
# of positive samples	107	86	119
# of negative samples	83	110	81

Table 7: Comparison of accuracy of ACODE (K = 3), WHYPER semantic analysis (WHYPER), and WHYPER keyword (WKW).

Resource	method	TP	TN	FP	FN	ACC	PPV	NPV
Contact	ACODE	96	63	20	11	0.837	0.828	0.851
	WHYPER	92	77	6	15	0.889	0.939	0.837
	WKW	95	46	37	12	0.742	0.720	0.793
Calendar	ACODE	77	98	12	9	0.893	0.865	0.916
	WHYPER	81	99	11	5	0.918	0.880	0.952
	WKW	84	60	50	2	0.735	0.627	0.968
Audio	ACODE	95	57	24	20	0.742	0.720	0.793
	WHYPER	103	69	12	16	0.860	0.896	0.812
	WKW	113	38	43	6	0.755	0.724	0.864

To validate the robustness of our approach, we use the external labeled dataset [27], which is provided by the authors of the WHYPER framework [11]. Since the dataset also includes the outcomes of the WHYPER framework, we can directly compare the performance of the two frameworks. Since the dataset consists of a set of labels for each sentence, we reconstructed original descriptions from the sentenses and assign labels to the descriptions; i.e., if a description consists of at least one sentence that declares the use of a permission, the description is labeled as positive, otherwise labeled as negative. Table 6 summarizes the dataset[2]. All the descriptions are written in English.

Table 7 shows the comparison of performance of the ACODE framework and the WHYPER framework in classifying descriptions. Our results show that the performance of the ACODE framework is comparable with that of the WHYPER framework. Especially, the delta for NPV, which is the most important metrics for our study, is less than 0.04 for all the three cases. We also notice that the keyword-based approach used in the WHYPER paper (WKW in the table) had high false positives. We conjecture that the high false positives are due to the nature of extracted keywords, which include some generic terms such as data, event, and capture.

Notice that the WHYPER dataset consists of higher fractions of positive descriptions, compared to ours. This may reflect the fact that the apps used for WHYPER study were collected from the top-500 free apps; i.e., it is likely the top apps were built by skilled developer and had informative descriptions. In contrast, our datasets consist of larger fractions of negative samples. Since our datasets were collected from entire app space, they consist of various apps, including the ones that failed to add informative descriptions due to the reasons that will be described in the next section. Despite this potential difference in the population of datasets, our framework established good accuracy among all the datasets.

In summary, we evaluated the accuracy of the ACODE framework using 5 of 11 permissions we considered[3]. In the following large-scale analysis, we assume that the ACODE framework establishes good accuracy for the rest of permissions as well. The

[2] We derived these numbers by analyzing the dataset [27]

[3] To be precise, we verified 5 of 11 permissions for English and 3 of 11 permissions for Chinese.

Table 8: Summary of Android apps used for this work.

	English	Chinese	Data collection periods
Official (Google Play)	100,000	0	Apr 2012 – Apr 2014
Third-party (Anzhi)	0	74,506	Nov 2013 – Apr 2014
Third-party (Nduoa)	0	25,494	Jul 2012 – Apr 2014

potential effect of the assumption will be discussed in Section 5.5.

5. ANALYSIS OF CODES AND DESCRIPTIONS

Using the ACODE framework, we aim to answer our research question **RQ** shown in Section 1. We first describe the details of the data sets we used for our analysis, in section 5.1. Then, we apply our code analysis to the apps and extract apps with *callable* APIs/URIs of permissions ($C2$, see figure 1) in section 5.2. Using the extracted apps with callable APIs/URIs of permissions, section 5.3 aims to quantify the fractions of apps with text descriptions that successfully inform users about the use of privacy-sensitive resources for each resource. In section 5.4 we aim to answer the research question **RQ**. We discuss in-depth analysis to understand the reasons of failures for text descriptions classified as $C4$ in informing users about access permissions. Finally, Section 5.5 discusses the limitations of our analysis and evaluation.

5.1 Data sets

We collected Android apps from the official marketplace [28] and two other third-party marketplaces [29, 30]. All these marketplaces have huge user bases. Note that these were all free apps. Although we might see some disparity between free and paid apps, we leave this issue open for future research.

After collecting mobile apps, we first pruned samples that are corrupt or have zero length text descriptions. From the rest of the samples, we randomly picked 100,000 apps for each type of markets. Table 8 summarizes the data sets we collected. Among 200,000 apps, only 1,831 apps were duplicated in package names between the two markets. To simplify the interpretation of analyses, we assigned different languages, English and Chinese, to the official and third-party marketplaces. Note that we have already shown that our text description classification scheme works well for both languages.

5.2 Extracting apps with callable APIs/URIs of privacy-sensitive resources

Table 9 presents the results of our code analysis. Overall, many applications require permission of location. As we will detail later, many of these are apps that use ad libraries. Interestingly, the popularity of personal data resource requirements is almost identical across markets. The most popular is location, second is contact, third is accounts, and fourth is calendar. Generally, third-party markets tend to require/use more permissions than the official market. This may correlate to the existence of defense mechanisms installed on the official marketplace – Bouncer [31].

Another useful finding we can extract from the results is that over privilege ($C0 - C1$) is observed commonly across the categories. Also, there are non-negligible numbers of apps that have code to use permissions but cannot be called ($C1 - C2$). This often occurs when a developer incorporates an external library into an app; the library has many functions, including APIs/URIs of permissions, but the app does not actually call the APIs/URIs. Our code analysis can prune these applications from further analysis.

Overprivilege ratios are especially high for account and contact permissions in the third party marketplaces and for camera, calendar, and kill background processes permissions in both markets. Careful manual inspection revealed that these cases can be attributed to misconfiguration on the part of developers; i.e., the Intent issue discussed in section 3.2. Such apps were pruned by the second filter. We also note that these apps do not need to declare permissions because the permissions are misconfigurations. These observations agree with the work performed by Felt et al. [3]. Although our scheme pruned those applications, the pruning did not affect the analysis because the pruned apps are unlikely to exhibit special characteristics in their text descriptions.

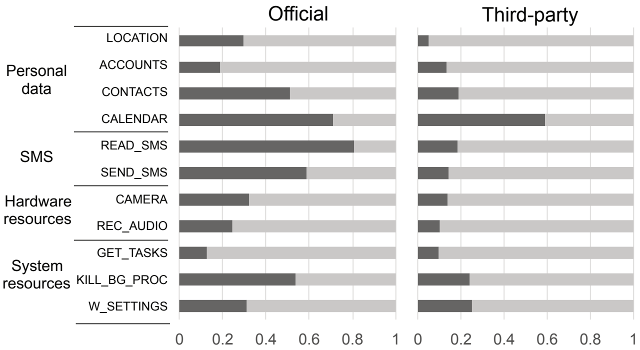


Figure 5: Fractions of descriptions that refer to a permission. Populations are $C2$ apps shown in Table 9; e.g., of the 18,165 of official market apps with callable functions that request location permission, roughly 30% of them mentioned the use of location in the description.

5.3 Analysis of apps with callable APIs/URIs for a permission.

Using apps that include callable APIs/URIs for a permission ($C2$ in Table 9), we analyzed their text descriptions. Figure 5 presents the results. We first notice that fractions of positive text descriptions are higher for official market apps. This can be considered natural, given that official market is more restrictive. We also notice that some resources such as `CALENDAR` for both markets and `SMS` permissions and the `KILL_BG_PROC` (kill background process) permission for the official market are well described in their descriptions.

For the official market, `GET_TASK` and `ACCOUNTS` were the permissions that were less described (15–20%). In contrast, `READ_SMS` and `CALENDAR` were the permissions that were well described (70–80%). These results are consistent with intuition that permissions that are directly associated with user actions tend to be well described. Overall, our impression is that for the official market, the fractions of proper descriptions are higher than expected. Thus, if the descriptions of remaining apps were improved, the text description could serve as a good source of information to let users know about sensitive resources.

Finally, we note that the descriptions of apps collected from official market was only English, while the descriptions of apps collected from third-party market was only Chinese. Therefore, we cannot tell if the observed differences are due to the market or the language. We leave the issue for future work.

5.4 Answers to the Research Question

To answer the research question **RQ**, we performed the manual inspection to the extracted apps that fail to refer to use of permis-

Table 9: Numbers of extracted apps for each category.

Official market apps

	Personal data				SMS		Hardware resources		System resources		
	Location	Accounts	Contacts	Calendar	SMS (read)	SMS (send)	Camera	Audio	Get tasks	Kill bg processes	Write setting
Permission (C0)	25026	9943	6962	1893	1352	3471	10232	5204	4646	409	2873
API/URI (C1)	23390	6948	6177	333	526	2043	7173	4621	3433	248	1954
Callable (C2)	18165	3933	4238	100	287	1567	6141	3297	1737	208	1744

Third-party market apps

	Personal data				SMS		Hardware resources		System resources		
	Location	Accounts	Contacts	Calendar	SMS (read)	SMS (send)	Camera	Audio	Get tasks	Kill bg processes	Write setting
Permission (C0)	40278	6585	9907	394	7686	16204	14581	10745	37436	7457	15249
API/URI (C1)	36885	3148	4863	98	4668	13807	6934	8354	19147	1158	11564
Callable (C2)	32122	1542	3429	66	4185	12355	6139	6147	15447	957	1029

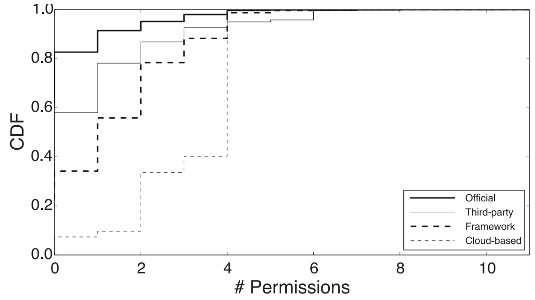


Figure 6: CDFs of number of permissions per application. The 11 permissions listed in Table 1 are used.

sions. The methodologies of the manual inspection are described below. Given a permission, e.g., `Camera`, we fist identify Java classes that include the APIs associated with the permission. From the identified class, we can extract a package name such as `/com/google/android/foo/SampleCameraClass.java`, which is segmented into a set of words, com, google, android, foo, and SampleClass. By analyzing the package name words for apps that fail to refer to use of the permission, we can find intrinsic words that are associated with specific libraries such as "zxing" used for handling QR code or service names such as "cordova", which is an app building framework. In addition, we can analyze developer certificates included in app packages. We also apply dynamic analysis of the apps when we need to check how the permission is used. Using the methodologies, we classified such apps into the four categories. For each category, we extracted reasons why text descriptions fail to refer to permissions.

(1) **App building services/frameworks**

Through the analysis of package names of apps, we noticed that many of apps were developed with cloud-based app building services, which enable a developer to create a multi-platform app without writing code for it. Examples of cloud-based app building services are SeattleClouds, iBuildapp, Appsbar, appbuilder, and biznessapps. Similarly, many of apps were developed with mobile app building frameworks, which also enable a developer to create a multi-platform app easily. Examples of such mobile app building frameworks are Apache Cordova (Phonegap) and Sencha. These services/frameworks provide a simple and intuitive interface to ease the processes of building a mobile app.

Among many such services/frameworks, we found a few services/frameworks that generate apps that *unnecessarily* install many permissions, and put callable APIs/URIs for the permissions into the code. Since a developer using such a service/framework cannot change that setting, it is likely that even the developer is not aware of the fact that app install the permissions with callable APIs/URIs; hence, it is less likely the developer writes about the permissions in the description.

Figure 6 shows CDFs of number of permissions per application. First, apps collected from the official market have small number of permissions among the 11 permissions; i.e., more than 80% of apps had zero permissions. They had other generic permission such as Internet. Second, we considered an intrusive cloud-based app building service and one of the popular app building frameworks. Both cases tend to install a large number of permissions. Especially, roughly half of the apps that were built with the intrusive cloud-based app building service had a fixed number of permissions (4 out of 11). We carefully inspected these apps, and found that many permissions such as record audio were unnecessarily installed by the services/frameworks.

We revealed that the apps built by the intrusive cloud-based app building services are popular in official market, but not popular in third-party market. In the official market, more than 65% of apps that failed to refer to use of `record audio` were developed with these services. Similarly, more than 25% of apps that failed to refer to use of `contact list` were developed with these services. We also observed non-negligible number of such apps in other resources; i.e., 5% for location and 10% of camera. For app building frameworks, one of the frameworks accounted for more than 28% of apps that failed to refer to use of `record audio` in the third-party market. In fact the permission was not necessary for the apps.

We also note that unnecessarily installed permissions on a framework such as phonegap, which is HTML5-based mobile app building framework, could bring additional threats because such permission can be abused through various channels of Cross-Site Scripting attacks [32].

(2) **Prolific developers**

Through the analysis of distributions of number of apps per developer certificate, we noticed that a very few number of developers accounted for a large number of descriptions without mention of privacy-sensitive resources. We call such developers "prolific developers". For instance, five prolific developers published 47% of third-party market apps that fail to refer to `send SMS`. We applied eleven popular commercial anti-virus scanners to the apps with SMS permission, and checked whether either of scanner detected the types of application. If at least one scanner detected an app as malware/adware, we marked it as malware/adware. We found that majority of the apps with unmentioned SMS permission were malware/adware and have been removed from the market later. There are other cases. Three prolific developers published 38% of third-party market apps that fail to use of `kill background processes`. Another three prolific developers published 32% of third-party market apps that fail to use of `write setting`. We carefully inspected these apps, and found that they do not have any reasons to use the permissions. Although not conclusive, we conjecture that these prolific developers likely reuse their own code for building a large number of apps; i.e., they tend to include unnecessary permissions/code.

(3) **Secondary functions**

Through the careful analysis of descriptions that failed to refer to permissions, we found several secondary functions that tend to be unmentioned. For instance, several apps have functions to share information with friends, e.g., scores of games. In many cases, such functions require to access contact list. However, such activity is often unmentioned in the descriptions because it is an optional function. Another example is map-based apps that require to access the `write setting` permission to enable location positioning service such as GPS or Wi-Fi. Such map-based apps accounted for 44% of apps that failed to refer to `write setting`. Among several cases, the most notable one was barcode reader, which requires access to `camera` device. Although there are several barcode reader apps, majority of apps with barcode reader function are shopping apps or social networking apps. Since the barcode reader is not a primary function for those apps, it tends to be unmentioned in their descriptions. To study the impact of such cases, we extracted apps that use barcode libraries such as ZXing [33] or ZBar [34]. We found that in the official market, more than 53% of apps that failed to refer to use of camera had barcode reader libraries in their code. In the third-party market, more than 66% of such apps had barcode libraries. Mobile application distribution platform providers may want to support exposing the use of privacy-sensitive resources by functions that tend to be unmentioned.

(4) **Third-party libraries**

There are some third-party libraries that need to use privacy-sensitive resources. For instance, it is well known that ad libraries make use of resources of location or account information for establishing targeted advertisement [5]. Another example of third-party libraries are log analysis libraries and crash analysis libraries. These libraries make use of `get task` permission and location information. We analyzed apps that have callable location APIs/URIs and text descriptions that do not refer to the location permission. We found that in the official market, more than 62% of such apps use ad libraries. In the third-party market, more than 80% of such apps used ad libraries. Similarly, in the third-party market, more than 20% of apps that failed to refer to location permission used access analysis libraries. Thus, if a developer uses these third-party libraries, it is likely that the description of the app fails to refer to the permission unless the developer explicitly expresses it.

5.5 Threats to Validity

This section discusses several limitations of our analysis and evaluation.

5.5.1 Static code analysis

Although we developed an algorithm to check whether privacy-sensitive APIs/URIs are callable, we are aware of some limitations. First, although the algorithm can detect the callability of APIs/URIs, we cannot precisely ensure that they are actually called. Second, our static code analysis cannot dynamically track assigned program code at run-time, such as reflection. Third, as Poeplau et al. [35] revealed, some malware families have the ability to self-update; i.e., after installation, an app can download the new version of itself and load the new version via `DexClassLoad`. Employing dynamic code analysis could be a promising solution to these problems. However, other challenges may include scalability and the creation of test patterns for UI navigations [36, 37]. As we mentioned earlier, we adopted static analysis because our empirical study required analysis of a huge volume of applications. On the other hand, we note that static code analysis has a chance to extract *hidden* functions that cannot be explored by a dynamic analysis. We leave these challenges for our future work.

5.5.2 Accuracy of the keyword-based approach

As we mentioned earlier, we evaluated the accuracy of the ACODE framework using 5 of 11 permissions we considered. Our assumption is that the ACODE framework establishes good accuracy for the rest of 6 permissions. However, there may be a concern that the keyword-based approach works better for some permissions more than others. We note that some of the results derived in Section 5.4 were based on permissions for which we did not evaluate the accuracy; e.g., `SEND_SMS`, `KILL_BG_PROC`, and `GET_TASKS`. Therefore, the results might have threats to validity. A simple solution to address the concern is to extend the labeled dataset, however, we were not able to perform the additional experiments due to the high cost of labeling descriptions written in two languages. Although not conclusive, we note that we have validated that the descriptions were correctly classified through the manual inspection, using randomly sampled apps; i.e., the obtained results were partially validated.

6. DISCUSSION

In this section, we discuss the feasibility and versatility of the ACODE framework. We also outline several future research directions that are extensions of our work.

6.1 User experience

In this study, we asked participants to read whole sentences carefully, regardless of the size of the text description. In a real-user setting, users might stop reading a text description if it is very long. Studying how the length of text descriptions or the placement of permission-related sentences affect user awareness is a topic for future work. In addition to text descriptions, mobile software distribution platforms provide other information channels, such as meta data or screenshots of an app. As users may also pay attention to these sources of information, studying how these sources provide information about permissions is another research challenge we are planning to address.

6.2 Cost of analysis

Because this work aims to tackle with a huge volume of applications, we adopt light-weight approaches; static code analysis (instead of dynamic code analysis) and keyword-based text analysis (instead of semantic analysis). In the followings, we detail the cost of our approach. The cost of data analysis with the ACODE framework can be divided into two parts: the static code analyzer and the text descriptions analyzer. For the static code analyzer, the most expensive task is the function call analysis because we first need to build function call trees to study whether an API is callable. Our empirical study showed that the task of function call analysis for an application took 6.05 seconds on average. We note that the tasks can be easily parallelized. By parallelizing the tasks with 24 of processes on a commodity PC, we were able to process 200 K apps within a single day. For the text description analyzer, collecting label was the most expensive task. On average, a single participant labeled 1,500 of descriptions within 10 hours. However, once we get the performance evaluation of our approach, we do not need to employ the task again because our work does not need manually-labeled samples. Since we adopt keyword-based approach, analyzing hundred thousands of descriptions was quite fast.

Overall, all the tasks can be completed within a single day, and we can further accelerate the speed if this is desired. As our objective is not to perform the analysis in real-time, we believe that the cost of performing analyses with the ACODE framework is affordable.

6.3 Permissions that should or should not be mentioned.

Android OS manages several permissions with a protection level defined as "dangerous," which means "a higher-risk permission that would give a requesting application access to private user data or control over the device that can negatively impact the user [38]." Ideally, users should be aware of all these dangerous permissions. The dangerous permissions can be broadly classified into two categories: for users and for developers. Permissions for users include read/write contacts, access fine location, read/write calendar, read/write user dictionary, camera, microphone, Bluetooth, and send/read SMS. The three resources analyzed in this paper are the permissions aimed at users. Permissions for developers include set debug app, set process limit, signal persistent processes, reorder tasks, write setting, and persistent activity.

Permissions for users are intuitively understandable. Thus, they should be described in the text descriptions. Permissions for developers are difficult for general users to understand; thus, describing them may be confusing. As describing these permissions could even distract users' attention from the text descriptions, they should *not* be mentioned in the text descriptions. For such dangerous permissions aimed at developers, we need to develop another information channel that lets users know about the potential threats in an intuitive way. We note that the ACODE framework can be used to identify dangerous permissions that are least mentioned. Knowledge of such permissions will be useful to develop a new information channel.

7. RELATED WORK

Researchers have studied mobile apps from various viewpoints, including issues of privacy, permission, and user behavior. In this section, we review the previous studies along four axes: system-level protection schemes, large-scale data analyses, user confidence and user behavior, and text descriptions of mobile apps.

7.1 System-level protection schemes

As a means of protecting users from malicious software, several studies have proposed install-time or runtime protection extensions that aim to achieve access control and application isolation mechanisms such as [39, 40, 41, 42, 43]. Kirin [39] performs lightweight certification of applications to mitigate malware at install-time based on a conservative security policy. With regard to install-time permission policies and runtime inter-application communication policies, SAINT [40] provides operational policies to expose the impact of security policies on application functionality, and to manage dependencies between application interfaces. TaintDroid [41] modifies the operating system and conducts dynamic data tainting at runtime in order to track the flow of sensitive data to detect when this data is exfiltrated. Quire [44] is defense mechanisms against privilege escalation attacks with inter-component communication (ICC). Finally, SEAndroid [43] brings flexible mandatory access control (MAC) to Android by enabling the effective use of Security Enhanced Linux (SELinux).

While the above studies improved the system-level security and privacy of smartphone, this work attempts to address the problem from a different perspective – understanding the effectiveness of text description as a potential source of information channel for improving users' awareness of privacy.

7.2 Large-scale data analyses

Several researchers have conducted measurement studies to understand how many mobile apps access to private resources and how they use permissions to do so [2, 3, 4, 5]. A survey report published by Bit9 [2] included a large-scale analysis of Android apps using more than 410,000 of Android apps collected from the official Google Play marketplace. Through the analysis, they revealed that roughly 26% of apps access personal information such as contacts and e-mail, 42% of apps access GPS, and 31% of apps access phone calls or phone numbers. Book et al. [5] analyzed how the behavior of the Android ad library and permissions have changed over time. Through the analysis of 114,000 apps collected from Google Play, they found that the use of most permissions has increased over time, and concluded that permissions required by ad libraries could expose a significant weakness in user privacy and security. From the perspective of dynamic code loading, Poeplau et al.[35] conducted an analysis of 1,632 popular apps, each with more than 1 million installations, and revealed that 9.25% of them are vulnerable to code injection attacks.

7.3 User confidence and user behavior

Several works on user confidence and user behavior discuss users' installation decisions [9, 45, 46, 47]. Refs. [46, 47] studied user behavior in security warnings, and revealed that most users continue through security warnings. Good et al. [45] conducted an ecological study of computer users installing software, and found that providing vague information in EULAs and providing short notices can create an unwarranted impression of increased security. Chin et al. [9] studied security and privacy implications of smartphone user's behaviors based on a set of installation factors, e.g., price, reviews, developer, and privacy. Their study implicates user agreements and privacy policies as the lowest-ranked factors for the privacy. As these studies on user confidence and behavior suggest, user agreements or privacy policies are not effectively informing consumers about privacy issues with apps. Centralized mobile software distribution platforms should provide mechanisms that improve privacy awareness so users can use apps safely and confidently. We believe that our findings obtained using the ACODE framework can be used to complement these studies.

7.4 Text descriptions

As mentioned in section 1, only a few works have focused on text descriptions of mobile apps [10, 11, 12, 13]. The WHYPER framework [11] is the pioneering work that attempted to bridge the semantic gap between application behaviors and user expectations. They applied modern NLP techniques for semantic analysis of text descriptions, and demonstrated that WHYPER can accurately detect text sentences that refer to a permission. Qu et al. [13] indicated an inherent limitation of the WHYPER framework, i.e., the derived semantic information is limited by the use of a fixed vocabulary derived from Android API documents and synonyms of keywords there. To overcome the issue, they proposed the AutoCog framework based on modern NLP techniques extracting semantics from descriptions without using API documents. The key idea behind their approach is to select noun-phrase based governor-dependent pairs related to each permission. They demonstrated that the AutoCog framework moderately improved performance as compared to the WHYPER framework. Gorla et al. [12] proposed the CHABADA framework, which can identify anomalies automatically by applying an unsupervised clustering algorithm to text descriptions and identifying API usage within each cluster. Like our work, CHABADA uses API functions to identify outliers. On the other hand, the aim of ACODE is *not* to find anomalies, but to quantify the effectiveness of text descriptions as a means of making users aware of privacy threats. To this end, using a simple keyword-based approach, the ACODE framework

Table 10: Comparison between related works.

	ACODE	WHYPER	AutoCog	CHABADA	Lin et al. [10]
objective	Understanding inconsistency between codes and descriptions	Identifying sentences that refer to a permission	Assessing description-to-permission fidelity of applications	Identifying outlier apps	Understanding user expectation on sensitive resources
# of apps	200,000	581	83,656	32,308	134
# of studied permissions	11	3	11	N/A	4
markets	Official, Third-party	Official	Official	Official	Official
languages	English, Chinese	English	English	English	English
code analysis	Function call tree analysis	Permission check	Permission check	API analysis	Permission check
description analysis	Keyword-based	Semantic analysis	Semantic analysis	Topic model	N/A

attempts to assess the reasons why text descriptions do not refer to permissions. As we revealed, the performance of our approach is comparable with that of the WHYPER framework. We also note that the ACODE framework is more fine-grained than CHABADA since ACODE checks whether API functions/URIs found in code are callable by employing function call analysis. Finally, Lin et al. [10] studied users' expectations related to sensitive resources and mobile apps by using crowdsourcing. They asked participants to read the provided screenshots and text description of an app, and asked several questions to investigate users' perceptions of the app as related to privacy-sensitive resources. They concluded that users' expectations and the purpose for using sensitive resources have a major impact on users' subjective feelings and their trust decisions. This observation supports the importance of improving users' privacy awareness on mobile software distribution platforms.

We summarize the differences among the above three studies, and our own in table 10. In addition to the technical differences shown above, our work is distinguishable from other studies in its large-scale empirical analysis, which spans across 11 of distinct permissions, two market places, and 200K of text descriptions written in two different natural languages.

8. CONCLUSION

By applying the ACODE framework to 200,000 apps collected from both official and third-party marketplaces, our analysis across the 11 distinct resources revealed four primary factors that are associated with the inconsistencies between text descriptions and use of privacy-sensitive resources: (1) existence of app building services/frameworks that tend to add API permissions/code unnecessarily, (2) existence of prolific developers who publish many applications that unnecessarily install permissions and code, (3) existence of secondary functions that tend to be unmentioned, and (4) existence of third-party libraries that access to the privacy-sensitive resources.

We believe that our work provides an important first step toward improving users' privacy awareness on mobile software distribution platforms. For instance, developers of app building services/frameworks can use our findings to check the behaviour and deployment of their products. Individual mobile app developers can pay attention to our findings when they write text descriptions or use third-party libraries. And mobile software distribution platform providers can pay attentions to all the potential reasons that lead to the inconsistencies between user expectations and developer intentions. Based on the findings revealed by the ACODE framework, they may be able to come up with new information channels that effectively inform users about the use of privacy-sensitive resources.

Acknowledgments

We are grateful to the authors of WHYPER framework [11] for sharing the invaluable datasets with the research community. We also thank Akira Kanaoka for inspiring us to start this work. Our special thanks are to Lira Park, Gracia Rusli, Ahro Oh, Suthinan Thanintranon, Karyu Chen, Xia Tian, Bo Sun, Jiarong Chen, Xuefeng Zhang, Hao Wang, Dan Li, Chen Wang for their assistance in collecting the labeled text descriptions used in this work. Finally, we thank the anonymous reviewers for their thoughtful suggestions for improving this paper. In particular, we thank our shepherd, William Enck for his valuable feedback.

9. REFERENCES

[1] S. Shen and B. Blau, "Forecast: Mobile App Stores, Worldwide, 2013 Update." `https://www.gartner.com/doc/2584918/forecast-mobile-app-stores-worldwide`.

[2] B. Report, "Pausing Google Play: More Than 100,000 Android Apps May Pose Security Risks." `https://www.bit9.com/research/pausing-google-play/`.

[3] A. P. Felt, E. Chin, S. Hanna, D. Song, and D. Wagner, "Android permissions demystified," in *Proceedings of the 18th ACM Conference on Computer and Communications Security*, CCS '11, pp. 627–638, 2011.

[4] A. P. Felt, E. Ha, S. Egelman, A. Haney, E. Chin, and D. Wagner, "Android permissions: user attention, comprehension, and behavior," in *Symposium on Usable Privacy and Security (SOUPS)*, 2012.

[5] T. Book, A. Pridgen, and D. S. Wallach, "Longitudinal analysis of android ad library permissions," in *IEEE Mobile Security Technologies (MoST)*, 2013.

[6] Y. Zhou and X. Jiang, "Detecting passive content leaks and pollution in android applications," in *20th Annual Network & Distributed System Security Symposium (NDSS)*, Feb. 2013.

[7] J. Kim, Y. Yoon, K. Yi, and J. Shin, "ScanDal: Static analyzer for detecting privacy leaks in android applications," in *MoST 2012: Mobile Security Technologies 2012*, May 2012.

[8] Future of Privacy Forum, "FPF Mobile Apps Study." `http://www.futureofprivacy.org/wp-content/uploads/Mobile-Apps-Study-June-2012.pdf`.

[9] E. Chin, A. P. Felt, V. Sekar, and D. Wagner, "Measuring user confidence in smartphone security and privacy," in *Symposium on Usable Privacy and Security (SOUPS)*, 2012.

[10] J. Lin, S. Amini, J. I. Hong, N. Sadeh, J. Lindqvist, and J. Zhang, "Expectation and purpose: Understanding users' mental models of mobile app privacy through crowdsourcing," in *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*, pp. 501–510, 2012.

[11] R. Pandita, X. Xiao, W. Yang, W. Enck, and T. Xie, "Whyper: Towards automating risk assessment of mobile applications," in *Proceedings of the 22Nd USENIX Conference on Security*, pp. 527–542, Aug 2013.

[12] A. Gorla, I. Tavecchia, F. Gross, and A. Zeller, "Checking app behavior against app descriptions," in *ICSE'14: Proceedings of the 36th International Conference on Software Engineering*, 2014.

[13] V. Rastogi, X. Zhang, Y. Chen, T. Zhu, and Z. Chen, "Autocog: Measuring the description-to-permission fidelity in android applications," in *Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security*, CCS '14, 2014.

[14] M. P. Lewis, ed., *Ethnologue: Languages of the World*. Dallas, TX, USA: SIL International, seventeenth ed., 2013.

[15] Y. Zhou, Z. Wang, W. Zhou, and X. Jiang, "Hey, you, get off of my market: Detecting malicious apps in official and alternative Android markets," in *19th Annual Network & Distributed System Security Symposium (NDSS)*, Feb. 2012.

[16] "Android asset packaging tool." http://www.kandroid.org/guide/developing/tools/aapt.html.

[17] "PScout: Analyzing the Android Permission Specification." `http://pscout.csl.toronto.edu/`.

[18] K. W. Y. Au, Y. F. Zhou, Z. Huang, and D. Lie, "Pscout: Analyzing the android permission specification," in *Proceedings of the 2012 ACM Conference on Computer and Communications Security*, CCS '12, pp. 217–228, 2012.

[19] "android-apktool." `http://code.google.com/p/android-apktool/`.

[20] "smali – An assembler/disassembler for Android's dex format." `https://code.google.com/p/smali/`.

[21] "androguard." `https://code.google.com/p/androguard/`.

[22] "Kyoto Text Analysis Toolkit." `http://www.phontron.com/kytea/`.

[23] "Natural Language Toolkit." `http://www.nltk.org`.

[24] "imdict-chinese-analyzer." `https://code.google.com/p/imdict-chinese-analyzer/`.

[25] M. Makrehchi and M. S. Kamel, "Automatic extraction of domain-specific stopwords from labeled documents," in *Proceedings of the IR Research, 30th European Conference on Advances in Information Retrieval*, ECIR'08, pp. 222–233, 2008.

[26] S. E. Robertson and K. S. Jones, "Simple, Proven Approaches to Text Retrieval," Tech. Rep. 356, University of Cambridge Computer Laboratory, 1997.

[27] "Whyper: Towards automating risk assessment of mobile applications." `https://sites.google.com/site/whypermission/`.

[28] "Google play." `http://play.google.com/`.

[29] "Anzhi.com." http://anzhi.com.

[30] "Nduoa market." `http://www.nduoa.com/`.

[31] J. Oberheide and C. Miller, "Dissecting the android bouncer." SummerCon, Brooklyn, NY., 2012. `http://jon.oberheide.org/files/summercon12-bouncer.pdf`.

[32] X. Jin, X. Hu, K. Ying, W. Du, H. Yin, and G. N. Peri, "Code injection attacks on html5-based mobile apps: Characterization, detection and mitigation," in *Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security*, CCS '14, pp. 66–77, 2014.

[33] "Official ZXing ("Zebra Crossing") project home." `https://github.com/zxing/zxing`.

[34] "ZBar bar code reader." `http://zbar.sourceforge.net/`.

[35] S. Poeplau, Y. Fratantonio, A. Bianchi, C. Kruegel, and G. Vigna, "Execute this! analyzing unsafe and malicious dynamic code loading in android applications," in *Proceedings of the Network and Distributed System Security Symposium (NDSS)*, 2014.

[36] C. Zheng, S. Zhu, S. Dai, G. Gu, X. Gong, X. Han, and W. Zou, "Smartdroid: An automatic system for revealing ui-based trigger conditions in android applications," in *Proceedings of the Second ACM Workshop on Security and Privacy in Smartphones and Mobile Devices*, SPSM '12, pp. 93–104, 2012.

[37] A. Gianazza, F. Maggi, A. Fattori, L. Cavallaro, and S. Zanero, "Puppetdroid: A user-centric ui exerciser for automatic dynamic analysis of similar android applications," *CoRR*, vol. abs/1402.4826, 2014.

[38] "Android developers guide: App manifest – permission." `http://developer.android.com/guide/topics/manifest/permission-element.html`.

[39] W. Enck, M. Ongtang, and P. McDaniel, "On lightweight mobile phone application certification," in *Proceedings of the 16th ACM Conference on Computer and Communications Security*, CCS '09, pp. 235–245, ACM, 2009.

[40] M. Ongtang, S. McLaughlin, W. Enck, and P. McDaniel, "Semantically rich application-centric security in android," in *Proceedings of the 2009 Annual Computer Security Applications Conference*, ACSAC '09, (Washington, DC, USA), pp. 340–349, IEEE Computer Society, 2009.

[41] W. Enck, P. Gilbert, B.-G. Chun, L. P. Cox, J. Jung, P. McDaniel, and A. N. Sheth, "Taintdroid: An information-flow tracking system for realtime privacy monitoring on smartphones," in *Proceedings of the 9th USENIX Conference on Operating Systems Design and Implementation*, OSDI'10, 2010.

[42] S. Bugiel, L. Davi, A. Dmitrienko, T. Fischer, A.-R. Sadeghi, and B. Shastry, "Towards taming privilege-escalation attacks on android," in *19th Annual Network and Distributed System Security Symposium (NDSS)*, 2012.

[43] S. Smalley and R. Craig, "Security Enhanced (SE) Android: Bringing Flexible MAC to Android," in *NDSS*, The Internet Society, 2013.

[44] M. Dietz, S. Shekhar, Y. Pisetsky, A. Shu, and D. S. Wallach, "Quire: Lightweight provenance for smart phone operating systems," in *20th USENIX Security Symposium*, 2011.

[45] N. Good, R. Dhamija, J. Grossklags, D. Thaw, S. Aronowitz, D. Mulligan, and J. Konstan, "Stopping spyware at the gate: A user study of privacy, notice and spyware," in *Symposium on Usable Privacy and Security (SOUPS)*, pp. 43–52, 2005.

[46] C. Bravo-Lillo, L. F. Cranor, J. Downs, and S. Komanduri, "Bridging the gap in computer security warnings: A mental model approach," *IEEE Security & Privacy*, vol. 9, no. 2, pp. 18–26, 2011.

[47] D. Akhawe and A. P. Felt, "Alice in warningland: A large-scale field study of browser security warning effectiveness," in *Proceedings of the 22Nd USENIX Conference on Security*, Security'13, 2013.

APPENDIX

Table 11: Top-100 Domain-specific Stop Words: English (top) and Chinese (bottom). The keywords are sorted from the first (top-left) to the last (bottom-right).

(1) :	(2) !	(3) app	(4))	(5) (
(6) -	(7) free	(8) ’s	(9) get	(10) feature
(11) time	(12) use	(13) one	(14) new	(15) ?
(16) also	(17) application	(18) game	(19) android	(20) phone
(21) like	(22) make	(23) find	(24) help	(25) ha
(26) version	(27) please	(28) “	(29) ”	(30) simple
(31) screen	(32) need	(33) easy	(34) best	(35) play
(36) see	(37) fun	(38) way	(39) want	(40) n’t
(41) device	(42) information	(43) many	(44) friend	(45) using
(46) take	(47) know	(48) download	(49) keep	(50) go
(51) u	(52) work	(53) enjoy	(54) ielts	(55) &
(56) full	(57) different	(58) world	(59) show	(60) support
(61) mobile	(62) let	(63) available	(64) 2	(65) level
(66) share	(67) 3	(68) give	(69) day	(70) score
(71) set	(72) 1	(73) check	(74) list	(75) right
(76) number	(77) email	(78) access	(79) great	(80) learn
(81) save	(82) user	(83) note	(84) every	(85) update
(86) well	(87) even	(88) much	(89) button	(90) view
(91) add	(92) smoked	(93) try	(94) live	(95) first
(96) video	(97) search	(98) home	(99) ad	(100) allows

(1) 款	(2) 中	(3) 游戏	(4) 上	(5) 一个
(6) 不	(7) 更	(8) 最	(9) 还	(10) 简单
(11) 会	(12) 都	(13) 人	(14) 手机	(15) 功能
(16) 好	(17) 需要	(18) 更新	(19) 大	(20) 玩
(21) 快	(22) 提供	(23) 试试	(24) 小游戏	(25) 操作
(26) 小	(27) 信息	(28) 非常	(29) 赶快	(30) 软件
(31) 很	(32) 点击	(33) 生活	(34) 时间	(35) 想
(36) 所有	(37) 挑战	(38) 选择	(39) 体验	(40) 各种
(41) 不同	(42) 玩家	(43) 可爱	(44) 种	(45) 类
(46) 界面	(47) 障碍	(48) 使用	(49) 起来	(50) 看
(51) 支持	(52) 壁纸	(53) 最新	(54) 有趣	(55) 画面
(56) 休闲	(57) 经典	(58) 疯狂	(59) 屏幕	(60) 后
(61) 益智	(62) 完成	(63) 乐趣	(64) 内容	(65) 控制
(66) 帮助	(67) 世界	(68) 应用	(69) 出	(70) 一下
(71) 喜欢	(72) 方式	(73) 下	(74) 收集	(75) 时
(76) 版	(77) 关卡	(78) 进行	(79) 好玩	(80) 美女
(81) 能够	(82) 朋友	(83) nbsp	(84) 用户	(85) v1
(86) 获得	(87) 一定	(88) 看看	(89) 错过	(90) 消除
(91) 图片	(92) 精美	(93) 能力	(94) 越	(95) 键
(96) 必备	(97) 一起	(98) 任务	(99) 平台	(100) 没有

Table 12: Top-10 Extracted keywords for each permission: English (top) and Chinese (bottom). The keywords are sorted from the first (left) to the last (right).

ACCESS_FINE_LOCATION	(1) gps	(2) location	(3) map	(4) restaurant	(5) direction
	(6) city	(7) event	(8) local	(9) service	(10) area
SEND_SMS	(1) sms	(2) message	(3) sent	(4) send	(5) text
	(6) call	(7) contact	(8) gps	(9) notification	(10) automatically
KILL_BACKGROUND_PROCESSES	(1) task	(2) kill	(3) manager	(4) protection	(5) running
	(6) memory	(7) process	(8) usage	(9) battery	(10) clean
READ_CALENDAR	(1) calendar	(2) reminder	(3) meeting	(4) event	(5) schedule
	(6) widget	(7) google	(8) call	(9) notification	(10) data
CAMERA	(1) camera	(2) scan	(3) photo	(4) event	(5) customer
	(6) direction	(7) restaurant	(8) gallery	(9) offer	(10) community
GET_TASKS	(1) lock	(2) security	(3) task	(4) apps	(5) battery
	(6) password	(7) usage	(8) running	(9) thank	(10) devices
WRITE_SETTINGS	(1) alarm	(2) ring	(3) bluetooth	(4) notification	(5) wifi
	(6) volume	(7) sound	(8) switch	(9) battery	(10) song
RECORD_AUDIO	(1) recording	(2) voice	(3) record	(4) audio	(5) wall
	(6) exclusive	(7) located	(8) direction	(9) client	(10) event
READ_CONTACTS	(1) sms	(2) call	(3) contact	(4) message	(5) text
	(6) send	(7) notification	(8) receive	(9) automatically	(10) service
GET_ACCOUNTS	(1) grab	(2) google	(3) youtube	(4) account	(5) calendar
	(6) contact	(7) feed	(8) fan	(9) join	(10) today
READ_SMS	(1) sms	(2) message	(3) incoming	(4) backup	(5) call
	(6) log	(7) contact	(8) text	(9) notification	(10) data

ACCESS_FINE_LOCATION	(1) GPS	(2) 周边	(3) 附近	(4) 定位	(5) 优惠
	(6) 商家	(7) 酒店	(8) 会 员	(9) 导航	(10) 文
SEND_SMS	(1) 重生	(2) 妻	(3) 宠	(4) 嫁	(5) 总裁
	(6) 妃	(7) 男人	(8) 穿越	(9) 竟然	(10) 世
KILL_BACKGROUND_PROCESSES	(1) 狐	(2) 清理	(3) 进程	(4) 垃圾	(5) 省 电
	(6) 骚扰	(7) 管理器	(8) 耗	(9) 卸载	(10) 内存
READ_CALENDAR	(1) 日程	(2) 日历	(3) widget	(4) 谷歌	(5) 部件
	(6) 优先	(7) 日期	(8) Google	(9) 震动	(10) 彩信
CAMERA	(1) 码	(2) 扫描	(3) 相机	(4) 专用	(5) 淘
	(6) 拍照	(7) 拍 摄	(8) 商品	(9) 购物	(10) 店
GET_TASKS	(1) 结局	(2) 文	(3) 爱情	(4) 爱上	(5) 亲
	(6) 离 开	(7) 梦	(8) 幸福	(9) 真的	(10) 终于
WRITE_SETTINGS	(1) 铃声	(2) 总裁	(3) 修	(4) 仙	(5) 夜
	(6) 竟然	(7) 嫁	(8) 女人	(9) 男人	(10) 却
RECORD_AUDIO	(1) 完结	(2) 本文	(3) 结局	(4) 文	(5) 恋
	(6) 语音	(7) 爱情	(8) 爱上	(9) 亲	(10) 离 开
READ_CONTACTS	(1) 通话	(2) 来 电	(3) 拨号	(4) 通讯	(5) 短信
	(6) 号码	(7) 打 电话	(8) 备份	(9) 话费	(10) 拨打
GET_ACCOUNTS	(1) Google	(2) 谷歌	(3) Free	(4) 备份	(5) 日历
	(6) 云端	(7) Facebook	(8) 丢失	(9) 百度	(10) 呼叫

On the Impact of Touch ID on iPhone Passcodes

Ivan Cherapau, Ildar Muslukhov, Nalin Asanka, Konstantin Beznosov
University of British Columbia, Vancouver, Canada
{icherapau,ildarm,nalin,beznosov}@ece.ubc.ca

ABSTRACT

Smartphones today store large amounts of data that can be confidential, private or sensitive. To protect such data, all mobile OSs have a phone lock mechanism, a mechanism that requires user authentication before granting access to applications and data on the phone. iPhone's unlocking secret (a.k.a., *passcode* in Apple's terminology) is also used to derive a key for encrypting data on the device. Recently, Apple has introduced *Touch ID*, that allows a fingerprint-based authentication to be used for unlocking an iPhone. The intuition behind the technology was that its usability would allow users to use stronger passcodes for locking their iOS devices, without substantially sacrificing usability. To this date, it is unclear, however, if users take advantage of Touch ID technology and if they, indeed, employ stronger passcodes. It is the main objective and the contribution of this paper to fill this knowledge gap.

In order to answer this question, we conducted three user studies (a) an in-person survey with 90 participants, (b) interviews with 21 participants, and (c) an online survey with 374 Amazon Mechanical Turks. Overall, we found that users do not take an advantage of Touch ID and use weak unlocking secrets, mainly 4-digit PINs, similarly to those users who do not use Touch ID. To our surprise, we found that more than 30% of the participants in each group did not know that they could use passwords instead of 4-digit PINs. Some other participants indicated that they adopted PINs due to better usability, in comparison to passwords. Most of the participants agreed that Touch ID, indeed, offers usability benefits, such as convenience, speed and ease of use. Finally, we found that there is a disconnect between users' desires for security that their passcodes have to offer and the reality. In particular, only 12% of participants correctly estimated the security their passcodes provide.

1. INTRODUCTION

Smartphones have become our primary devices for accessing data and applications. With more than a billion smartphones sold in 2014 and more than 2 billion active subscribers, global smartphone user base is expected to grow to 5.6 billion by 2019 [15]. Smartphones are already used for online banking, accessing corporate data, operations that used to be only in the domain of desktops and laptops. This results in sensitive and confidential data being stored and accessed on smartphones. High mobility and small size of smartphones alter the common threat model we used for desktop and laptops devices. In particular, it is much easier to steal smartphones due to their size, and then to access data-at-rest [29].

Adopted by all mobile OS developers, the state of the art in protecting data-at-rest is to encrypt it. In order to avoid the problem of storing an encryption key together with the encrypted data, the key encryption key is commonly derived from the secret used for unlocking the device. Unfortunately, users employ weak unlocking secrets (a.k.a., "passcodes" in Apple's terminology), mainly due to usability-related considerations [32]. Being most common unlocking secretes, personal identification numbers (PINs) are not only susceptible to shoulder surfing attacks, but can also be easily brute-forced [34]. At the same time, PINs are considered unusable by more than 20% of smartphone users [32]. In particular, usability issues pushed these users to disable smartphone lock completely, which leaves hundreds of millions of such users unprotected [31].

Several device manufactures, such as Apple and Samsung, have recently introduced biometric authentication for unlocking smartphones. As a case in point, with the release of iPhone 5S in 2013, Apple has introduced a fingerprint sensor integrated into the "home button". Branded as Touch ID, the sensor authenticates a user, once she touches the button. As stated in the iOS security white paper [4], the key advantage of Touch ID is that it *"makes using a longer, more complex password far more practical because users won't have to enter it as frequently"* and *"the stronger the user password is, the stronger the encryption key becomes. Touch ID can be used to enhance this equation by enabling the user to establish a much stronger password than would otherwise be practical."*

These claims appear to be based on the assumption that the usability of a password largely depends on the frequency of its usage and that users will use stronger passwords, as a result of the decrease in usage frequency. Recent research, however, casts doubts on this assumption. In particular, several findings suggest that users tend to create low-entropy passwords, regardless of how frequently they have to input them [8, 18, 35]. Thus, it is unclear if and how Touch ID impacts the choice of users' passcodes. It is the main focus and the contribution of this paper to fill this knowledge gap.

In order to understand the impact of Touch ID sensor on users' passcode selection, we focused on testing our main hypothesis (H_1^{alt}) – *"There is a difference in passcode entropy between those who use Touch ID and those who do not."* For assessing passcode's strength, we used *zero-order entropy*, which estimates the search space of a secret, assuming that each character is chosen randomly and independently. Zero-order entropy served the purpose of comparing the strength of two passcode groups, without having access to actual passcodes. The results of our study revealed that even

Symposium on Usable Privacy and Security (SOUPS) 2015, July 22–24, 2015, Ottawa, Canada.

with zero-order entropy, which overestimated the real complexity of passcodes, the strength of the participants' passcodes was such that made brute-force attacks practical. For brevity, throughout this paper we refer to zero-order entropy as "entropy".

To test H_1^{alt}, we performed three user studies. First, we conducted an in-person survey with 90 iPhone owners in shopping malls and other public places in Vancouver, Canada. We opted for an in-person survey in order to verify accurately the self-reported data, such as the passcode length and the method of the phone unlocking. Results of the survey did not reveal statistically significant difference in the passcode entropies between those who did and who didn't use Touch ID. Furthermore, the 95% confidence interval suggested that if, hypothetically, there were a difference, then its absolute value could not be larger than 3.35 bits.

In order to understand why users are not adopting stronger passwords when Touch ID is available, we followed up with an interview study of 21 participants. Its results led us to identify possible reasons for users to stick with 4-digit PINs. Finally, to corroborate findings of the first two studies, we conducted an online survey with 374 Amazon Mechanical Turks. Overall, we confirmed statistical results of the first study and measured prevalence of reasons for using 4-digit PINs. In particular, more than 30% of the participants were unaware that passwords are available on iPhones, around 35% of the participants preferred PINs, as they are easier to remember, and more than half of the participants used PINs because they are easier to use (e.g., faster to type). In addition, we narrowed down the 95% confidence interval for a theoretical difference in passcode entropies between the two groups down to 1.91 bits.

Overall this paper makes the following contributions:

- We question the validity of the assumption that such phone unlocking methods as Touch ID would nudge users to use higher-entropy passcodes. We did not find any significant difference in passcode strengths between the two groups. Furthermore, the 95% confidence interval for the differences in mean entropy shows that even if there were a statistically significant difference, it would not be greater than 1.91 bits. In the light of observed average entropy (approximately 16 bits), such a difference would result in passcodes of 18 bits of entropy, translating to about 4.5 hours of extra work for an adversary performing an on-device brute-force guessing attack on an iPhone [4].

- We investigate why Touch ID has not resulted in stronger passcodes. In particular, we find that more than 30% of users do not know that they can use passwords, rather than PINs. Others use PINs due to the usability benefits over passwords, e.g., easy to remember or faster to type.

- Finally, we find a significant mismatch between the desires for protection the majority of iPhone owners report and the actual strength of their passcodes. In particular, the preferences of only 12% of participants matched the provided level of protection, while others preferred significantly higher protection. For instance, 48% desired their passcodes to protect the data for more than 40 years, which is far from reality.

The rest of the paper is organized as follows. We first provide background and discuss related work in Sections 2 and 3. Next, we present our research question and our approach at answering it in Section 4. Then we describe our studies: in-person survey in Section 5, interviews in Section 6, and MTurk survey in Section 7. We discuss results in Section 8 and conclude in Section 9.

2. BACKGROUND

We begin this section with a description of a practical brute-force attack on iOS device passcode. Then, we explain how Touch ID works. We conclude by describing zero-order entropy.

2.1 Data Protection and Brute-force Attack

To protect data confidentiality, iOS encrypts each file with a unique *per-file key*. Per-file key is then encrypted with one of four *class keys*. Each of the four class keys is available during various contextual settings, e.g., on the first unlock after booting. These class keys are protected with a combination of the user's passcode and the *device key*, a unique per-device key embedded in the crypto-chip. In order to extract this device key, an adversary can attempt to reverse engineer the crypto chip, which is an expensive task in terms of time and resources required. An alternative option for an adversary would be to mount an *on-device* guessing attack on the passcode. An adversary uses the crypto-chip directly in an *on-device* attack, in order to try passcode candidates and eventually to decrypt class keys. To decrease the effectiveness of such attacks, the crypto-chip in iPhones and iPads is calibrated to take at least 80 ms for each passcode attempt.

In order to mount an *on-device* attack, an adversary needs to run arbitrary code on the target device. This can be achieved by compromising the *boot-chain* [1], which would allow bypassing iOS kernel's limitation on the number of available passcode guessing attempts [42]. For example, the current version of iOS (8.3), if configured so, would limit the number of guesses to 10, and wipeout the device afterwords. It takes some time, effort, and luck to find an exploitable bug in the boot-chain. While no flaws are known in the current iOS, such flaws have been found in earlier versions.

To summarize, due to the feasibility of on-device unlimited guessing attacks, the protection of the data-at-rest on iOS devices could any day end up hinging on the security of their passcodes.

2.2 Touch ID

Touch ID is a biometric authentication sensor based on a high definition fingerprint scanner embedded into *"home button"* on iPhones and iPads. This sensor allows users to unlock their devices by simply touching the home button. Although Touch ID allows to unlock a device without typing in a passcode, users are still required to set passcodes on their devices, before being able to use Touch ID. The main reason for such a strict requirement lays in data-at-rest encryption, which needs a source of entropy that is not stored on the device itself. User's device unlocking secret serves this purpose.

A passcode can be either (1) a simple 4-digit PIN[1] or (2) a longer one, with up to 37 characters selected from the alphabet of 77 symbols, to which we refer in this paper as "password". The user can chose to set up either a PIN or a password as her unlocking secret. We use term "passcode" as a general reference for an unlocking secret, unless we want to distinguish between PINs and passwords.

When a device with Touch ID enabled boots, it prompts the user to provide the correct passcode. At this stage, the internal memory of Touch ID is clear, i.e., immediately after reboot users are not able to use Touch ID sensor. Once the user provides the correct passcode, the iOS is able to recover actual data encryption keys and uses them to decrypt and encrypt data. If the device is locked, OS erases certain types of keys from RAM, which will require either the correct passcode or successful unlocking with Touch ID, in order to recover these keys on unlock. The unlocking flow with Touch ID enabled is shown in Figure 1.

When a user locks the device that has Touch ID enabled, iPhone's

[1] Apple security white paper defines it as a "simple passcode".

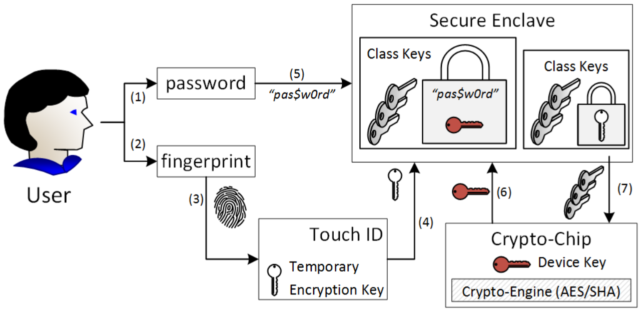


Figure 1: Unlocking flow with Touch ID enabled. When the user locks the device, the class encryption keys are wrapped by a random temporary encryption key (TEK). To unlock the device, the user has two options, she can either (1) type in her passcode, or (2) use Touch ID. When the user uses Touch ID, it authenticates the user by matching her fingerprint with saved fingerprints (3). If the authentication is successful, the sensor releases the TEK to the Secure Enclave (4), which allows decrypting class keys and sending them to the crypto-chip (7). If, the user fails to authenticate for five times with Touch ID, or does not unlock device for 48 hours, the Touch ID sensor flushes the TEK, which leaves typing in the passcode as the only option for unlocking the device. Without Touch ID, the user types her passcode (1), which is sent to the Secure Enclave (5). The combination of the device key (6) and password (5) are used to decrypt class keys and send them to the crypto-chip (7).

CPU generates a random *temporary encryption key* (TEK), which protects certain class keys by "wrapping" them (a cryptographic operation somewhat similar to encryption). It then sends the TEK to Touch ID and deletes class keys from RAM. After that, there are two options for the iOS to recover the wrapped class keys (1) receive the TEK from Touch ID once the user successfully authenticates to the sensor, or (2) receive the correct passcode from the user, then derive the correct encryption key from a combination of the passcode and the device key, and then "unwrap" class keys. When the user touches the Touch ID sensor, the sensor tries to authenticate the user based on the fingerprint. If the authentication attempt is successful, the sensor releases the TEK to the Secure Enclave, which is located in the CPU. If, however, the user fails to authenticate with the fingerprint for five times, or has not unlocked the device for 48 hours, the Touch ID sensor flushes the TEK, which leaves passcode as the only option for unlocking an iPhone.

We decided to focus on Touch ID, because it is deployed on an existing and popular mobile platform, adopted by millions of users worldwide. We did not study Android fingerprint and face recognition because the former is a new technology that first appeared in April 2014 [20] and the latter has not become widely adopted by the users, probably due to usability [7] and security issues [16].

2.3 Zero-order Entropy

The strength of an authentication secret is defined by the effort an attacker needs to spend on guessing it. In simple terms, this effort is assumed to be proportional to the size of the search space the attacker needs to check in order to find the secret. One such metric is *zero-order entropy*, measured in bits and calculated as

$$L * log_2 N$$

where L is the length of the password and N is the character set size. For example, the length of iPhone's PIN in iOS 8.3 is four and the character set size is 10, hence, its zero-order entropy is 13.28 bits. That is, zero-order entropy measures the size of the whole search space of all possible secrets of a given length and the size of a given alphabet, with the assumption that each character is selected randomly and independently from all other characters.

Of course, zero-order entropy, as a metric, suffers from several limitations, when it's applied to human-chosen secrets, like passwords and PINs. The most important one is that it does not measure the secret strength accurately. Recent research has shown that users tend to select highly predictable passwords and often use dictionary words as ones [9, 17]. Such predictability makes the search space smaller, i.e., the work of an attacker easier. This implies that the zero-order entropy measures the upper bound of the attacker's work. In other words, it overestimates the actual work.

3. RELATED WORK

Authentication mechanisms have been studied extensively for many years [8, 26], however, text-based passwords remain the most commonly used authentication mechanism and the security's weakest link [9, 22, 27]. Florencio and Herley [17] conducted a study on web password use and reuse with half a million users over a three months period. Their results suggest that web users employ and re-use low-entropy passwords on websites. Weir *et al.* [40] analyzed a set of leaked passwords. The authors showed that popular passwords were also weak and "123456" was very common among users. To prevent users from choosing passwords that are too easy for an attacker to guess, system administrators often enforce password composition policies [27]. Such a policy might require users to use a password that contains non-alphanumeric symbols, lower and upper case letters, and numbers. Using a password policy that is too strict, however, might backfire and push users to write down passwords or store them on some other devices [27].

Two recent studies examined smartphone locking behaviours using conventional authentication mechanisms. Harbach et al. found that users activate their phones 85 times and unlock their phones 50 times per day on average and that most of users did not see any threat to the data on their phones [21]. Egelman et al. also found a strong correlation between locking behaviours and risk perceptions, but the authors believe that users underestimate actual the risks [14]. In contrast, we focused on studying the effect that Touch ID makes on users unlocking password selection and the reasons for such an effect.

Biometrics-based authentication modality has also received considerable attention from the research community in recent years [2, 30, 38]. Although usability of a biometric system is still an important factor in adoption [33, 37], such authentication methods could potentially remedy common drawbacks of text-based passwords. For example, users do not need to remember anything [7]. Indeed, recent studies showed that the usability of biometric-based phone unlocking is important for users [13]. Crawford and Renaud [12], however, have showed that users are willing to try biometric authentication mainly for its usability benefits. In addition, Breitinger *et al.* [10] suggest that 87% of users are in favour of fingerprint authentication. Others have found that the presence of a biometric factor in a two-factor authentication system can lead users to picking weaker credentials, in comparison with a password-only authentication system [41]. In contrast, we focus on how Touch ID impacts users' choice of iPhone passcodes in a single-factor authentication system.

Indeed, there are many reasons to use fingerprint for authentication. To start with, it is unique to each individual, and it is almost impossible to find two people with an identical fingerprint pattern [4]. Individuals' fingerprint patterns never change during their life span [39]. Fingerprint sensor can improve the security and the convenience for users, if used in smartphones [19], because there are many limitations of smartphones' screens and keyboards [19, 25] that make password-based authentication/unlocking undesirable. For instance, text entry on constrained keyboards is prone to errors, time-consuming and frustrating. In particular, Lee and Zhai showed that error rate for typing on virtual keyboards, i.e., keyboards drawn on a screen, is 8% higher than on hardware keyboards for desktops [28]. In addition, Bao *et al.* [6] found that the average typing speed for an 8-character alphanumeric password on mobile devices is three times slower than on desktop computers.

Finally, recent research suggests that users tend to use weak 4-digit PINs over alphanumeric passwords in smartphones [24, 32]. Users justify such choice by ease of use of PINs, in comparison to passwords, especially in cases when one has to unlock their device with high frequency for day-to-day activities [31]. Unfortunately, it is clear today that a 4-digit PIN provides virtually no security for data-at-rest [4, 36]. To make the matter worse, even within the search space of 4-digit PINs, users make highly predictable choices. For example, Amitay [3] analyzed over 200,000 iPhone PINs and discovered that "1234" is the most common PIN, followed by "0000" and "2580". Considering the software limitation on the number of allowed unlocking attempts (i.e., 10 attempts in iOS) through the user interface, one can try the top 10 PINs and still achieve 15% success rate without the need to go for an *on-device* brute-force attack.[2] That is, one in seven iPhones can be unlocked by just trying the top 10 PINs. It seems that the main intuition behind the design of Touch ID was to reduce the number of times the user must type her authentication secret to unlock the device [4]. Bhagavatula et al. found that most Touch ID users perceive it as more usable and secure than a PIN [7]. To the best of our knowledge, we are the first to assess whether users take an advantage of Touch ID by using stronger passcodes.

4. METHODOLOGY OVERVIEW

The main research question (RQ_M) of our study was "How availability of Touch ID sensor impacts users' selection of unlocking authentication secrets". To answer this research question, we have formulated the following hypotheses to be tested:

- H_1^{null} – *Use of Touch ID has no effect on the entropy of passcodes used for iPhone locking.*
- H_1^{alt} – *Use of Touch ID affects the entropy of passcodes used for iPhone locking.*
- H_2^{null} – *Availability of Touch ID has no effect on ratio of users who lock their iPhones.*
- H_2^{alt} – *Availability of Touch ID increases the ratio of users who lock their iPhones.*

We conducted three user studies, starting with a study based on in-person surveys. This study allowed us to test our hypotheses. In addition, it allowed us to clarify areas with the lack of understanding and refine our follow-up studies. We followed the first study with an interviews, in order to gain deeper insights into passcode selection by users. In particular, we focused on understanding why users do not take advantage of Touch ID, i.e., understanding users' reasoning for not adopting stronger passcodes when Touch ID is available. Finally, to corroborate our data from the first study and to measure the prevalence of the reasons for using weak passcodes, we conducted the third study in a form of an online survey. This study gave us a larger and diverse subject pool for testing our set of hypotheses and provided descriptive statistics on reasons for using weak passcodes.

In the first and third studies, we chose zero-order entropy for estimating the strength of participants' passcodes, even though it has limitations, as discussed in Section 2.3. There were several reasons for this choice. First, evaluation of the passcode's guessability would require access to plaintext passcodes, which we chose not to obtain for ethical considerations. Second, zero-order entropy served well the purpose of our study in comparison of two groups, i.e., with and without Touch ID, in terms of work the attacker needs to do. Finally, the results of our study showed that even if we overestimated the passcodes strength, the actual workload for a brute-forcing attacker is still practical.

We obtained ethics approval from our university's behavioural research ethics board for all three studies.

5. STUDY I: IN-PERSON SURVEY

5.1 Methodology

In our first study, we chose to use an in-person survey of iPhone users for several reasons. First and foremost, this choice allowed us to verify answers related to participants' unlocking behaviour and the authentication secret being used. In addition, an in-person nature of the study allowed us to follow-up unforeseen answers with additional questions. We strived to recruit a pool of diverse participants, hence we approached people in public locations, such as shopping malls and coffee shops. Each participant signed a consent form and received $10 as a compensation for participation.

5.1.1 Study Design

To facilitate faster data collection in public locations with limited and unreliable access to the Internet, we used an iPad with our own survey app. All answers were stored locally on the iPad, and for some of the questions we also validated participants' answers by asking participants to show us some elements of their unlocking process and other relevant data. In particular, we validated the type of the unlocking method used, by asking them to show the locked screen. We also validated the length of the password (for those who used it) by asking participants to show us the unlocking screen after the password has been typed but before they clicked on the *enter* button. This allowed us to validate their answer about the password length by our researcher counting the number of stars in the password field. In addition, participants were asked to navigate to the settings of the auto-lock screen on their iPhones and show us the value of the auto-lock timeout. Finally, by asking each participant who claimed to use Touch ID to unlock their device with a fingerprint, we were able to confirm that they, indeed, used it.

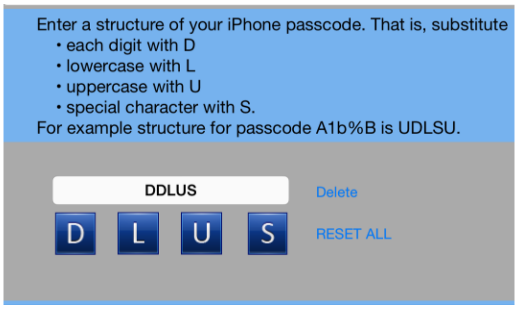
Most of the survey questions were either open-ended or contained option "other", which allowed participants to provide their own answer if needed. The questionnaire guide is provided in Appendix A.1 and consists of the following parts:

Part 1 Demographic questions (e.g., age, gender, education, income, occupation).

Part 2 Security and privacy concerns related questions, e.g., we asked participants if they had any sensitive, private or valuable information on their iPhones.

Part 3 Questions on the experience participants had so far with their smartphones, including if they locked their previous smartphones.

[2] This is a simpler approach that does not require execution of arbitrary code on the device.



Enter a structure of your iPhone passcode. That is, substitute

- each digit with D
- lowercase with L
- uppercase with U
- special character with S.

For example structure for passcode A1b%B is UDLSU.

DDLUS

Figure 2: Passcode structure question in Study I.

Part 4 Passcode metrics questions. In this part, we asked participants to provide us a structure of their unlocking passcodes. In order to preserve confidentiality of their passcodes, we asked participants to substitute each character in their passcodes with the mnemonic of the character type: **D** - digits, **L** - lower-case letters, **U** - upper-case letters, **S** - special characters. We refer to such encodings of passcodes as "masks". The screenshot of this question is shown in Figure 2. We chose this approach for two reasons. First, it allowed us to assess entropy. Second, this approach did not require participants to reveal their passcodes to us.

Part 5a This section was only relevant to the owners of iPhone 5s, 6, and 6 Plus. Here, we asked questions related to Touch ID's usability and reasons for its adoption.

Part 5b This section was only relevant to the owners of iPhone 5 and older models. Here, we asked about their perception of biometric authentication methods such as Touch ID.

In order to test our questionnaire, we conducted a pilot study with 12 participants. Based on the results of the pilot study, we revised several questions in the questionnaire and added an attention check question (#28 in Appendix A.1). Most of the changes we made were aimed at improving questions' clarity and readability.

5.1.2 *Participant Recruitment*

We recruited participants in public places such as shopping malls, libraries and coffee shops in the downtown area of Vancouver. We approached prospective participants who had iPhones with them and invited them to participate in our study. We chose this recruitment method mainly because we were interested in the general population of iPhone users. We recruited participants who were iPhone users and 19 years old or older. Although the main focus of our study were owners of Touch ID (iPhone models 5S, 6, and 6 Plus), we also recruited owners of older models. Participants that used Touch ID were assigned to Touch ID group, while the rest to non-Touch ID group. Note, that those iPhone 5S, 6 and, 6 Plus owners who did not use Touch ID, were assigned to the non-Touch ID group.

5.2 Results

In this section, we report the results of our in-person survey. We first report participants' demographics, then provide findings for all participants and for each group separately. Finally, we report the results of statistical tests for H_1 and H_2.

Participant Demographics. Overall, we recruited 93 participants. We, however, had to exclude responses from 3 participants who failed password length verification. Thus, the results presented in this section are based on **90** participants.

Out of 90 participants, 30 were female. The minimum and maximum age was 19 and 71 years, and the average age was M = 29 (SD = 12). Among all participants, 41 used Touch ID sensor and 49 did not. The majority of our participants was experienced iPhone users, i.e., they owned an iPhone for more than two years. Only 12 participants owned iPhones for less than a year. Almost all of the participants (81) had owned another smartphone before the current one. Most of our participants (69) stated that they unlock their iPhones at least once per hour. In addition, we found that 32 participants had lost their smartphones before, and 15 participants were victims of smartphone theft. On average, participants completed survey in around 5.5 minutes (SD = 2 minutes) in non-Touch ID group, and in around 7 minutes (SD = 3 minutes) in Touch ID group. Demographics summary is provided in Table 2 (column "Study I").

Reasons To Lock Or Not To. Overall the participants use various reasons for locking or not locking their iPhones. Some of the reasons were driven by *a possible attacker*, e.g., 58 participants locked their devices to prevent strangers from access, and four participants locked their devices to protect data if they get mugged, 23 participants used locked their iPhones to control access by their family and/or friends. In addition, we found that some participants used social norms to rationalize locking, e.g., 12 participants locked their devices because their friends did the same.

Other reasons could be attributed to either (1) *usability problems* of device locking, voiced mainly by those who did not lock their device, or (2) the *necessity to have certain features* that were either enabled or prevented by device locking. The four participants who did not lock their device stated the following reasons: (a) locking a phone makes it impossible to use it in emergency cases, (b) locking iPhone makes it impossible to contact the owner in case the device is lost, and (c) unlocking process takes too much time. Only two participants, out of the four who did not lock their iPhones, stated that they did not care about security of their data.

Use of PINs and Passwords. Out of the 90 participants, 86 locked their phones, with 66 employing 4-digit PINs, and 20 using passwords. Third of the participants (36) used the same passcode for their iPhones as in their previous smartphones. In addition, 52 participants stated that they shared their passcode with someone else, and 53 stated that they knew passcodes for smartphones owned by others.

Touch ID Group. The Touch ID group included 41 participants, with 29 of them using 4-digit PINs. The majority of them agreed that they liked using Touch ID. In particular, 26 participants found that setting up Touch ID was easy or very easy, and 29 participants stated that the use of Touch ID was easy or very easy (see Appendix A.2 for more details). The majority of the participants (30) had never had any issues with Touch ID, and, overall, Touch ID participants considered Touch ID as a convenient, secure, quick, and easy to use unlocking mechanism.

Touch ID participants also voiced their concerns with fingerprint scanning sensor. In particular, three participants had problems with sharing their iPhones. Others saw Touch ID sensor as a threat due to the ability of an attacker to unlock the device, while the owner is sleeping (e.g., P9 "... [I] might be sleeping and someone might use my finger to unlock [my iPhone] ...").[3] Some participants were even afraid that an attacker might fake their fingerprints, in order to access the device later. Seven participants worried about privacy of their fingerprints, due to the lack of clarity on whether Apple stores their fingerprints somewhere else. For example, one of the

[3] Exactly the same story has happened in December 2014, when a boy unlocked the iPhone of his sleeping father with his father's thumb [11].

Table 1: Passcode average entropies for Touch ID and non-Touch ID groups in Study I. While non-Touch ID group had 49 participants, four of them did not use any passcode to lock their phones and were excluded in computing entropies.

	Touch ID	Non-Touch ID
Mean	15.88 bits	15.61 bits
SD	6.93 bits	7.45 bits
N	41	45

participants (P11) stated that she was afraid about "Apple leaking my fingerprint and someone can impersonate me" and "fingerprint being used for purposes other than to just unlock my phone."

Non-Touch ID Group. The non-Touch ID group included 49 participants, where 37 participants used PINs and eight used passwords to unlock their iPhones. Four participants did not lock their phones and were excluded from computing average entropy of passcodes in this group. While 13 in this group had Touch ID available, they did not use it.

We observed that participants perceived fingerprint authentication as a security improvement. For example, "anyone can figure out a password but people can't copy your fingerprint" (P69), "for those with sensitive info on phones more security is desirable" (P78), "it is easy, accurate and secure" (P5), "it's safer" (P19), "more secure than 4 digit password" (P33), "no one can fake my fingers" (P89), "I will use Touch ID so my friends don't get in my phone" (P45). Although their iPhones did not have fingerprint scanners, more than one-third of participants believe that Touch ID is the most secure unlocking method. Surprisingly, only three participants from non-Touch Group were willing to use a longer alphanumeric password alongside with the Touch ID.

5.2.1 *Hypothesis Testing*

To test H_1, we first compared proportions of participants that used PINs and passwords in both groups. Then we compared mean values of entropies in both groups. Analysis of proportions did not reveal any statistically significant difference (χ-squared = 1.01, p = 0.32). For computing entropy of participants' passcodes, we obtained the length of the passcodes and the alphabet size from the masks our participants provided (Figure 2). The results of Mann-Whitney U test did not reveal any statistically significant difference between mean values of entropies in Touch-ID and Non-Touch ID groups (W = 15708, p = 0.70), see Table 1. Thus, we were unable to reject H_1^{null}.

In addition, statistical analysis of the mean values of entropies gave us a confidence interval, i.e., the possible interval of the difference. This allowed us to assess the biggest possible difference in entropies in case a statistically significant difference is found, by recruiting larger participant pool. In this case the 95% confidence interval for the difference between the means was from -3.35 up to 2.81, or 3.35 bits at most.

If we consider a hypothetical scenario in which the Touch ID group has a higher entropy, and we simply failed to find it due to small size of the participant pool, and considering the observed mean entropy value of 15.88 bits, we can assess that the possible maximum entropy with 95% confidence is 19.23 bits. Taking into account the design of the data encryption in iPhones, i.e., that each passcode guessing attempt takes at least 80ms, we can show that 19.23 bits of entropy corresponds to roughly 14 hours. In comparison, it would take only 1.1 hour to brute-force passcodes in non-Touch ID group with average entropy of 15.61 bits.

We tested H_2 hypothesis with Chi-squared test (χ-squared = 0, p = 1.0). We were unable to reject H_2^{null}, and hence we conclude that Study 1 failed to show an effect of Touch ID on users' preference to lock their iPhone.

5.3 Limitations

There were several limitations that might have negatively impacted our ability to find a statistically significant difference between passcodes of Touch ID and non-Touch ID groups. First, we might not have obtained large enough sample size. Second, our participant pool had a fairly large bias towards the 19-34 age group. Third, since we obtained only passcode exact length and the types of the characters in each position, but not the characters themselves, this coarse granularity of the data did not allow us to observe the difference. Fourth, as we did not control for or collect data on how technically and security savvy our participants were, we might have had one of the two groups with participants heavily skewed on these traits. In order to address these limitations and gain a deeper insight into why users are sticking with 4-digit PINs we decided to proceed with an interview-based study.

While we included an attention check question (see question 28 in Appendix A.1), we realized after running the survey that the question was poorly worded. So, we have decided not to exclude participants based on their response to this question, because most of those who failed the question likely did not understand it. We paraphrased the question and used it in Study III (see question 36 in Appendix C.1).

6. STUDY II: INTERVIEWS

We followed the in-person survey with an interview study in order to gain a better understanding of users' reasoning to stick with weak passcodes. Our main objective was to answer research question (RQ_1) "Why Touch ID users do not employ stronger passcodes for smartphone locking?"

6.1 Methodology

We designed our study with the focus on qualitative data collection. We used semi-structured interviews since they gave us the freedom to explore new topics, as they emerged. We used theoretical sampling, rather than random sampling, because (as common with explorative enquiries) we were interested in the diversity and richness of the participants' answers, rather than in the generalizability of the findings. A pilot study with eight participants revealed the necessity for real life scenarios in several questions, and we revised the interview guide accordingly. We randomized the order of interview questions, in order to reduce bias due to the order of the questions. Two first interviews were conducted by two researchers together in order to ensure that all important questions were asked and well understood by the participants. Each participant was compensated $10 for a 20-minute interview. We audio recorded all interviews and two researchers coded each interview independently. After each coding session, the coders discussed any disagreements until they reached consensus. Overall, we coded 211 responses into 55 unique codes. Researchers disagreed on the coding of 5 responses, achieving inter-rate agreement of 91%.

6.1.1 *Participant Recruitment*

We recruited participants by directly approaching them in public places such as shopping malls, libraries, and coffee shops in Vancouver. Our inclusion criteria were participants of age 19 years and older who used Touch ID on their iPhones. After the 17th interview, we did not observe any new codes and decided not to schedule new participants, hence we stopped interviewing after 21 participants. Saturation analysis of new concepts with each additional interview is shown in Figure 3.

Table 2: Participants' demographics for the three studies.

		Study I		Study II	Study III	
Parameter	**Value**	#	%	#	#	%
Gender	Female	30	34	10	220	59
	Male	60	66	11	154	41
Age	19 to 24	43	48	7	110	29
	25 to 34	29	32	4	195	52
	35 to 44	8	9	2	49	13
	45 to 54	2	2	2	17	5
	55 to 64	6	7	3	2	1
	65 or older	2	2	3	1	0
	Mean	29		30	N/A	
	Median	30		27	N/A	
Education	High school	30	34	5	19	5
	College degree	22	24	5	129	35
	Bachelor	28	31	8	151	40
	Master or PhD	7	8	3	75	20
	Other	3	3	0	0	0
Income	Less than 20K	25	28	2	67	18
	20K-50K	29	32	3	97	26
	50K-80K	16	18	7	70	19
	80K-120K	8	9	6	99	26
	Above 120K	5	6	0	41	12
	Prefer not to answer	7	8	3	0	0
Industry	Construction	2	2	2	1	0
	Trade	2	2	3	8	2
	Transportation	3	3	1	6	2
	Finance and real estate	7	8	3	23	6
	Professional services	5	6	6	67	17
	Business and building	11	12	0	18	5
	Educational services	4	4	2	51	13
	Health care and social	5	6	2	52	13
	Inform./culture/recreation	3	3	0	16	4
	Accommodation and food services	6	7	3	19	5
	Public administration	1	1	0	9	2
	Other	45	41	3	104	27
Role	Individual Contributor				122	33
	Team Lead				35	9
	Manager				46	12
	Senior Manager				7	2
	Management / C-Level				9	2
	Partner				5	1
	Owner				18	5
	Volunteer				4	1
	Intern				12	3
	Student				57	15
	Other				59	16
Locking method	PIN	66	73	19		
	Password	20	22	2		
	None	4	5	0		
Locked with PIN/Password/None	non-Touch ID				177/6/18	
	Touch ID				166/7/0	

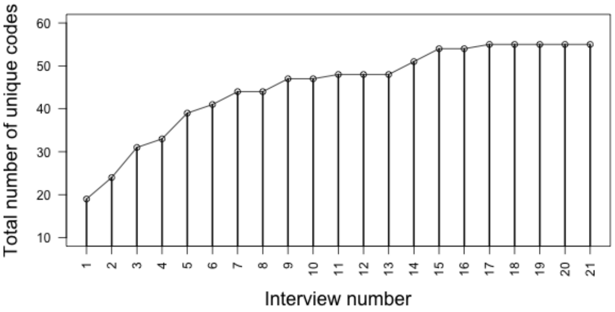


Figure 3: The total number of unique codes for each additional interview in Study II. We reached saturation around 17th interview.

6.1.2 Procedure

After agreeing to be interviewed and showing us their iPhone 5s, 6, or 6 Plus, each participant was asked to read and sign a consent form. The interviewer explained that the purpose of the interview was to investigate how users interact with their iPhones. Interviews followed the interview guide reproduced in Appendix B and consisted of the following parts:

Using Touch ID: In the first part of the interviews, we asked participants to describe why they use Touch ID, how they thought Touch ID works, whether it's possible to use Touch ID without setting up PIN or password, and why and how Touch ID impacts the iPhone security, in case the phone gets stolen.

Locking Behaviour: We asked participants whether they locked their iPhones or not and also, what method they used (PIN or password). We verified their answers by asking them to unlock their iPhones. We asked why they chose to use PIN or password. We also asked participants about their passcode sharing behaviour.

iPhone Data: Then we asked participants about the most valuable data in their iPhones, what data they considered to be confidential or sensitive, and who they cared protecting it against.

Data Protection: We asked participants for how long they wanted their data to be protected, in case their iPhones get stolen.

6.2 Results

6.2.1 Participant Demographics

Overall, we recruited **21** participants, out of which 10 were females, and the average age was 29 (SD = 12.4). Only one participant used a password, while all others used a PIN. All participants had owned an iPhone for over a year. Almost all participants had owned another smartphone before the current one. In addition, 16 participants lost their smartphones before, including the six participants who also were victims of smartphone theft. Participants' demographics are summarized in column "Study II" of Table 2.

6.2.2 Reasons for using PINs

The most common reason for using 4-digit PINs was a wrong perception of Touch ID impact on data security when a device is lost or stolen. In particular, nine participants did not understand how Touch ID works, which led to confusion about the relationship between passcode and Touch ID. They believed that Touch ID "somehow" protects data-at-rest when a device is stolen, i.e., would not allow to decrypt data without a correct fingerprint.

> *"I guess Touch ID will protect my phone. They cannot open my phone without my finger. So it [Touch ID] will definitely help." [P1]*

Another evidence of participants' confusion was that they incorrectly understood how Touch ID and passcode work together. That is, they assumed that using Touch ID, in addition to having a passcode, increases security of data-at-rest, while in reality it does not. In addition, some participants thought that Touch ID provides higher security, compared to passcode. They justified such an answer by stating that users tend to use dictionary words as passwords, while random digits are usually used for PINs. For instance:

> *"Touch ID is more secure than PIN or password because it's unique for the owner" [P3]*

> *"people often choose their dogs' names or middle names or something similar as their passwords" [P11]*

The second most common factor for using a PIN was the lack of knowledge about the ability to use passwords on iPhones. Six participants were not aware that they could use a password for unlocking their iPhones. For instance:

> [After the participant was explained what a passcode is and how to use it.] *"Really? I even did not know that you could do this [use a password]. That is good to know. I will look at it today" [P4]*,

Two participants stated that they used PINs because the sales staff who helped with setting up their iPhones in Apple Stores, showed only the PIN option to the participants. As a result, they believed that this was the only option available:

> *"When I bought my iPhone, they asked me to set up a PIN. That is why I am using PIN" [P5]*

> *"They [Apple store customer service employee] only gave me a PIN code option..." [P14]*

Also, five participants admitted that they got habituated to use PINs from their previous devices, and continued to use PINs on the new iPhones. In addition, participants stated that they did not want to remember a new password, so they just decided to use the old PIN on the new device:

> *"Because on my old phone I was lazy to think about password back then, so now I just stuck with PIN. There is really no major reason; it is just the way it is. I am just too used to this number and I am just too lazy to memorize a new set of numbers." [P1]*

Unsurprisingly some participants stated that they decided to use PINs because it is easier to use, faster to type and easier to remember in comparison to passwords. Indeed, similar results have been shown in previous research, e.g. [31]. In addition, five participants stated that they did not store any sensitive information on their iPhones, thus, they did not care about the extra level of security a password can provide. They believed that PINs are good enough to protect their phones and did not see a good reason to switch to passwords:

> *"PIN is easier. I do not want to type the whole password in. If I lose my phone, it is not a big deal for me. There is nothing important on it" [P15]*

Finally, seven participants reused their PINs across multiple devices or accounts in order to reduce the amount of information they needed to remember. Several participants stated that, because they shared their iPhones with others, PINs were easier to share for them than passwords, for instance:

> *"Simplicity I guess. As I said before, I am not the only person who uses my iPhone. So PIN is easy of access for other users. It is easier to give someone 1234 PIN than 'Charlie-unicorn' is weird, capitals, asterisks, etcetera" [P8]*

In summary, participants provided various reasons for sticking with 4-digits PINs. In particular, some participants did not know that they can use alphanumeric passwords, others were only shown how to setup and use PINs, when they were assisted by the salespeople when purchasing their iPhones. Other participants justified the use of PINs by the fact that they had low requirements for the security of data-at-rest on their iPhones. Some participants were habituated to use PINs from previous devices or wanted to reuse PINs across various devices and accounts. Understandably, participants stressed the usability benefits of PINs over passwords, as one of the reasons to use the former. In particular, they stated that PINs are faster, easier to use and memorize. More critically, our participants misunderstood how Touch ID works and how it impacts the security of data-at-rest, in cases when an iPhone is lost or stolen. Finally, PINs were more convenient than passwords for sharing iPhones with others.

6.2.3 Passcode Sharing Behaviour

Eight participants shared their passcodes with others for several reasons. First, some participants were pressed to share:

> *"I share [PIN] with my girlfriend because she forced me to!" [P2]*

Second, participants trusted others with their data, and, thus shared their passcode:

> *"I share with my boyfriend because I trust him and sometimes he uses my phone, too" [P19]*

> *"I share it with my best friend because I trust her and if she has my phone and needs to look at it, she can do that" [P10]*

Finally, participants shared their passcodes with others because of concerns with emergency situations, when someone close needs access to the phone or its data. For instance:

> *"I share with my girlfriend because if something happens with me, at least she knows the code and can unlock the device" [P9]*

To summarize, the participants shared their passcodes to enable emergency access to their phones, or because they trusted others with the data on their phones, or because they were pressed to share their phones.

6.3 Limitations

Our interview study has several limitations. As with most qualitative enquiries, the results of the interviews are not generalizable. The results of the analysis might have been impacted by our biases. We strived to minimize this bias by using separate coders and discussing the disagreements. Finally, the participants might have misunderstood some questions. To reduce chances of such misunderstanding, we conducted a pilot study with eight participants, with the main purpose of testing the interview questions. We alleviated some of these limitations by conducting our third study.

7. STUDY III: ONLINE SURVEY

The results of the first study suggest the lack of any practically significant impact of Touch ID on passcode selection, prompting us to investigate why users don't choose stronger passcodes, provided that they need to type them rarely if they use Touch ID. While the findings from the second study offered possible reasons for sticking with 4-digits PINs, the study did not allow us to assess the prevalence of these reasons in a representative sample of the iPhone users. Our third study aimed at addressing exactly this limitation. We designed it in a form of an online survey, so that we could recruit a larger and more representative sample in order (a) to corroborate statistical results from the first study, and, (b) to measure qualitatively the prevalence of reasons for iPhone users not employing stronger passcodes.

7.1 Methodology

The online survey closely resembled in its structure our in-person questionnaire (Section 5.1). We just added questions for collecting descriptive statistics about the reasons for not using stronger passcodes. Appendix C.1 provides our online survey.

We recruited participants on Amazon Mechanical Turk (MTurk) [23] between February and March 2015. We limited MTurk workers to the US participants with HIT approval rate at 90% and above. Before running the study, we conducted a pilot with 149 MTurk participants to test the data collection in general and the survey questions in particular.

In comparison with the first two studies, which were conducted in-person, the online survey made it challenging to validate whether or not a participant had an iPhone and used the unlocking mechanism as she claimed to. To mitigate this concern, the participants were asked during the survey to submit two photos: (1) a photo of their iPhone reflection in a mirror taken with the front-facing camera, and (2) a screenshot of the unlocking interface. Examples of verification photos that our participants submitted are shown at Figure 4. We later used these photos to validate the claimed iPhone model (i.e., iPhone 4, 4S, 5S) and the locking mechanism. In addition, we also asked participants to provide us with the model number, e.g., ME302C/A,[4] which has one-to-one correspondence with the marketed model, e.g., iPhone 5S. We excluded responses from all those participants who either did not provide us with photos or who provided photos that did not match their choices in the survey. Finally, we also used attention check question, similarly to the one we used in Study I, in order to check if the participant read instructions carefully. This time, it was revised to improve the wording (see question 36 in Appendix C.1). We paid $1.00 to all the participants, including those who failed the attention check question or iPhone model verification or unlocking mechanism verification.

7.2 Results

7.2.1 Demographics

Overall, we recruited 1,219 participants and assigned them to Touch ID and non-Touch ID groups, depending on whether they reported using Touch ID or not. At the end, responses from 374 participants were taking into account during the data analysis, 31% of the ones who were recruited.

Non-Touch ID group. 698 participants have started the survey in the non-Touch ID group, and 550 finished it. On average it took each of them about 16.3 minutes (SD = 7.5 minutes) to finish the survey. Note that we excluded seven participants that spent more

[4] Device model can be found in the Model field of iPhone's Settings in `General->About` section.

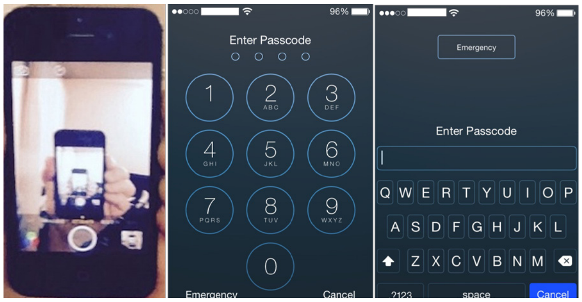


Figure 4: Examples of verification photos that participants sent us. From left to right, (1) a photo of an iPhone taken with front facing camera in a mirror, (2) a screenshot of PIN based iPhone unlock interface, and (3) a screenshot of password-based iPhone unlock interface.

than an hour finishing the survey. 317 participants failed to submit correct photos of the iPhone and screen shots of the locking interface, which left us with 226 eligible participants. Finally, 25 out of 226 participants failed the attention check question, which reduced the non-Touch ID group size to **201** participants, i.e., 37% of all participants that finished the survey.

Touch ID group. 521 participants have started the survey, and 445 finished it. On average it took about 15.7 minutes (SD = 6.2 minutes) for participants to answer the questions. Similarly to non-Touch ID group, we excluded unqualified participants. In particular, we excluded five participants that spent over an hour to finish the study, and all the participants who failed to submit a proper proof of an iPhone and locking mechanism screenshot. We also excluded all the participants who failed the attention check question, which reduced our participant pool down to **173** participants, 39% of those who finished.

The participants' demographics are shown in column "Study III" of Table 2. We recruited participants from various occupations, ranging from agriculture to public administration. The participants' job titles also included various positions, such as managers, students, team leaders and others. Our participants had diverse education levels, including 75 participants with Ph.D. or Masters degrees. More than 50% of the participants were between 25 and 34 years old. Finally, our participants had various income levels.

7.3 Testing Hypotheses

In H_1, we hypothesized that, due to the usability of Touch ID, users would switch from PIN to passwords with a bigger search space, in order to increase the work required for a brute-force attack. We first used Chi-square test to check if the proportions of users who used PINs and passwords in both groups were different. The results of the statistical analysis did not reveal any statistically significant difference ($\chi = 0.01$, p = 0.92).

The 95% percentile confidence interval for the difference between the means of passcode entropies in two groups was [-1.91, +0.95]. That implies that in case if, hypothetically, there is a difference and we just failed to reveal it, due to small sample size, then with 95% confidence we can state that the difference between mean entropies of passcodes in Touch ID and non-Touch ID groups would be 1.91 bits at most. Analysis with t-test did not reveal any statistically significant difference (t= -0.66, p = 0.51) between the non-Touch ID (M = 14.13 bits, s = 5.04) and Touch ID (M = 14.61 bits, s = 8.20) groups. Due to the results of these statistical tests, we could not reject H_1^{null}.

Similarly to Study I, we estimated the amount of work an attacker will need to do, on average, in order to brute-force the whole passcode space for Touch ID group, assuming the best case scenario for defenders, i.e., iPhone users. Considering the observed average passcode entropy in Touch ID group (14.61 bits) and the maximum possible difference between the groups (i.e., 1.91 bits), we can easily obtain the maximum possible average entropy in the Touch ID group (with 95% confidence), which is 16.52 bits.[5] Considering that for testing each passcode candidate on iPhones, an attacker must spend at least 80ms, they can brute-force the whole search space of 16.52 bits in size in about 2 hours.

In order to test H_2 hypothesis, we split all 18 participants in the non-Touch ID group who did not lock their device on those who had Touch ID (4) and who did not (14). The results of Chi-square test did not reveal any statistically significant difference (χ = 3.78, p = 0.05) between the proportions in the two groups. Thus, we could not verify the correlation between the presence of Touch ID on the phone and the user's willingness to lock their device with a passcode.

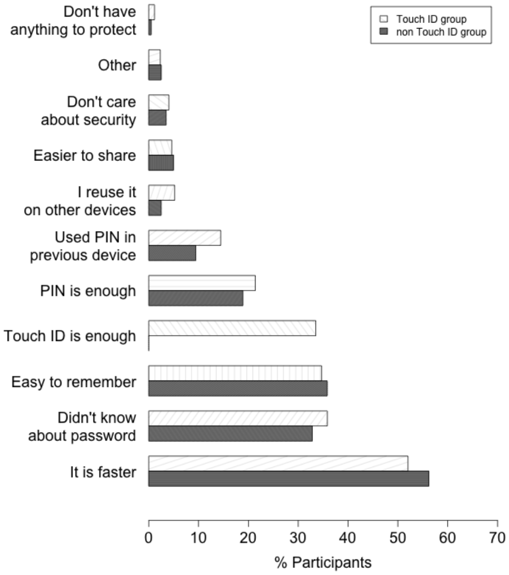


Figure 5: Reasons for using PIN instead of password for each group.

7.4 Reasons for using PIN

In both groups, we asked users for reasons why they used a PIN rather than a password. A summary of participants' answers is shown in Figure 5. Note, that for this analysis we excluded the last option, i.e., "Touch ID is enough", from both groups, since it was only present in Touch ID group. Our analysis did not reveal any statistically significant difference in distributions of answers between the two groups (χ-squared = 4.88, p = 0.85).

The results of the statistical analysis suggested that users in both groups use similar reasons for using PINs. We found that the top most three reasons were either related to usability of PINs, i.e., "It is faster" and "It is easier to remember", or to the gap in knowledge, i.e., "Did not know about the password". Finally, in Touch ID group, more than 25% of participants stated that Touch ID was

[5] As with Study I, this was an overestimation and real difference of search spaces is likely smaller. We chose to overestimate the search space to show the upper bound, i.e., the maximum work an attacker needs to do on average.

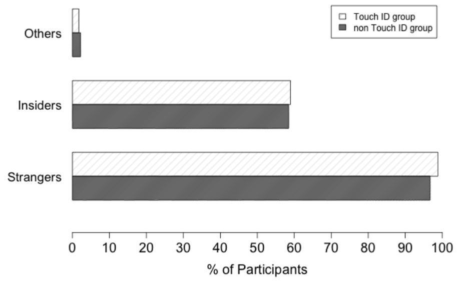


Figure 6: Distribution of actors (insiders and strangers) who our participants locked their iPhones against.

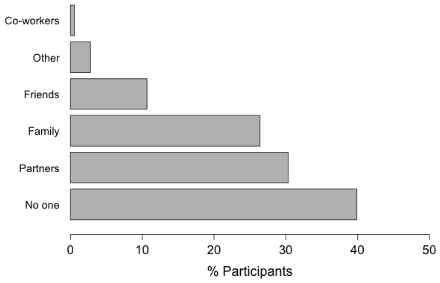


Figure 7: Distribution of passcode sharing with various groups of people (N = 374).

good enough for them from the security perspective.

7.5 Reasons for using Touch ID

Participants selected speed, convenience, and ease of use as the top three reasons for using Touch ID. Furthermore, more than 50% of participants stated that *security* provided by Touch ID was one of the reasons to use it. This suggests that the key factors that drive the adoption of Touch ID are due to its usability and perceived security. The summary of participants' answers is provided in Figure 10.

7.6 Who Users Lock their iPhones Against

The distribution of participants' answers to the question that asked who they locked their iPhone against is shown at Figure 9 in Appendix C.2. Our analysis did not reveal any statistically significant difference between Touch ID and Non-Touch ID groups (χ-squared = 9.98, p = 0.13). Interestingly, almost all participants in both groups stated that they wanted to protect their device against *strangers* (see Figure 6). At the same time, participants were also concerned with *insiders*. For instance, around 40% in both groups locked their device against co-workers, around 30% locked their phone against friends and family members, and around 20% locked their phones against classmates and roommates. These results are in line with previously reported findings [32].

We also asked participants for how long they would want their data to be protected in case if someone steals their iPhone and tries to brute-force the passcode, in order to decrypt data. See question 27 in Appendix C.1.1 for the options that we gave to choose from. Note that for the observed average passcode entropy (i.e., about 15 bits, see Section 7.3 for details), we can show that it takes less than 44 minutes to search through the whole password space. Comparing the results with what users desired, we found, surprisingly, that the preferences of only 12% of our participants matched the strength of the actual protection. The remaining 88%, however, preferred the data to be protected for more than an hour. Even more, 48% of participants wanted the data to be protected for 40 years or *indefinitely*.

7.7 Passcode Sharing Behaviour

We asked our participants who they shared their iPhone passcodes with (Figure 7). We did not find any statistically significant difference in sharing habits between non-Touch ID and Touch ID groups (χ-squared = 3.00, p = 0.70), thus, in our report we combined both groups. Overall, we found that 40% of participants did not share their passcodes with anyone. Others shared with different categories of related people. In particular, more than 25% of participants shared their passcodes with a partner or other family members. About 10% shared their passcodes with friends, while almost no one shared their passcodes with co-workers. In addition, 61% of all participants stated that they knew unlocking secret of someone else's smartphone.

7.8 Limitations

Online surveys have several limitations. First of all, in this study we asked Amazon Mechanical Turk participants to take a picture of their phone and send us a screenshot. This requirement introduced a bias in our survey toward more technically savvy users. We tried to mitigate this limitation by providing detailed instructions on how to make a screenshot. Second, Mechanical Turk users do not necessarily represent general iPhone users, thus, any generalization of the results from this study should be done carefully.

8. DISCUSSION

We first discuss the main result of the work, that is, the lack of correlation between the use of Touch ID and passcode entropy. We then proceed with a discussion of reasons why users do not take advantage of Touch ID and continue using weak passcodes. Finally, we conclude with a discussion of possible approaches to address the low adoption of stronger passcodes.

8.1 No Effect

Surprisingly, we did not find any statistically significant difference between the entropies of passcodes of those users who use Touch ID and those who do not. In addition, the results of our study suggest that the availability of Touch ID does not increase the ratio of users who lock their devices. At least, the effect is so small that we could not measure it. Under the assumption that use of Touch ID does result in the increase of passcode entropy by 1.91 bits (cf. see Section 7.3), our estimates show that, on average, an attacker would need to spend around 2 hours to brute-force the whole space of passcodes in an on-device attack. For the observed average entropies in both groups, i.e., around 15 bits, the attacker would only need 44 minutes to search through the whole space, which would meet desires of only 12% of our participants from Study III.

8.2 Reasons for Using 4-digit PINs

The second and the third studies allowed us to get a better understanding of reasons for users to stick with PINs. In particular, the results suggest that the main factors are (a) the lack of awareness that one can use password, and (b) usability considerations, e.g., ease of remembering, sharing, and typing. For instance, we found that more than 30% of our participants did not know that they can use passwords instead of PINs. Currently, during device initialization with iOS 8.3, one can setup only a PIN, even if Touch ID sensor is turned on. If the user wants to switch from PIN to password, she must do so by navigating to the corresponding settings, and only after her iPhone setup is finished. Even more, our interview study revealed that some users have been guided by salespersons with setting up the device lock, hence, have not explored the passcode setup options. These findings suggest that currently the password option lacks visibility. First, this option should be made available during the setup process. Second, users should be told about this option, if they are assisted by a salesperson at the time of setup.

The remaining participants, approximately 70%, used PINs due to their higher usability. For example, more than 50% of participants stated that they used PINs, as they are faster to type than passwords. Furthermore, about 45% used PINs because they are easier to remember. This suggests that more research is needed to find a usable authentication method that allows users to create secrets that are stronger than PINs yet just as memorable. In addition, new methods should have speed and accuracy comparable to PINs. For instance, an investigation of passcode-composition policy affects, similar to the one by Komanduri et al. [27], can be conducted with a focus on smartphone unlocking. Also, an option of providing users with feedback on passcode strength might be a promising direction for future research.

Finally, we found that over 55% of participants share their passcodes with someone else, such as family members, friends, partners, etc. Participants stressed that they did so in order to enable those people to access their devices in case of emergencies. Given that Touch ID allows registering up to five fingers, it would be interesting to see if Touch ID could actually facilitate such sharing, possibly in a more secure way. In addition, our participants indicated that they were concerned that locking an iPhone makes it impossible to call from it a dedicated number, specified by the owner, when a lost phone is found. This suggests that certain features are still missing from the current mobile OSs.

8.3 Recommendations

We envision several approaches for improving the current state of passcode selection, when Touch ID available. First, considering that the user can only use a PIN during the setup of a new iPhone, Apple should allow or request users to create stronger passcodes when they set Touch ID. Also, if sales personnel helps users to setup their iPhones, they should explain to the customers the weaknesses of PINs and let them know about the password option.

Second, the results of our study suggest that most users do not understand how Touch ID works and how it impacts the security of the data-at-rest. In particular, our participants did not understand that Touch ID is just another path in the unlocking procedure and has no impact on the physical security of their iPhones. One possible way to address this lack of understanding is by providing a better system image that facilitates the development of an adequate mental model. For example, showing that the time span of the data-at-rest protection depends only on the passcode might be one such improvement.

Third, the feedback on passcode strength can also be improved. Results of our investigation suggest that currently the preferences of only 12% of users roughly match the strength of the actual protection provided by their passcodes. It would be interesting to see if feedback on passcode strength might help users to choose appropriate passcodes.

Fourth, persuasion might be an effective option. For example, iPhone can show statistics to the Touch ID user on how often they actually use their passcode and suggest choosing a stronger passcode. Also, in order to alleviate the difficulty of retaining infrequently used passcodes in long-term memory [5], the OS can ask the user to type their passcode once every 2-3 days, in locations where it is easy to do so, e.g., at home or in office, but not on a bus, or in a car, or while the user is walking. Finally, one can employ gamification methods to motivate the choice of stronger passcodes, e.g., the user can get something (app, music, game, iCloud storage) for free as a reward.

Last but not least, our findings suggest that current mobile OSs miss important features that could impact users' choice with regards to locking their devices. In particular, the owner should be able to specify another phone number that can be called from a locked phone if someone finds it, in order to facilitate a return.

In addition, the ability to share device in a usable and secure fashion appears to be another important factor that impacts users' choice of passcodes. By providing a secure and usable way to share a smartphone, developers can enable users to pick stronger passcodes, yet being able to share their devices easily.

9. CONCLUSION

In this paper, we presented our investigation of Touch ID's impact on passcodes used for unlocking iPhones. To characterize the impact, we conducted three user studies (a) an in-person survey with 90 subjects, (b) an interview-based study with 21 participants, and (c) an online survey with 374 subjects. The results of user studies did not reveal any correlation between the use of Touch ID and the strength of users' passcodes. In particular, we observed that the average entropy was 15 bits, which corresponds to 44 minutes of work for an attacker to brute-force the whole search space, in order to find the correct password. Surprisingly, the preferences of only 12% of our participants matched the strength of the actual protection provided by passcodes. We also found that more than 30% of participants did not know that they can use alphanumeric passwords to lock their iPhones.

Based on the results of our investigation, we suggest research directions to improve the awareness of Touch ID users of the impact of stronger passcodes on data-at-rest security and to increase the visibility of the password option. We plan to investigate the proposed research directions in future work.

10. ACKNOWLEDGMENTS

We would like to thank our reviewers and our colleagues for their help and constructive feedback on earlier versions of this paper. Constructive suggestions of our SOUPS shepherd Simson Garfinkel were very helpful in improving the paper.

11. REFERENCES

[1] D. Abalenkovs, P. Bondarenko, V. K. Pathapati, A. Nordbø, D. Piatkivskyi, J. E. Rekdal, and P. B. Ruthven. Mobile forensics: Comparison of extraction and analyzing methods of ios and android. *Master Thesis, GjÃ¿vik University College*, 2012.

[2] A. A. Al-Daraiseh, D. Al Omari, H. Al Hamid, N. Hamad, and R. Althemali. Effectiveness of iphone's touch id: Ksa case study. *(IJACSA) International Journal of Advanced Computer Science and Applications*, 6(1):154–161, 2015.

[3] Amitay. Most common iphone passcodes. http://danielamitay.com/blog/2011/6/13/most-common-iphone-passcodes, June 2011. last accessed March 8, 2015.

[4] Apple. iOS Security, 8.1 and up. http://www.apple.com/business/docs/iOS_Security_Guide.pdf, 2014. Accessed April 26, 2015.

[5] A. D. Baddeley. *Human memory: Theory and practice*. Psychology Press, 1997.

[6] P. Bao, J. Pierce, S. Whittaker, and S. Zhai. Smart phone use by non-mobile business users. In *Proceedings of the 13th International Conference on Human Computer Interaction with Mobile Devices and Services*, pages 445–454. ACM, 2011.

[7] C. Bhagavatula, B. Ur, K. Iacovino, S. M. Kywe, L. F. Cranor, and M. Savvides. Biometric authentication on iphone and android: Usability, perceptions, and influences on adoption. *USEC '15*, February 2015.

[8] J. Bonneau. The science of guessing: analyzing an anonymized corpus of 70 million passwords. In *Security and Privacy (SP), 2012 IEEE Symposium on*, pages 538–552. IEEE, 2012.

[9] J. Bonneau, C. Herley, P. C. Van Oorschot, and F. Stajano. The quest to replace passwords: A framework for comparative evaluation of web authentication schemes. In *Security and Privacy (SP), 2012 IEEE Symposium on*, pages 553–567. IEEE, 2012.

[10] F. Breitinger and C. Nickel. User survey on phone security and usage. In *BIOSIG*, pages 139–144, 2010.

[11] CNN. iPhone encryption stops FBI, but not this 7-year-old. http://money.cnn.com/2014/12/01/technology/security/apple-iphone-encryption-fingerprint, December 2014. last accessed June 13, 2015.

[12] H. Crawford and K. Renaud. Understanding user perceptions of transparent authentication on a mobile device. *Journal of Trust Management*, 1(1):7, 2014.

[13] A. De Luca, A. Hang, E. von Zezschwitz, and H. Hussmann. I feel like i'm taking selfies all day! towards understanding biometric authentication on smartphones. In *CHI'15*, Seoul, Korea, 2015.

[14] S. Egelman, S. Jain, R. Portnoff, K. Liao, S. Consolvo, and D. Wagner. Are you ready to lock? understanding user motivations for smartphone locking behaviors. In *Proceedings of the 2014 ACM SIGSAC Conference on Computer & Communications Security, CCS*, volume 14, 2014.

[15] Ericsson. Ericsson mobility report. http://www.ericsson.com/res/docs/2014/ericsson-mobility-report-june-2014.pdf, June 2014. last accessed June 25, 2013.

[16] R. D. Findling and R. Mayrhofer. Towards face unlock: on the difficulty of reliably detecting faces on mobile phones. In *Proceedings of the 10th International Conference on Advances in Mobile Computing & Multimedia*, pages 275–280. ACM, 2012.

[17] D. Florencio and C. Herley. A large-scale study of web password habits. In *WWW '07: Proceedings of the 16th International Conference on World Wide Web*, pages 657–666, New York, NY, USA, 2007. ACM.

[18] D. Florêncio and C. Herley. Where do security policies come from? In *Proceedings of the Sixth Symposium on Usable Privacy and Security*, SOUPS '10, pages 10:1–10:14, New York, NY, USA, 2010. ACM.

[19] M. Gao, X. Hu, B. Cao, and D. Li. Fingerprint sensors in mobile devices. In *Industrial Electronics and Applications (ICIEA), 2014 IEEE 9th Conference on*, pages 1437–1440. IEEE, 2014.

[20] Google. Ice cream sandwich. https://developer.android.com/about/versions/android-4.0-highlights.html, March 2011. last accessed March 8, 2015.

[21] M. Harbach, E. von Zezschwitz, A. Fichtner, A. D. Luca, and M. Smith. It's a hard lock life: A field study of smartphone (un)locking behavior and risk perception. In *Symposium On Usable Privacy and Security (SOUPS 2014)*, pages 213–230, Menlo Park, CA, July 2014. USENIX Association.

[22] C. Herley and P. Van Oorschot. A research agenda acknowledging the persistence of passwords. *Security & Privacy, IEEE*, 10(1):28–36, 2012.

[23] Amazon Mechanical Turk. https://www.mturk.com/, 2005.

[24] M. Jakobsson and R. Akavipat. Rethinking passwords to adapt to constrained keyboards. *Proc. IEEE MoST*, 2012.

[25] M. Jakobsson, E. Shi, P. Golle, and R. Chow. Implicit authentication for mobile devices. In *Proceedings of the 4th USENIX conference on Hot topics in security*, HotSec'09, Berkeley, CA, USA, 2009. USENIX Association.

[26] S. Karthikeyan, S. Feng, A. Rao, and N. Sadeh. Smartphone fingerprint authentication versus pins: A usability study (cmu-cylab-14-012). *CMU-CyLab*, pages 14–012, July 31 2014.

[27] S. Komanduri, R. Shay, P. G. Kelley, M. L. Mazurek, L. Bauer, N. Christin, L. F. Cranor, and S. Egelman. Of passwords and people: measuring the effect of password-composition policies. In *Proceedings of the 2011 annual conference on Human factors in computing systems*, CHI '11, pages 2595–2604, New York, NY, USA, 2011. ACM.

[28] S. Lee and S. Zhai. The performance of touch screen soft buttons. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 309–318. ACM, 2009.

[29] I. Lookout. Lost and found: The challenges of finding your lost or stolen phone. http://blog.mylookout.com/2011/07/lost-and-found-the-challenges-of-finding-your-lost-or-stolen-phone/. last accessed August 18, 2011.

[30] V. Matyáš and Z. Říha. Biometric authentication—security and usability. In *Advanced Communications and Multimedia Security*, pages 227–239. Springer, 2002.

[31] I. Muslukhov, Y. Boshmaf, C. Kuo, J. Lester, and K. Beznosov. Understanding users' requirements for data protection in smartphones. In *Workshop on Secure Data Management on Smartphones and Mobiles*, 2012.

[32] I. Muslukhov, Y. Boshmaf, C. Kuo, J. Lester, and K. Beznosov. Know your enemy: the risk of unauthorized access in smartphones by insiders. In *Proceedings of the 15th*

international conference on Human-computer interaction with mobile devices and services*, MobileHCI '13, pages 271–280, New York, NY, USA, 2013. ACM.

[33] M. A. Sasse. Red-eye blink, bendy shuffle, and the yuck factor: A user experience of biometric airport systems. *Security & Privacy, IEEE*, 5(3):78–81, 2007.

[34] F. Schaub, R. Deyhle, and M. Weber. Password entry usability and shoulder surfing susceptibility on different smartphone platforms. In *Proceedings of the 11th International Conference on Mobile and Ubiquitous Multimedia*, page 13. ACM, 2012.

[35] R. Shay, S. Komanduri, P. G. Kelley, P. G. Leon, M. L. Mazurek, L. Bauer, N. Christin, and L. F. Cranor. Encountering stronger password requirements: user attitudes and behaviors. In *Proceedings of the Sixth Symposium on Usable Privacy and Security*, SOUPS '10, pages 2:1–2:20, New York, NY, USA, 2010. ACM.

[36] A. Skillen and M. Mannan. On implementing deniable storage encryption for mobile devices. In *Proceedings of the 20th Annual Network and Distributed System Security Symposium*, NDSS Symposium'13, San Diego, CA, USA, 2013.

[37] M. F. Theofanos, R. J. Micheals, and B. C. Stanton. Biometrics systems include users. *Systems Journal, IEEE*, 3(4):461–468, 2009.

[38] S. J. Tipton, D. J. White II, C. Sershon, and Y. B. Choi. iOS security and privacy: Authentication methods, permissions, and potential pitfalls with touch id. *International Journal of Computer and Information Technology*, 03(03), May 2014.

[39] T. Trimpe. Fingerprint basics. http://sciencespot.net/Media/FrnsScience/fingerprintbasicscard.pdf, June 2009. last accessed March 5, 2015.

[40] C. S. Weir, G. Douglas, M. Carruthers, and M. Jack. User perceptions of security, convenience and usability for ebanking authentication tokens. *Computers & Security*, 28(1):47–62, 2009.

[41] H. Wimberly and L. M. Liebrock. Using fingerprint authentication to reduce system security: An empirical study. In *Security and Privacy (SP), 2011 IEEE Symposium on*, pages 32–46. IEEE, 2011.

[42] J. Zdziarski. Identifying back doors, attack points, and surveillance mechanisms in iOS devices. *Digital Investigation*, 11(1):3–19, 2014.

APPENDIX

A. STUDY I: SUPLEMENTAL MATERIALS

A.1 Questionnaire

A.1.1 Inperson Interaction Script

1. Introduce yourself, your affiliation and give an overview of the study: "The purpose of this study is to investigate how users interact with iPhones. We aim to investigate users' motivation for choosing passwords and using fingerprint unlock. You will be asked to answer the questionnaire on iPad. It will take approximately 15 minutes. Please feel free to provide any comments and feedback on the study".
2. Verify that the participant has iPhone.
3. After the participant read and agreed with the consent form, asked her to read and sign a payment receipt and hand her a honorarium payment of $10.
4. After a participant completed the survey, conduct short exit interview asking PIN users "Why do you use 4-digit PIN, not alphanumeric password?" and password users "Why do you use alphanumeric password, not PIN?".
5. Verify the length of the password and auto-lock time.
6. Debrief.

A.1.2 Questions for both conditions

1. What is your age? [6]
2. What is your gender?
   (a) Female
   (b) Male
   (c) Prefer not to answer
3. What is your highest level of completed education?
   (a) High school
   (b) College degree
   (c) Bachelor
   (d) Master or PhD
   (e) Other, please specify
4. What industry have you worked for the past 6 months?
   (a) Agriculture
   (b) Forestry, fishing, mining, quarrying, oil and gas
   (c) Utilities
   (d) Construction
   (e) Manufacturing
   (f) Trade
   (g) Transportation and warehousing
   (h) Finance, insurance, real estate and leasing
   (i) Professional, scientific and technical services
   (j) Business, building and other support services
   (k) Educational services
   (l) Healthcare and social assistance
   (m) Information, culture and recreation
   (n) Accommodation and food services
   (o) Public administration
   (p) Other
5. What is the annual income of your household?
   (a) Less than $20,000
   (b) Above $20,000, below $50,000
   (c) Above $50,000, below $80,000
   (d) Above $80,000, below $120,000
   (e) Above $120,000
   (f) Prefer not to answer

[6] Questions that does not have suggested possible answers are open-ended questions

6. Have you ever lost your smartphone?
   (a) Yes
   (b) No
7. Have you been a victim of smartphone theft?
   (a) Yes
   (b) No
8. In your opinion, what unlocking method is more secure?
   (a) Multi-character password
   (b) 4-digit PIN
   (c) Fingerprint unlock (Touch ID)
   (d) Eye recognition
   (e) Face recognition
   (f) None of them
   (g) I have no idea
9. You are willing to use face recognition authentication
   (a) Strongly disagree
   (b) Disagree
   (c) Agree
   (d) Strongly agree
   (e) I don't know
10. Please explain your answer to the previous question.
11. What is the model of your iPhone?
    (a) 5s, 6 or 6 Plus
    (b) 5c or earlier model
    (c) I am not sure
    (d) Other, please specify
12. Do you use the same password for your iPhone as you used in your previous smartphone?
    (a) Yes
    (b) No
    (c) N/A
    (d) Prefer not to answer
13. How often do you change your PIN or password?
    (a) Weekly
    (b) Monthly
    (c) Every six months
    (d) Once a year
    (e) Never
    (f) I don't know
14. Enter a structure of your iPhone password. That is, substitute each digit (single digit number) with D, lowercase with L, uppercase with U, special character with S. For example structure for password A1b%B is UDLSU.
15. For how long have you been using an iPhone during last 5 years?
    (a) Less than a year
    (b) 1 to 2 years
    (c) 2 to 3 years
    (d) Over 3 years
16. Does your iPhone store any sensitive or confidential information?
    (a) Yes
    (b) No
    (c) I have no idea
17. What is the worst thing that could happen to your iPhone?
    (a) My iPhone gets broken or stolen, but I recover my data, so nobody will get access to my data
    (b) Someone get access to the data on my iPhone
    (c) Someone misuses my apps and account
    (d) Other, please specify
18. On average, how frequently do you unlock your iPhone?
    (a) Once a day
    (b) Few times a day
    (c) Once per hour
    (d) Few times per hour
    (e) I have no idea
19. What is your iPhone auto lock time (how long the screen stays on if the device is not being used)?
    (a) Never
    (b) 1 min
    (c) 2 min
    (d) 3 min
    (e) 4 min
    (f) 5 min
    (g) I don't know
20. A simple password is a 4-digit number. Do you know how to turn simple password off in the settings?
    (a) Yes
    (b) No
21. Have you ever shared your iPhone password with anybody else?
    (a) Yes
    (b) No
    (c) Maybe
22. Do you know anybody else smartphone security lock?
    (a) Yes
    (b) No
    (c) Maybe
23. What motivates you to lock your iPhone? Select all that apply.
    (a) My friends lock their phones
    (b) Locking prevents strangers from using my iPhone
    (c) It's easy to lock
    (d) Locking controls when my family or friends can use my iPhone
    (e) Other, please specify
24. (alternative) Why do you choose not to lock your iPhone? Select all that apply.
    (a) Information on my iPhone is useless
    (b) In case of loss, I can easily be contacted
    (c) It is too much effort
    (d) In case of emergency, others can use my iPhone
    (e) None of the above
    (f) Other, please specify
25. What kind of smartphone did you own before iPhone?
    (a) Android
    (b) Windows Phone
    (c) iPhone
    (d) BlackBerry
    (e) None of them
    (f) Other, please specify
26. What security lock have you used for your old smartphone?
    (a) Multi-character password
    (b) 4-digit PIN
    (c) Fingerprint unlock (Touch ID)
    (d) Pattern Lock
    (e) Face recognition
    (f) I didn't use a lock
    (g) I didn't have a smartphone
    (h) Other, please specify
27. Enter a structure of your previous smartphone password. That is, substitute each digit (single digit number) with D, lowercase with L, uppercase with U, special character with S. For example structure for password A1b%B is UDLSU.

28. Please select the option ’no answer’ for this question. How long did you feel this survey was?
    (a) Very long
    (b) Long
    (c) Neither short nor long
    (d) Very short
    (e) No answer

A.1.3 Questions for Touch ID group

1. How hard was it to set up Touch ID?
    (a) Very difficult
    (b) Difficult
    (c) Decent
    (d) Easy
    (e) Very easy
2. Is it easy to use Touch ID?
    (a) Very difficult
    (b) Difficult
    (c) Decent
    (d) Easy
    (e) Very easy
3. Why do you use Touch ID?
    (a) Convenience
    (b) Novelty
    (c) Security
    (d) Time
    (e) Ease of use
    (f) Reliability
    (g) Privacy
    (h) Cool to use
    (i) Fun to use
    (j) Other, please specify
4. Have you ever had issues with using Touch ID?
    (a) Yes
    (b) No
    (c) I don’t know
5. In your own experience, what situations are best suited for using Touch ID? Select all that apply. Answers are in random order for each survey.
    (a) Driving
    (b) Walking
    (c) Sitting
    (d) When using only one hand
    (e) When it’s dark
    (f) When the owner is intoxicated
    (g) Other, please specify
6. What situations are NOT suitable for using Touch ID? Select all that apply. Answers are in random order for each survey.
    (a) Driving
    (b) Walking
    (c) Sitting
    (d) When using only one hand
    (e) When it’s dark
    (f) When the owner is intoxicated
    (g) Other, please specify
7. Does use of Touch ID affect your privacy?
    (a) Yes
    (b) No
    (c) I don’t know
8. What is your major security or privacy concern about Touch ID?
9. What kind of limitations do you experience because of using Touch ID?
10. What kind of situations Touch ID should be temporarily disabled according to your own experience?
11. You feel that it is easy to circumvent Touch ID
    (a) Very difficult
    (b) Difficult
    (c) Decent
    (d) Easy
    (e) Very easy
12. Would you recommend using Touch ID to your friend?
    (a) Yes
    (b) Maybe
    (c) No
13. Please explain your answer to the previous question.
14. Overall, how satisfied are you with using Touch ID?
    (a) I hate it
    (b) I dislike it
    (c) I’m OK with it
    (d) I like it
    (e) I love it!

A.1.4 Questions for non-Touch ID

1. Have you ever used a biometric authentication system?
    (a) Yes
    (b) No
    (c) I don’t know what is biometric authentication
    (d) I’m not sure I used biometric authentication
2. In general, what are your major security or privacy concerns about biometric authentication?
3. You are willing to use face recognition authentication
    (a) Strongly disagree
    (b) Disagree
    (c) Agree
    (d) Strongly agree
    (e) I don’t know
4. Please explain your answer to the previous question.
5. You are willing to use fingerprint authentication
    (a) Strongly disagree
    (b) Disagree
    (c) Agree
    (d) Strongly agree
    (e) I don’t know
6. Please explain your answer to the previous question.
7. Would you start using longer alphanumeric password alongside with using of fingerprint scanner?
    (a) Yes
    (b) Maybe
    (c) No
    (d) I don’t know

A.1.5 Final instructions for both groups

Please follow the instructions in the order given below:

1. Lock your iPhone.
2. Turn your iPhone on.
3. Swipe to unlock.
4. Enter your password (DO NOT PRESS ‘DONE’).
5. Show your masked password to the researcher (we just want to count number of characters).
6. Navigate to the ‘Settings’–>‘General’ and show the auto-lock interval to the researcher.

Thank you for your participation!

A.2 Additional Results

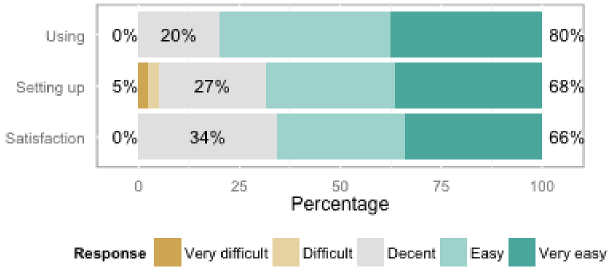


Figure 8: Touch ID participants' answers for questions "How hard was it to set up Touch ID?", "Is it easy to use Touch ID?" and "Overall, how satisfied are you with using Touch ID?" (n = 41).

B. STUDY II: INTERVIEW GUIDE

B.1 Agenda

1. Introduce yourself, your affiliation and give an overview of the study: "The purpose of the study is to investigate how users interact with iPhones. We aim to investigate users' motivation for choosing passwords and using fingerprint scanner. t will take approximately 15 minutes. Please feel free to provide any comments and feedback on the study".
2. Verify that a participant has iPhone 5S, 6 or 6 Plus with her.
3. Ask her to unlock her iPhone without using Touch ID.
4. Ask the participant to read and sign the consent form.
5. Turn on audio recording.
6. When interview is over, turn off audio recording.
7. Ask the participant to fill out a demographics form.
8. Ask the participant to sign a receipt form.

B.2 Questions

1. Lets talk about your use of Touch ID:
   (a) Why do you use Touch ID?
   (b) How do you think Touch ID works?
   (c) Do you know if you can use Touch ID without a password/PIN?
   (d) How do you think Touch ID impacts the security of your device in case it gets stolen? [Ask to elaborate. Clarify that after Touch ID recognizes the fingerprint, it restores PIN or password and unlocks device using PIN or password]
2. Password vs. PIN code section:
   (a) Can I ask you if the password/PIN code that unlocks your iPhone is being used anywhere else? [Other Devices, Web-Sites, Credit Cards, other online services]
   (b) Do you share your password/PIN with anyone else, like family members, friends of colleagues? [YES] Why do you do that?
   (c) Do you know how to switch iPhone lock from PIN to password? [Please, show me how to do that]
   (d) Did you change your password/PIN after you started using Touch ID enabled iPhone? Why [for both cases]?
   (e) Why do you use PIN, not password? (OR Why do you use password, not PIN?)
3. Let's talk about how you use your iPhone:
   (a) What is the most valuable in your phone for you? How about your data? [Ask to elaborate on data types]
   (b) Is there any data that you consider to be confidential, private or sensitive? [Ask to provide some examples]
   (c) Who do you care protecting your private data against? [Strangers, Co-workers, Friends, Family]
4. Lets consider the following scenario: "Someone stole your iPhone. He is trying to get into it to get access to your data by guessing your PIN or password. Also, he is very careful, and removed SIM card so that your iPhone is not connected to the Internet." For how long would you like your iPhone to be able to protect your [sensitive, confidential, private] data in hands of such criminal?

C. STUDY III: SUPLEMENTAL MATERIALS

C.1 Survey Questions

C.1.1 Questions for both groups

1. What is the model of your iPhone?
   (a) 3G, 3GS, 4, 4S, 5 or 5c
   (b) 5s, 6 or 6 Plus
   (c) I don't know
   (d) Other, please specify
2. What is the model number of your iPhone? You can find the model number in the About screen on your iPhone. Choose Settings > General > About.
3. How often do you change your PIN/password?
   (a) Hourly
   (b) Daily
   (c) Weekly
   (d) Monthly
   (e) Every six months
   (f) Once a year
   (g) Never
   (h) I don't use either PIN or password
   (i) I don't know
4. When did you change your iPhone PIN password last time?
   (a) 1-2 hours ago
   (b) 1-2 days ago
   (c) 1-2 weeks ago
   (d) 3-4 weeks ago
   (e) 1-2 months ago
   (f) 3-6 months ago
   (g) 6-12 months ago
   (h) More than 12 months ago
   (i) Never
5. When did you change last but one iPhone PIN/password?
   (a) 1-2 hours ago
   (b) 1-2 days ago
   (c) 1-2 weeks ago
   (d) 3-4 weeks ago
   (e) 1-2 months ago
   (f) 3-6 months ago
   (g) 6-12 months ago
   (h) More than 12 months ago
   (i) Never
6. For how long in total have you been using iPhone?
   (a) Less than a year
   (b) 1 to 2 years
   (c) 2 to 3 years
   (d) Over 3 years
7. What is the worst thing that could happen to your iPhone?
   (a) My iPhone gets broken, but I recover my data
   (b) My iPhone gets broken, but I do not recover my data

(c) Someone steals my iPhone and gets access to my iPhone data, my apps or my accounts
(d) Other, please specify

8. On average, how frequently do you unlock your iPhone?
   (a) Once a day
   (b) A few times a day
   (c) Once per hour
   (d) A few times per hour
   (e) I have no idea
9. What is your iPhone auto lock time (i.e. how long does the screen stay on if the device is not being used)? You can find iPhone auto lock time in Settings > General > Auto-Lock.
   (a) Never
   (b) 1 min
   (c) 2 min
   (d) 3 min
   (e) 4 min
   (f) 5 min
   (g) I don't know
10. Do you use 4-digit PIN or alphanumeric password for unlocking your iPhone?
   (a) PIN
   (b) Password > Please enter the structure of your iPhone password. That is, substitute each single digit number with D, lowercase with L, uppercase with U, special character with S. For example the structure for password A1b%B is UDLSU
   (c) Neither
11. What motivates you to lock your iPhone? Select all that apply.
   (a) My friends lock their phones.
   (b) Locking makes my iPhone inaccessible in case I lose it.
   (c) Its easy to lock
   (d) Locking gives me control over when my family or friends want to use my iPhone
   (e) Other, please specify
12. (Optional) Why do you choose not to lock your iPhone? Select all that apply.
   (a) Information on my iPhone is not sensitive and I do not care if others look into it
   (b) In case of loss, I can easily be contacted
   (c) It is too much effort to lock
   (d) In case of emergency, others can use my iPhone to call my family and friends
   (e) I never lose sight of my iPhone, it's always with me
   (f) Other, please specify
13. Do you use the same PIN/password for your iPhone as you used in your previous smartphone?
   (a) Yes
   (b) I did not use PIN/password in my previous smartphone.
   (c) This is my first phone.
   (d) No
14. Do you use your iPhone PIN/password anywhere else (for web sites, credit cards, other online services)?
   (a) Yes
   (b) No
15. Do you share your iPhone PIN/password with anyone else, e.g. family members, friends of colleagues?
   (a) Yes > Who do you share you iPhone PIN/password with? Family, Friends, Co-workers, Partners, No one, Other.
   (b) No
   (c) Other, please specify
16. Do you know anybody else smartphone security lock?
   (a) Yes
   (b) No
17. Does your iPhone store any sensitive or confidential information?
   (a) Yes
   (b) No
   (c) I don't know
18. Who do you care protecting your private data against?
   (a) Strangers
   (b) Co-workers
   (c) Friends
   (d) Family
   (e) Classmates
   (f) Roommates
   (g) Other, please specify
19. What kind of smartphone did you owe or use right before your current iPhone?
   (a) Feature phone
   (b) Android
   (c) Windows Phone
   (d) iPhone
   (e) BlackBerry
   (f) None
   (g) Other, please specify
20. What security lock have you used for your old smartphone? Select all that apply.
   (a) Alphanumeric password > Enter the structure of your previous smartphone password. That is, substitute each single digit number with D, lowercase with L, uppercase with U, special character with S. For example the structure for password A1b%B is UDLSU.
   (b) Long PIN (PIN with 5 or more digits)
   (c) 4-digit PIN
   (d) Fingerprints (Touch ID)
   (e) Pattern
   (f) Face recognition
   (g) I didn't use a lock
   (h) I didn't have a smartphone
   (i) Other, please specify
21. In your opinion, what unlocking method provides the best security for your iPhone?
   (a) Alphanumeric password
   (b) 4-digit PIN
   (c) Fingerprint scanner (Touch ID) + 4-digit PIN
   (d) Fingerprint scanner (Touch ID) + alphanumeric password
   (e) Other, please specify
22. Do you know that you can use alphanumeric password for unlocking your iPhone?
   (a) Yes > Please, provide exact steps how you can turn on alphanumeric password
   (b) No
23. Please, rate your agreement with the following statements. PIN is good enough for unlocking the iPhone
   (a) Strongly disagree
   (b) Disagree
   (c) Neutral
   (d) Agree
   (e) Strongly agree
24. My iPhone is more secure if I use Touch ID than PIN/password alone.
   (a) Strongly disagree
   (b) Disagree
   (c) Neutral
   (d) Agree

(e) Strongly agree

25.

For PIN participants: Why do you use 4-digit PIN, not alphanumeric password?

(a) Touch ID is enough to protect my iPhone, so I do not see a reason why I should use a password
(b) I didn't know that there is an alphanumeric password option
(c) PIN is easier to remember
(d) PIN is faster to type
(e) PIN is easier to share
(f) I continue with PIN, because I used PIN in my previous smartphone(s)
(g) PIN provides enough security for my iPhone
(h) I use the same PIN for multiple devices or accounts
(i) I do not care about security of my iPhone
(j) I do not have any sensitive data on my iPhone that I need to protect
(k) Other, please specify

For password participants: Why do you use alphanumeric password, not 4-digit PIN?

(a) Password is more secure than PIN.
(b) My company requires me to use password.
(c) I continue with password, because I used password in my previous smartphone.
(d) Other, please specify

26. What do you think the most common way for an attacker to break into your iPhone?

(a) Guessing (aka brute-forcing) PIN/password to unlock your iPhone
(b) Using social engineering to learn your PIN/password
(c) Shoulder surfing
(d) Other, please specify:

26. Lets consider the following scenario: "Someone has stolen your iPhone. He is trying to get into your iPhone to get access to your data. She is doing so by guessing your PIN/password. Also, she is very careful, and removed SIM card so that your iPhone is not connected to the Internet. Thus, you can not remotely wipe or 'kill' your iPhone." For how long would you like your iPhone to be able to protect your data in hands of such criminal?

(a) SLIDEBAR [0-1h-3h-6-12-1d-2d-3d-1w-2w-1m-2m-6m-1y-2y-5y-10y-20y-40y-indefinitely]

27. What is your gender?

(a) Female
(b) Male
(c) Prefer not to answer

28. What is your age?

(a) 19-24
(b) 25-34
(c) 35-44
(d) 45-54
(e) 55-64
(f) 65 or older

29. What is your highest level of completed education?

(a) High school
(b) College degree
(c) Bachelor
(d) Master or PhD
(e) Other, please specify

30. What industry have you worked for the past 6 months?

(a) Agriculture
(b) Forestry, fishing, mining, quarrying, oil and gas
(c) Utilities
(d) Construction
(e) Manufacturing
(f) Trade
(g) Transportation and warehousing
(h) Finance, insurance, real estate and leasing
(i) Professional, scientific and technical services
(j) Business, building and other support services
(k) Educational services
(l) Healthcare and social assistance
(m) Information, culture and recreation
(n) Accommodation and food services
(o) Public administration
(p) Other services, please specify

31. What is your job title?

32. What is the annual income of your household?

(a) Less than $20,000
(b) Above $20,000, below $50,000
(c) Above $50,000, below $80,000
(d) Above $80,000, below $120,000
(e) Above $120,000
(f) Prefer not to answer

33. Have you ever lost your smartphone?

(a) Yes
(b) No

34. Have you ever been a victim of smartphone theft?

(a) Yes
(b) No

35. Have you ever experienced a situation when somebody has unauthorizedly used your iPhone for data access or making a call?

(a) Yes
(b) No

36. You have almost completed the survey. We have to make sure that our data are valid and not biased. Specifically, we are interested in whether you read instructions closely. Please select the option 'no answer' for this question. How long did you feel this survey was?

(a) Very long
(b) Long
(c) Neither short nor long
(d) Very short
(e) No answer

C.1.2 Questions for non-Touch ID group

1. Biometrics authentication is used in computer science as a form of identification and access control. Examples include fingerprint and face recognition Have you ever used a biometric authentication system?

(a) Yes
(b) No
(c) I'm not sure I used biometric authentication

2. In general, what are your major security or privacy concerns about biometric authentication?

3. Please, rate your agreement with the following statements. I am willing to use face recognition authentication

(a) Strongly disagree
(b) Disagree
(c) Neutral
(d) Agree
(e) Strongly agree

4. I am willing to use fingerprint authentication like Touch ID

(a) Strongly disagree
(b) Disagree
(c) Neutral

(d) Agree
(e) Strongly agree

5. I am willing to use a longer alphanumeric password alongside the fingerprint scanner such as Touch ID
   (a) Strongly disagree
   (b) Disagree
   (c) Neutral
   (d) Agree
   (e) Strongly agree

C.1.3 Questions for Touch ID group

1. Why do you use Touch ID? Select all that apply.
   (a) Convenience
   (b) Novelty
   (c) Security
   (d) Time/speed
   (e) Ease of use
   (f) Reliability
   (g) Privacy
   (h) Efficiency
   (i) Cool to use
   (j) Fun to use
   (k) Other, please specify
2. Please, rate your agreement with the following statements. PIN is good enough for unlocking the iPhone
   (a) Strongly disagree
   (b) Disagree
   (c) Neutral
   (d) Agree
   (e) Strongly agree
3. My iPhone is more secure if I use Touch ID than PIN/password alone.
   (a) Strongly disagree
   (b) Disagree
   (c) Neutral
   (d) Agree
   (e) Strongly agree
4. It was difficult for me to set up Touch ID
   (a) Strongly disagree
   (b) Disagree
   (c) Neutral
   (d) Agree
   (e) Strongly agree
   (f) I did not set it up
5. It is easy for me to use Touch ID
   (a) Strongly disagree
   (b) Disagree
   (c) Neutral
   (d) Agree
   (e) Strongly agree
6. Overall, I am satisfied with using Touch ID
   (a) Strongly disagree
   (b) Disagree
   (c) Neutral
   (d) Agree
   (e) Strongly agree

C.2 Additional Results

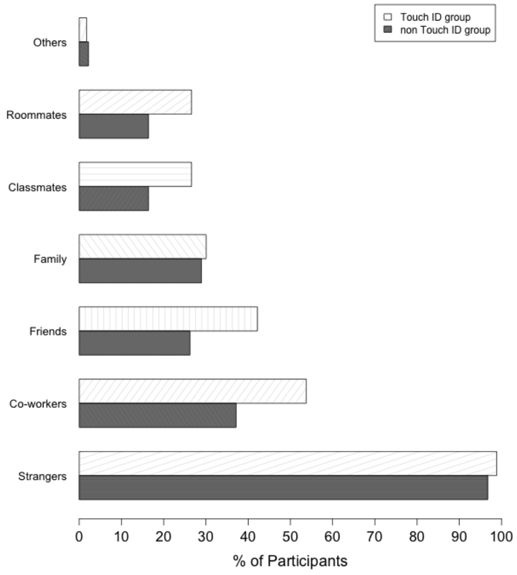


Figure 9: Actors who users lock their iPhones against, for Touch ID (n = 173), and non-Touch ID (n = 201) groups.

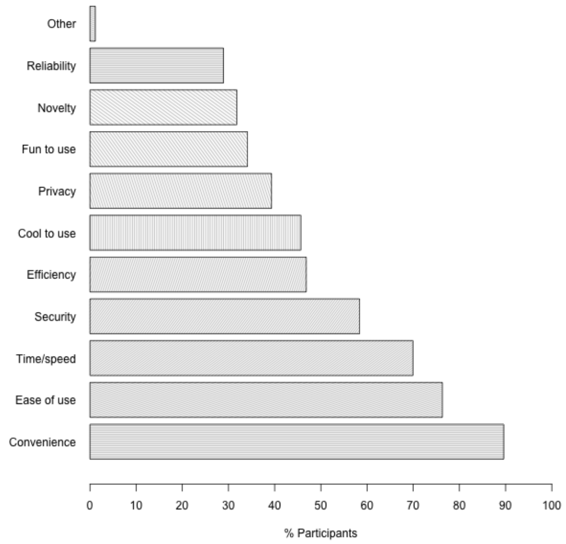


Figure 10: Reasons for using Touch ID (n = 173).

Learning Assigned Secrets for Unlocking Mobile Devices

Stuart Schechter
Microsoft
stuart.schechter@microsoft.com

Joseph Bonneau
Stanford University & EFF
jbonneau@cs.stanford.edu

ABSTRACT

Nearly all smartphones and tablets support unlocking with a short user-chosen secret: *e.g.,* a numeric PIN or a pattern. To address users' tendency to choose guessable PINs and patterns, we compare two approaches for helping users learn assigned random secrets. In one approach, built on our prior work [16], we assign users a second numeric PIN and, during each login, we require them to enter it after their chosen PIN. In a new approach, we re-arrange the digits on the keypad so that the user's chosen PIN appears on an assigned random sequence of key positions. We performed experiments with over a thousand participants to compare these two repetition-learning approaches to simple user-chosen PINs and assigned PINs that users are required to learn immediately at account set-up time. Almost all of the participants using either repetition-learning approach learned their assigned secrets quickly and could recall them three days after the study. Those using the new mapping approach were less likely to write down their secret. Surprisingly, the learning process was less time consuming for those required to enter an extra PIN.

1. Introduction

Text passwords are no longer the dominant form of device authentication. Sales of smart phones, which almost universally support numeric PINs, far exceed sales of PCs. Tablets are also poised to overtake PCs in sales [53]. Regardless of whether these devices run Android, iOS, or Windows, they support authentication via simple device-unlock secrets, namely numeric PINs or graphical passwords. Even devices with fingerprint unlock typically fall back to secret-based authentication periodically for additional security or when fingerprints cannot be read.

While PINs and text passwords are both static user-chosen secrets, the reduced length and character set allow PINs to be entered in less time and on smaller screens, meeting usability requirements for mobile-device unlocking that text passwords cannot. Although mobile-device unlocking has stricter usability requirements, the consequences of having a mobile device compromised may be as dire as for computers with keyboards. Many companies allow their employees to access email and other services from their mobile devices. Mobile devices also now serve a critical role as second factors in website authentication, whether through access to users' emails and text messages or through dedicated applications for generating one-time codes.

To protect device-unlock secrets, most operating systems restrict guessing, either by limiting the frequency with which unlocks can be attempted, by erasing devices if too many consecutive unlock attempts fail, or both. Yet attackers may succeed even when restricted to a few guesses. They may exploit users' tendency to choose easy-to-remember but common numeric sequences (*1234*), repeat the same key (*9999*), or choose a path of adjacent keys (*2580*). They may guess the four-digit birth-year of the user or the users' loved ones—an estimated 25% of screen unlock codes are based on a date of some form, and 7% of respondents in a prior study admitted to using their own birthday as their banking PIN [18]. Leaked data on PINs chosen to unlock an iPhone application suggests an attacker with three guesses could expect as high as a 9.23% chance of success [18].

If, instead of choosing their own device-unlock secret, users were assigned a random four-digit PIN attackers would only have a 0.03% chance of guessing the PIN in three attempts. Furthermore, assigning secrets prevents users from re-using a PIN they have previously used elsewhere—though it does not prevent users from later re-using their assigned PIN elsewhere. If users are allowed to choose their own PINs, they may re-use the same PIN that they use for their ATM card, their voicemail, their frequent-flyer account, or their gym locker (in which they may store their phone or other mobile device). Indeed, over a third of respondents in a prior survey reported re-using their banking PINs for some other purpose [18].

We recently demonstrated a prototype ceremony for teaching users a random 56-bit secret, encoded as 12 characters or 6 words, using spaced repetition[16]. We integrated the ceremony into an existing website login process. Each time participants logged into the study website and verified their password, we displayed the text of the secret we had assigned to them and required them to type it into a text field. Each time they logged in, we added a progressively longer delay delay before revealing the secret for them to copy. Eventually, most participants began to recall and type their secrets reliably from memory.

While the prior approach was successful for participants using keyboards, we questioned whether users would accept the requirement to enter two PINs on a mobile device while learning an assigned code. Users expect their phones and tablets to unlock quickly, with minimal finger movement and delay. An in-situ study estimated typical users unlock their

Symposium on Usable Privacy and Security (SOUPS) 2015, July 22–24, 2015, Ottawa, Canada.

phones nearly 50 times per day [35] and the time taken to unlock is already a significant annoyance to users.

In this paper we compare the prior approach to a new approach that does not require users to type additional keys to learn an assigned secret.

2. Learning Assigned Secrets via Mappings

Whereas our previous repetition-learning approach [16] requires users to enter the assigned secret after their chosen secret, our new approach teachers users a new assigned secret while they enter their chosen secret. The key idea is to assign users a random sequence, but provide their chosen secret as a guide to highlight the correct sequence. This is done by choosing a new random mapping from digits to positions on the keypad for each digit of the user's chosen PIN. The user simply needs to press each digit of their chosen PIN in sequence on four random-looking keypads, with the keypad appearing to shuffle before the entry of each digit. We illustrate our approach in Figure 5.

It is important that while the mapping of digits to keys changes with each key entered (between indexes in the sequence) it remains the same from login to login. This ensures that the user will be pressing the same sequence of key positions with each login. The assigned secret is this (random) sequence of positions on the keypad. We also provide letters for each position on the keypad which do not change—the assigned secret can equivalently be thought of as the sequence of letters needed to be pressed.

Essentially, our approach allows the user to enter their chosen secret and assigned secret at the same time. By using a random mapping, we ensure that even if users' chosen secret are drawn from an arbitrarily skewed distribution (including, as a degenerate case, if all users choose the exact same PIN) the distribution of assigned secrets will always be a uniform distribution over all possible sequences.

Since the pattern of keys in the assigned secret is the same each time a user logs in, we hypothesized that users would learn their assigned patterns from habit. To encourage users to learn, and to detect when learning had occurred, we add a delay before the digits appears that grows as the learning progresses. We draw arrows from key to key as users enter their assigned key sequences, making the visual pattern more salient. In the event that the same key appears two or more times in a row, we use within-key circular arrows. Line segments earlier in the sequence appear faded relative to those later in the sequence.

As with the prior scheme, attackers who obtain the device during the teaching period need only guess the user's chosen PIN in order to authenticate as the user. However, after learning the mapping can be destroyed and, unless attackers were already able to obtain it, knowledge of users' likely PIN choices will yield no benefit in guessing the assigned sequence of keys.

3. Related Work

3.1. Random passwords and PINs

It is now well-established in the research literature that humans will choose a skewed distribution of passwords or other secrets when given free choice [15, 42, 56]. This effect is robust across demographic groups [17, 49] and is impacted only marginally when users are more motivated to pick a strong password [15], are given stricter composition policies [42, 44] or are nudged to choose better passwords [70].

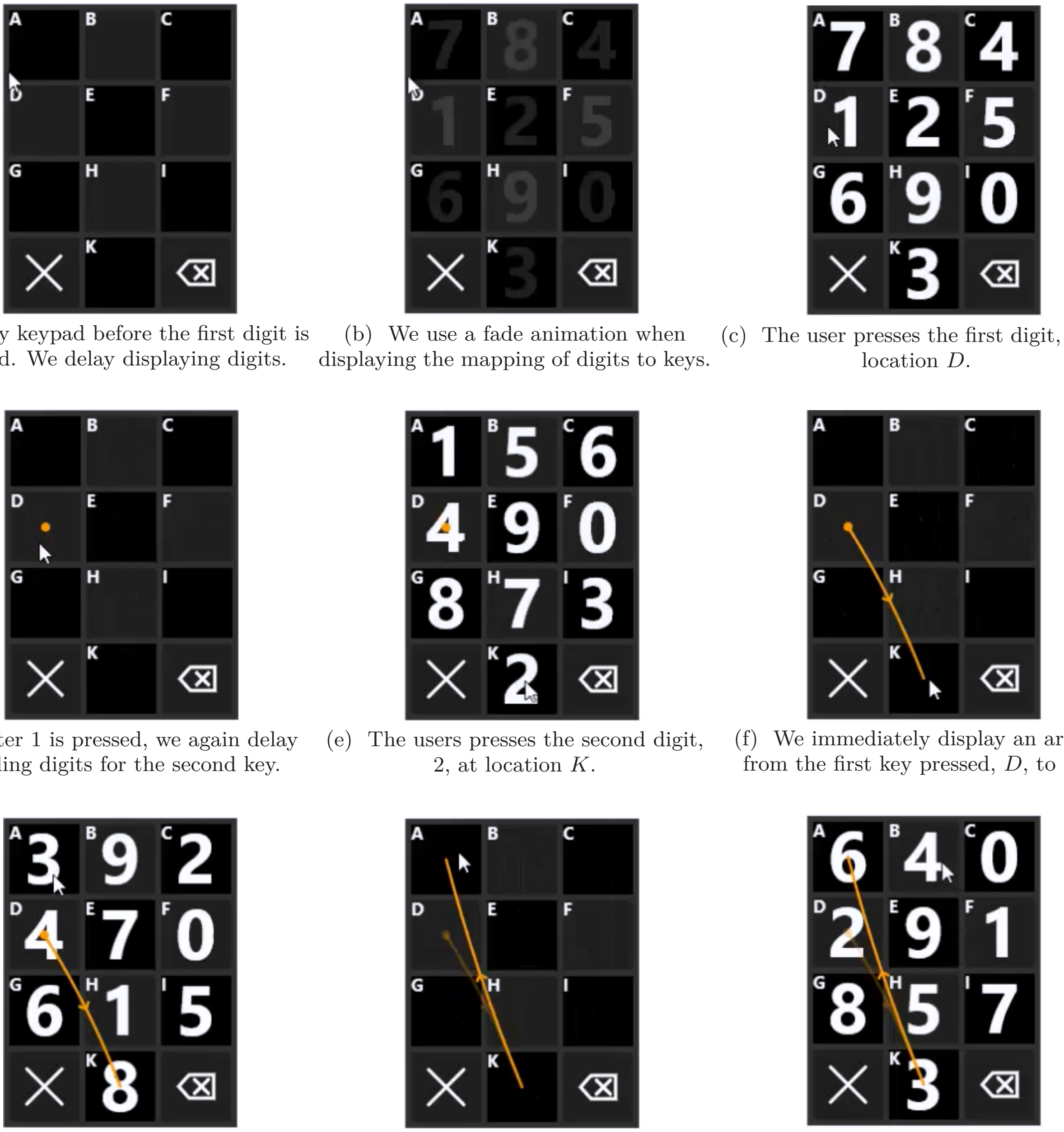
Random passwords. In response to persistent problems with weak human-chosen text passwords, a number of schemes have been proposed for encoding random passwords in such a way as to make memorization easier. Surprisingly, the few studies which have directly compared recall rates of user-generated passwords to assigned passwords have not found statistically strong evidence that users are less likely to remember assigned passwords than self-chosen passwords when no learning period is used [19, 63, 81, 85]. Various encodings have been proposed ranging from generating random but pronounceable nonsense words [1, 34, 80], choosing a list of random words from a dictionary [6, 46], generating a random grammatical sentence [7, 41] or even generating a random song [55]. No studies have actually validated that these encodings are more memorable than random character strings. Two studies which compared users' ability to recall random passwords under different encodings found no conclusive differences in memorability between random alphanumeric strings, random pronounceable strings or randomly generated passphrases [48, 62].

Spaced repetition. The hypothesis emerging from these results is that strong passwords, whether user-chosen or assigned, are not highly memorable without a learning period. Over a century of psychological research supports that spaced repetition [10, 20, 31] is the most reliable way to form long-term memories. While many other factors have been identified which affect the rate of memory formation, such as the depth of neural processing required during rehearsals [24] or the encoding of information in multiple forms [57], repetition is the most powerful and robust effect.

In recent work, we demonstrated the promise of spaced repetition for learning strong, 56-bit random text passwords for authentication [16], achieving 80–90% recall after a learning period of at most 15 days was followed by a period of at least three days during which the secret was not used. A subsequent study by Blocki et al. [13] demonstrated the effects over a longer period of time and with multiple passwords being memorized, observing recall rates close to 80% over a period of 180 days for a more complicated interface with graphical prompts for multi-word passphrases.

Numeric PINs. Relatively little has been published on numeric PINs. Historically, banking PINs were machine-chosen for technical reasons as well as security ones. Banks gradually began allowing user-chosen PINs in the 1980s as a marketing gimmick—they are now predominant. Bonneau et al. presented perhaps the only publicly-available estimates of the distribution of human-chosen PINs based on leaked data from an iPhone application developer, leaked web password data and surveys [18]. Their work highlighted that users' tendency to pick dates is the biggest source of skew in the data, consistent with research on dates chosen in text passwords [75]. Little work has focused on the memorability of random PINs; one exception is a pilot study by Huh et al. [38] which found memorability declines for longer PINs, although this could be improved by chunking them into smaller groups.

Graphical password schemes. A large variety of graphical passwords schemes have also been proposed [12, 59, 66]. Graphical passwords attempt to capitalize on the human brain's relatively strong visual memory, though different schemes have different goals. Traditionally the three ma-



A B C D E F G H I K
A 7 B 8 C 4 D 1 E 2 F 5 G 6 H 9 I 0 K 3
A 7 B 8 C 4 D 1 E 2 F 5 G 6 H 9 I 0 K 3
A B C D E F G H I K
A 1 B 5 C 6 D 4 E 9 F 0 G 8 H 7 I 3 K 2
A B C D E F G H I K
A 3 B 9 C 2 D 4 E 7 F 0 G 6 H 1 I 5 K 8
A B C D E F G H I K
A 6 B 4 C 0 D 2 E 9 F 1 G 8 H 5 I 7 K 3

(a) They keypad before the first digit is entered. We delay displaying digits.

(b) We use a fade animation when displaying the mapping of digits to keys.

(c) The user presses the first digit, 1, at location D.

(d) After 1 is pressed, we again delay revealing digits for the second key.

(e) The users presses the second digit, 2, at location K.

(f) We immediately display an arrow from the first key pressed, D, to K.

(g) The user presses the third digit, 3, at location A.

(h) We connect the arrow path from $D \rightarrow K \rightarrow A$.

(i) The user completes the sequence by pressing digit 4 at location B.

Figure 1. In our mapping-based approach to learning a random secret, a user chooses her own PIN: in this example the four digits 1234. However, when she enters her PIN, the mapping of digits to keys on the keypad changes each time she enters a digit of her PIN; the letters at the top left of each key remain fixed. We choose the mapping of digits to keys so that the sequence of keys of the user's chosen PIN map to a randomly-chosen sequence of four keys: in this example the keys represented by $D \rightarrow K \rightarrow A \rightarrow B$. To encourage the user to learn to enter her key sequence without looking for the digits of her PIN, we increase the delay before we reveal the mapping of digits to keys with each login. Once the user has learned to enter the sequence of keys without seeing the digits, we can erase the mapping of digits to keys.

jor categories are recognition-based (searchmetric) schemes, in which a user recognizes previously-seen images [30, 64]; click-based (locimetric) in which users select points of interest in one or more images [21, 22, 79]; and recall-based or free-drawing schemes (drawmetric) in which the user draws an image or pattern [40, 67, 74, 78]. Note that secrets are typically assigned in recognition-based schemes, whereas recall- and click-based schemes tend to employ user-chosen secrets. A number of studies have demonstrated that user choice in graphical password schemes proposed for web authentication suffers from predictable choices much as do text password schemes [25, 68, 72, 73, 83].

Viewing a PIN-entry keypad as a visual stimulus with regions that users can press, our scheme could be considered a free-drawing recall-based graphical password. To our knowledge, no prior research has looked at how users memorize randomly-assigned secrets for free-drawing schemes.

3.2. Device authentication

User authentication for mobile devices, often simply called “device authentication” or “device (un)locking”, is a burgeoning field of research. Our study appears to be the first utilizing spaced repetition to help users memorize a random secret for device authentication, but the literature suggests a number of interesting further research questions.

Device authentication habits. Several studies have looked at why, how, and how often users unlock their devices [32, 35, 71, 76] through a combination of surveys and telemetry on user devices. Collectively, these studies have found that between 40% [35] and 70% [32, 71] of users lock their phones, with a consistent preference for graphical unlock mechanisms over numeric PINs [32, 35, 71]. This dislike for numeric PINs was noted at least as early as 2002 [23], with most users choosing not to activate PIN activation on early generation (non-touchscreen) phones when this was the only option. Interestingly, despite this preference users are actually able to unlock more quickly and reliably using numeric PINs than graphical schemes [76].

A key challenge of unlock mechanisms compared to text passwords is the very high rate at which they are used; Harbach et al.’s telemetry study found an average of nearly 48 unlocks per day out of 80 total device activations (with some not requiring an unlock due to recent use or only performing a non-sensitive action) [35]. The frequency of unlocks, particularly as many are seen as unnecessary by users, motivates our goal of making learning as lightweight as possible. Parallel work in progressive and multi-level authentication [36, 60] aims to limit the number of unlock actions required of the user by delaying them until a security-critical action is attempted. This work is orthogonal to ours as our learning could be performed whenever a device authentication is needed, though we might note that if the rate of authentications per day were to become too low the speed of learning a new secret would decrease.

Security. Uellenbeck et al. [69] provide the only public estimates of the difficulty of guessing unlock patterns for Android’s default scheme, a 3 × 3 variant of Tao et al.’s PassGo scheme [67]. By collecting patterns from a large number of users in an experimental setting and devising a dictionary to attack them, they estimate that this scheme provides roughly 8–10 bits of security for the median user and thus is roughly comparable to random 3-digit PINs. Thus all of our experimental treatments represent a security upgrade over the baseline Android scheme.

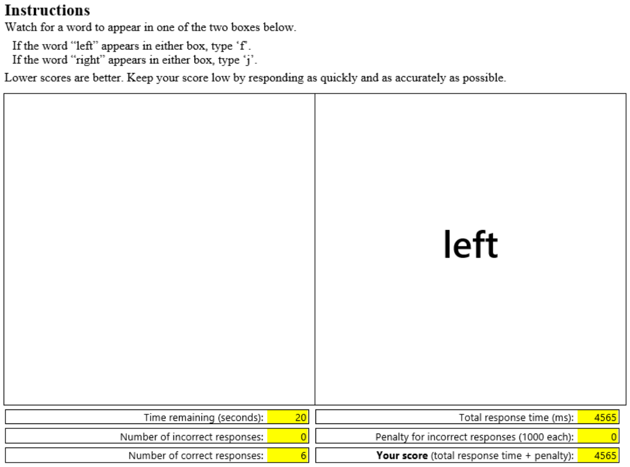
Windows touchscreen devices have used a modified click-based scheme, with a background image on which users enters a series of clicks, drags, or circles. Zhao et al. [83] studied the security of this scheme and found, depending on the background image in use, a dictionary with roughly 2^{18}–2^{20} items was sufficient to compromise the majority of users’ patterns. They also found that a smaller dictionary of 2^{10} items was sufficient to compromise over 10% of user’s patterns, indicating that even this stronger scheme still has many users picking predictable patterns for which any of our experimental treatments would be a security upgrade.

Strength meters have been proposed to nudge users towards choosing more difficult-to-guess unlock patterns [3, 65], but their effectiveness has not been established.

Other attacks on touchscreen authentication. Both PINs and graphical patterns are vulnerable to smudge attacks [9] or fingerprint attacks [82], in which residue from the user’s fingers indicates where the user touches their screen during unlock. Defending against these attacks requires either randomizing the physical pattern input during any given authentication [2, 47, 61, 77], or switching to authentication schemes which do not require touching the screen such as gaze-based authentication [26, 45] or gesture-based authentication [8, 50, 51, 58]. PINs and graphical patterns are also both vulnerable to “shoulder-surfing” or physical observation attacks [54]. A number of authentication schemes have been proposed to defend against smudge attacks and shoulder-surfing [28, 29], though none of these schemes has seen practical deployment and they all appear to impose additional burden on the user.

Other device authentication mechanisms. Other research has attempted to replace explicit device unlocking completely. Physical biometrics deployed for smartphones include Apple’s Touch ID fingerprint sensor and Android’s Face Unlock face recognition scheme. User studies of these mechanisms find that users generally prefer using a fingerprint sensor and many find using face recognition annoying or impractical in certain situations (e.g. in a dark room) [11, 52]. Interestingly, convenience and perceived speed are the dominant factors, with increased security not being a major factor motivating adoption. Indeed, Apple’s Touch ID is always configured to allow PIN or password authentication as a fallback; Apple’s own documentation states that the primary goal of the feature is to allow users to use a stronger password since they won’t have to enter it as often [5]. Similarly, Android’s Face Unlock can be overridden by PIN entry. Thus, even with increased deployment of biometrics they are currently only a convenience and helping users remember stronger unlock codes is an important goal for security.

Behavioral biometrics [39, 43] capture a user’s implicit actions using the device to detect if a different human appears to be using the device. For example, much research has shown that fine details of a user’s touchscreen use can identify them [14, 27, 33, 37, 84]. However, this approach hasn’t been deployed and it appears to inherently have sufficiently high false negative rate to require a more reliable backup authentication mechanism. Thus, our research is orthogonal to the challenge of using either explicit or implicit biometric signals to decrease the number of authentication requests imposed on the user.



Instructions

Watch for a word to appear in one of the two boxes below.

If the word “left” appears in either box, type ‘f’.
If the word “right” appears in either box, type ‘j’.

Lower scores are better. Keep your score low by responding as quickly and as accurately as possible.

left

Time remaining (seconds): 20
Number of incorrect responses: 0
Number of correct responses: 6
Total response time (ms): 4565
Penalty for incorrect responses (1000 each): 0
Your score (total response time + penalty): 4565

Figure 2. In the attention game, we asked players to press a key on the side of the keyboard that matches the word on the screen, regardless of which side the word appears on. Scores are based on response time and accuracy.

4. Methodology

In order to observe participants repeatedly entering a secret, we needed an excuse to cause them to authenticate. We asked participants to login to a study to perform the same distractor task used in our prior study: the attention game illustrated in Figure 2. We shortened the game to five attention trials (30 seconds). In each trial, we randomly choose a side of the screen (left or right) and a word (‘left’ or ‘right’). We display the chosen word in the box on the chosen side. We ask participants to type a letter on the left side of the keyboard if they see the word ‘left’ and on the right side of the keyboard if they see the word ‘right’, and to ignore the side of the screen that the word appears in.

We recruited by offering the attention game as a Human Intelligence Task (HIT) on Amazon’s Mechanical Turk. We paid US$0.10 and restricted our task to workers from the US. When workers completed the attention game we offered them the opportunity to become participants in our study. We offered $9 for 50 repetitions of the game: $0.10 per game plus a $4 bonus for completing all 50 and a survey at the end. We issued payments automatically on an hourly basis.

We required participants to wait 30 minutes between each game and gave them a total of 8 days to complete all 50 games. For each of the 50 games, participants would log in, play the attention-test game for 30 seconds, and then be shown a timer that counted down the 30 minutes until they could play again (another login would be required). This allowed us to collect data on no less than 49 authentications over 2–8 days. Participants could reload the study page and log in before they were allowed to play another game, but they would be forced to log in again after the time expired. As a result, the number of logins sometimes exceeded the number of games played.

4.1. Treatments

The process through which participants logged in to play the game depended on their treatment group. We created a total of ten treatments, which we list in Table 1 along with the probability that a participant would be randomly assigned to each treatment. We set the probabilities such that most participants (65%) would be placed in one of our four *primary* treatments: (1) used only a user-chosen PIN; (2) used only an assigned PIN to be memorized when setting up the account; (3) used a user-chosen PIN augmented with a second assigned PIN learned during a learning period, mirroring our prior work; and (4) used a user-chosen PIN mapped to a random sequence of keys—our new approach, as described in Section 2. These primary treatments were the focus of our a-priori hypotheses for which we perform statistical testing. The remaining six auxiliary treatments explore variants of the mapping-based approach.

Treatment		Length	Keys	p
Primary				
(1)	***User-Chosen***	**4**	**10**	.10
(2)	***Assigned***			.15
(3)	***Second-PIN***			.20
(4)	***Mapping***			.20
Auxiliary (variants of *Mapping*)				
(5)	*4x20 Mapping*	4	20	.05
(6)	*6x10 Mapping*	6	10	.05
(7)	*6x20 Mapping*	6	20	.05
(8)	*Instructionless*	4	10	.05
(9)	*Arrowless*	4	10	.10
(10)	*6x20 Arrowless*	6	20	.05

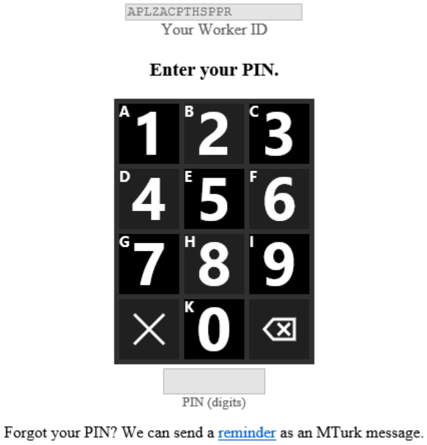
Table 1. Treatments followed by the probability that a participant would be assigned to that treatment (p). We assigned 65% of participants to our four *primary* treatments, shown in boldface: user-chosen PINs (1), traditional assigned PINs (2), a second assigned PIN entered after a chosen PIN (3), and our new mapping-based approach (4). Our a-priori hypotheses, for which we planned and ran statistical tests, focused on these four treatments. We assigned the other 35% of participants to variants of *Mapping*, examing such factors as the number of keys in the sequence (length) and the number of keys on the keypad (keys).

Regardless of treatment, participants used a standard 10-digit numeric keypad with digits placed in sequential order (left to right, top to bottom) when signing up for the study. (The same keyboard that appears in Figure 3.) The back arrow at the bottom right can be used to backspace one digit and the ‘X’ at the bottom left can be used to clear all digits. For consistency, small letters used in some treatments appear at the top left of each key regardless of treatment. For treatments (1)-(3), participants continued to use this standard 10-digit PIN-entry keypad throughout the study.

For all treatments, we provided a feature that would send participants a PIN reminder. This reminder contained the participant’s assigned PIN only for the *Assigned* treatment (for which participants did not have a chosen PIN). For all other treatments (1,3-9), the reminder contained the participant’s *chosen* PIN.

(1) User-Chosen

Participants in the *User-Chosen* treatment chose their own four-digit PIN and entered it on the standard keypad throughout the study. Since most users currently choose their own PINs, this treatment serves as a baseline of user-acceptability, speed of entry, and memorability.



APLZACPTHSPPR

Your Worker ID

Enter your PIN.

PIN (digits)

Forgot your PIN? We can send a reminder as an MTurk message.

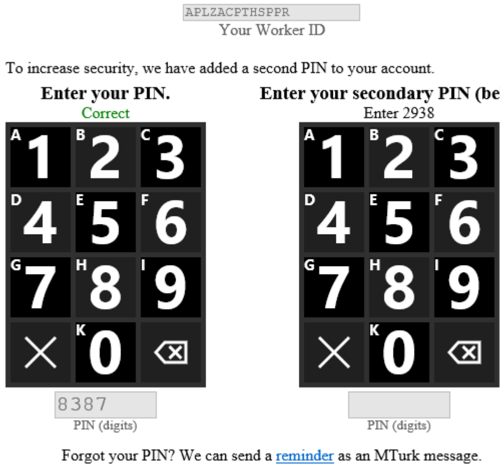
Figure 3. The PIN-entry keypad for an example participant in the *User-Chosen* treatment. The keypad would look the same for participants in the *Assigned* treatment.

(2) Assigned

For each participant in our *Assigned* treatment, we randomly generated a four-digit PIN and instructed the participant to memorize it immediately. So as to not overly disadvantage this treatment for measures of memorability, we asked participants to enter their PIN twice during sign-up.

(3) Second-PIN

We created the *Second-PIN* treatment to mirror the two-secret approach from our prior work. This treatment appeared the same as the *User-Chosen* treatment when participants signed up. However, each time participants logged in with their chosen PIN, we asked them to copy their second secret, a four-digit PIN we had generated at random and assigned to them, using a second keypad—see Figure 4. Above the PIN-entry keypad, we provided participants the following *primary* instruction for each login: "To increase security, we have added a second PIN to your account." For each attention game they had completed after the first[1] (a lower bound on the number of prior logins), we added a 1/3 second delay before revealing the secret. So, by the login for the eigth game, a participant would need to wait 2 seconds before she could see the second PIN to copy it. If the participant entered the correct digit during the period before we revealed the digits to copy, we would conclude that she had entered the digit from memory. Each time a participant entered a correct key before the delay expired, we would start the delay over so that the participant would have as much time to enter the next key from memory as she had for the last key.



APLZACPTHSPPR

Your Worker ID

To increase security, we have added a second PIN to your account.

Enter your PIN.

Correct

Enter your secondary PIN (below).

Enter 2938

8387

PIN (digits)

PIN (digits)

Forgot your PIN? We can send a reminder as an MTurk message.

Figure 4. The PIN-entry keypad for the *Second-PIN* treatment after the example participant has entered her user-chosen PIN.

After participants had completed five games, we added the following *supplemental* instruction:

> You do not need to wait for the second PIN to be written above the keypad to enter your PIN. If you recall the correct sequence of digits, you may enter it immediately.

To prevent participants from permanently tuning out the *supplemental* instruction text, we used a boldface font during logins following the 10th completed game and every 5 games after that (the 15th, 20th, ..., 45th). We removed the primary instruction after participants completed ten games. We removed both instructions immediately and permanently once the participant demonstrated that they had learned the assigned secret (by entering it before it was revealed).

(4) Mapping

The *Mapping* treatment is the baseline implementation of our new approach (Section 2), using a four-digit user-chosen secret (PIN) and an assigned secret (key/letter sequence) with 10,000 possible values. To assist learning, we employed the delay for *every* digit in the sequence. See Figure 1. As with the *Second-PIN* treatment, we used a 1/3 second additive delay.

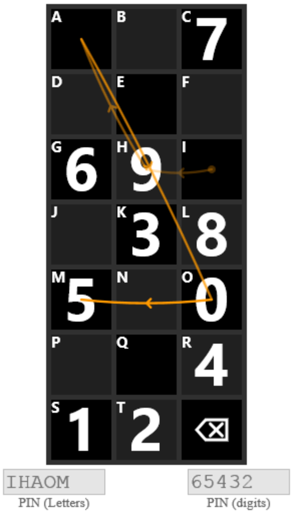
We provided participants with the following *primary* instruction (with the same timing rules as the *Second-PIN* treatment).

> To increase security, we have changed the positions of the digits on the keypad you will use to sign in. We will use the same set of positions each time you sign in. That means that the pattern you make on the keypad when you enter the PIN will be the same each time. Entering your numeric PIN also creates a sequence of letters, representing the letters on each key of your PIN. This sequence of letters also stays the same each time you enter your PIN.

We used the following *secondary* instruction:

> You do not need to wait for the digits to appear on the keypad to enter your PIN. If you recall the

[1] We adjusted timings based on the number of games played, and not the number of prior logins, because participants might log in more than once per game. For example, a participant could refresh the study website, log in, and find they needed to wait until being allowed to play another game. They would need to log in again when the next-game timeout expired. Our decision to increase delays based on the number of games participants had played may have caused login-counts to grow without as much delays as we would have liked. The alternative, triggering on the number of prior logins, would have increased delays even when logins did not have sufficient spacing between them to reinforce learning.



IHAOM
PIN (Letters)

65432
PIN (digits)

Figure 5. A six-digit mapping-based PIN on a 20-key pad. The user's chosen PIN was 654321. We assigned the user a random sequence encoded as letters $IHAOMS$. For each key the user enters, the mapping of letters to keys stays the same (see the top left corner of each key) but the digits move. We place digits so that the user's chosen PIN maps to the sequence of keys (letters) we assigned. In this screenshot we illustrate the moment at which a user had already typed 65432 at key positions $IHAOM$. Pressing the key labeled S with the 1 on it would complete the PIN. The arrows are an affordance to help users remember the pattern.

> correct pattern or sequence of letters before the digits appear on the keyboard, you may enter it immediately.

(5-10) Variants of Mapping

We created six more treatments to examine possible variations of the idea.

To test the effectiveness for memorizing more secure PINs (with more than 10,000 possible values), we created treatments (5)–(7) which encode larger secrets. In treatment (5), we expanded the PIN-entry keypad to include 7 rows. We also remove the clear function from the key on the bottom left so that digits could be placed on this key. This allowed us to double the number of usable keys from 10 to 20—see Figure 5 . The 10 digits were mapped onto these 20 keys with 10 keys left blank. Thus, while a participant would still choose a secret from a 4×10 space, the assigned secret (key/letter sequence) wass drawn from 4×20 (160,000) possible values—increasing security against guessing by a factor of 16. In treatment (6), participants chose a 6-digit PIN on a standard keypad and we assigned a six-digit secret with 6×10 possible values—increasing security against guessing by a factor of 100. In treatment (7), we combined the approaches of (5) and (6) to yield a 6×20 secret with 6,400,000 possible values—increasing security against guessing by a factor of 6,400 over the baseline *Mapping* treatment.

In treatments (8)–(10), we examined the impacts of taking away certain affordances in our design to see if they were actually needed, or just getting in the way.

When performing our prior work, we only told participants they *could* enter their secret without waiting for it to appear—we never asked them to. We were surprised how quickly they learned with such little guidance. We wondered whether it was necessary to provide any guidance at all. We created the *Instructionless* treatment (8) for which we removed both the primary instructions that explained the mapping of digits to keys and the secondary instructions that explained that the PIN could be entered before the digits appeared.

In treatments (9) and (10), we remove the arrow affordance from a 4-digit and 6x20 mapping treatments. As arrows may increase vulnerability to shoulder-surfing attacks, we would prefer to remove them if they had no benefit.

4.2. Study-completion survey

When participants logged in to complete their 50th attention game, we bypassed the game and immediately presented the completion survey.

Following standard demographic questions (language, gender, age, occupation, and level of education) we asked questions about the login process. We asked whether participants had entered their PIN "using a mouse, touch screen, or some other pointing device." We then asked if they had written or otherwise stored their PIN. To avoid confusion, we asked participants in the *Second-PIN* treatment only about their second (assigned) PIN. We then asked all participants in treatments using the mapping-based approach whether they had written/stored their assigned secret. We asked participants who reported storing their chosen or assigned secrets to explain how they had done so.

We asked participants using the mapping-based approach whether they remembered their secret as a visual pattern, a sequence of letters, or some combination of the two.

Finally, we asked participants in *all* treatments, "If you wanted to keep your phone or tablet secure, would you want to use a PIN like the kind you used to sign into our experiment's website?" For participants in treatments other than *User-Chosen*, we prefaced the question by explaining the security benefits of having an assigned PIN. For those in the *Second-PIN* treatment, we explained that a real system would allow users to discard their chosen PIN after learning—requiring only the second PIN.

We include our survey specification, with the exact wording used in the survey, as Appendix 10.

4.3. Follow-up study to test recall

Three days (between 72 and 73 hours) after each participant completed the main study, we emailed an invitation to participate in a follow-up for $0.50. The purpose of the follow-up was to determine if participants could recall their PIN after the learning period and having not used it for three days. We required only that participants log in to the study website, though we paid participants after a day's delay if they tried but failed to log in. As our goal was to measure memory after the learning period, participants in the mapping-based treatments were never shown the mapping of digits to keys, and those in the *Second-PIN* group were never shown their second PIN.

4.4. Hypotheses

We finalized the following seven hypotheses (four main hypotheses, three with two parts each) on the day we began our experiment, emailing a hash of the hypothesis statements to the program chairs as evidence that could be used to prove these were a-priori hypotheses formed before examining experimental reuslts (see Appendix B). We present each hypothesis starting with the intuition behind it, then informally, and finally as a formal statement that can be used as a specification for a hypothesis test.

H1a/b. Study-completion rates

One of the motivations for the mapping-based approach is to reduce the frustration that may result when users are forced to attempt to memorize a secret in one session (without assistance) or if they are forced to type extra keys. We hypothesized that this frustration would cause higher drop-out rates for affected groups.

> A smaller proportion of participants in *Mapping* will drop out than of participants in (a) *Assigned* and (b) *Second-PIN*.

H2a/b. Written/stored PINs

Another motivation for the mapping-based approach is that users may be less likely to write them down. We intuited that users would be more likely to write down a PIN they were forced to memorize without a gradual learning period. We further assumed a second PIN, presented compactly as a numeric string above the PIN-entry keypad, would be easier to write down than a sequence of keys and so users would be more likely to do so.

> A smaller proportion of participants in *Mapping* will report having written down their assigned secret than of participants in (a) *Assigned* and (b) *Second-PIN*.

H3. Learning time

We hypothesized that, since the mapping approach requires users to press only half as many keys, learning an assigned sequence via a mapping would consume less of users' time than learning a second numeric PIN.

> Participants who complete the study in *Mapping* will have spent less time learning their secret than those in *Second-PIN*.

We define the time spent learning a secret as the sum of the PIN-entry time of each user's authentication sessions up to, but not including, the first session in which the user entered their assigned secret without assistance (the revealing of the positions of digits for *Mapping* or the display of the second PIN). We measure the PIN-entry time from the instant the PIN-entry keypad appears until authentication is complete. We cap the time consumed by any one PIN-entry session at 60 seconds.

H4a/b. Sentiment

Finally, we thought that participants would prefer learning an assigned secret via a mapping to receiving no assistance, or to having to enter extra keys during each login.

> When asked if they would want to use this system, participants in *Mapping* will answer more positively than those in (a) *Assigned* and (b) *Second-PIN*.

Since participants had three possible responses to this question, we use an using ordinal scale to measure positive sentiment: *no*<*maybe*<*yes*.

Hypothesis testing

We measure differences in proportion using a *two-tailed* Fisher's Exact Test (FET) and differences in both times and ordinal responses using a *two-tailed* Wilcoxon test with the Mann Whitney U statistic (U). To correct for multiple testing when examining these seven hypotheses, the conservative Bonferroni method yields a threshold of significance α=.05/7=.0071).

4.5. Ethics

Since some participants in our prior experiment had figured out that authentication was a focus of our study, in this experiment we revealed that the PIN was *a* component of our study. We used the study sign-up process to inform participants about the research in place of a standard informed consent form. We did not volunteer to participants that we would later give them the opportunity to participate in a follow-up study.

Unlike our prior work, we informed participants in advance that we would pay for each attention game they played even if they did not complete all 50—though we did provide a significant bonus to those who completed the study. This is more consistent with ethical guidelines that participants should know they may leave an experiment at any time without penalty.

We paid participants at a rate designed to ensure they received at least the highest minimum wage in the US. We identified that this was research being performed by Microsoft Research. We responded to workers' requests quickly and, where terms of service allowed, monitored worker forums to identify any participant concerns we might address.

Our study was approved by Microsoft Research's institutional review system. The second author, who was not employed by Microsoft Research, contributed after the last participant had already been recruited and was not involved in the conduct of the experiments or analysis of raw data.

4.6. Known limitations

While we strove to mimic as many aspects of a typical device-authentication experience as possible, we could not do so perfectly. While many users use a PIN to authenticate to their devices more frequently than once every thirty minutes [35], others may perform fewer than 50 PIN-based authentications over an eight-day period (such as those who bypass most PIN-authentications using their fingerprint). Participants in our study may have been more or less motivated to learn the PINs than real-world users would be.

Participants may have wanted to please the researchers by giving a more positive answer to our sentiment question, which asked whether they they would want to use the PIN from the experiment for their own mobile device. For this reason, we do not compare participants' reported sentiments for the *User-Chosen* scheme to others; Participants may be more likely to believe that the less-familiar schemes that assign PINs are schemes the researchers want to succeed.

Whereas PIN-entry on modern mobile devices uses a touchscreen keypad, participants in our study used our on-screen keypad via whatever computer and input device was available to them. Peeking ahead to our results, only 27 of 782 participants who completed the study (3%) reported that they primarily used a touchscreen to enter their PIN(s). The great majority of participants, 729 (93%), reported using a mouse and 26 (3%) using some other device. We would anticipate device-unlock times would be shorter on touch screens, as users do not have to synchronize their hand movements with that of a mouse in order to press each key. While it's possible that using a device with a different form factor and input device impacted our between-group comparisons, we do not anticipate any reasons why one treatment group would disproportionately advantaged or disadvantaged.

5. Results

We offered our Human Intelligence Task (HIT) on Amazon's Mechanical Turk from 7:30PM EST on Sunday February 22, 2015 to 1:30PM on Wednesday February 25.[2] During this period, 1274 workers accepted the HIT and, of those, 1230 (97%) completed the HIT and saw the offer to sign-up for the study. Of those workers, 1016 loaded the sign-up page for the study (83% of those who completed the HIT). Since we assigned workers to a treatment group only after they arrived at the sign-up page, any departures prior to that should not be attributed to their treatment.

Of those workers who arrived at the sign-up page, 1001 (99%) completed the sign-up process to become study participants. Since more than four of every five workers who we paid to complete the attention-game HIT signed up to become study participants (1001/1230=81%), this recruiting strategy proved cost effective.

One factor contributing to the effectiveness of this recruiting approach was that many participants performed the HIT expressly because they had learned about the full study. The HIT appeared in discussions on forums for workers on Mechanical Turk, which ended up serving as feeders to our study. The forum that appeared to have the largest influence on our recruitment rate was MTurkGrind. After a pause in recruiting to ensure our funding would arrive in time to pay additional participants, we posted on that forum to let forum members know that the popular study was again open to new participants.

In monitoring these forums, we did not observe that any forum members had discovered that different participants were given different types of PINs, or any other "spoilers" that might have confounded the study. The one exception was that, towards the end, some posts revealed that the follow-up study was coming. For the most part, participants shared their best scores on the attention-game, encouraged each other, and shared their progress in completing the study.

5.1. Completion rates

Encouragement and competition on the forums may have caused a greater fraction of participants to complete the study, and possibly to do so at a faster rate, than they might have otherwise done. Another factor that may have raised completion rates is that we paid workers a higher wage than most requesters on Mechanical Turk. In the words of a participant who posted on the Turkopticon forum, "I think I got lucky to get in on this"[4].

With this in mind, it may not be surprising that we did not observe any significant differences in completion rates between treatment groups, meaning we had no support for Hypotheses 1a or 1b. As can be seen in Table 2, the proportion of workers arriving at the sign-up page who completed the study was *not* higher for the *Mapping* treatment (164 of 206, or 81%) than the *Assigned* treatment (128 of 158, or 82%, H1a: FET p=0.7395). In fact, a greater proportion of participants completed the *Assigned* treatment than the *User-Chosen* treatment (though the difference is well within the variance expected due to chance).

Treatment	Didn't sign up	Quit quickly	Quit later	Finished
Assigned	2 (1%)	10 (6%)	18 (11%)	128 (81%)
User-Chosen	4 (4%)	17 (16%)	5 (5%)	83 (76%)
Second-PIN	4 (2%)	19 (11%)	24 (13%)	132 (74%)
Mapping	3 (1%)	16 (8%)	23 (11%)	164 (80%)
4x20 *Mapping*	1 (2%)	3 (7%)	1 (2%)	40 (89%)
6x10 *Mapping*	1 (2%)	7 (11%)	7 (11%)	50 (77%)
6x20 *Mapping*	0 (0%)	1 (2%)	3 (5%)	52 (93%)
Arrowless	0 (0%)	3 (3%)	8 (9%)	80 (88%)
6x20 *Arrowless*	0 (0%)	1 (2%)	2 (4%)	54 (95%)
Instructionless	0 (0%)	2 (4%)	5 (10%)	43 (86%)

Table 2. Participants' progress through the main study. We track all workers who arrived at the sign-up page, and were assigned a treatment, as participants assigned a PIN might abandon sign-up. We say that participants quit quickly if they completed no more than three attention tests (which would require two logins) or quit later if they otherwise failed to finish the main study.

While fewer participants completed the *Second-PIN* treatment (132 of 179, 75%) as compared to *Mapping*, the difference was not significant (H1b: FET p=0.1731).

5.2. Re-use and writing down of secrets

Recall that, for all treatments other than *Assigned*, the only observable differences in behavior at the sign-in page was the length of the PIN we asked participants to choose. We asked all participants who chose a PIN, with the exception of those in *Second-PIN*, whether they had chosen (re-used) a PIN they already used elsewhere and whether they had written down, or otherwise stored, their chosen PIN.

When asked if the PIN they chose was one they had used before, 229 of the 409 (56%) participants with a four-digit PIN reported that it was, as did 59 of the 146 (40%) participants with a six-digit PIN. Since six-digit PINs are less common than four-digit PINs, it is likely that fewer participants had a six-digit PIN already memorized to reuse.

Of the 180 participants who claimed not to have re-used an existing four-digit PIN, 41 (23%) reported that they had written down or stored their new chosen PIN, as opposed to 26 of 87 (30%) for six-digit PINs.

In Table 3 we examine the proportions of participants who wrote down the random secret assigned to them, those who needed a reminder of their PIN (their chosen PIN in all treatments except the *Assigned* treatment), and those who were unable to recall and enter their secret during the follow-up study.

In the *Assigned* treatment, 62 of 128 participants (48%) either wrote/stored their secret or later required a reminder. Surprisingly, if our participants are to be believed, the majority successfully memorized their PIN simply by entering it twice on the keypad of the study sign-up page! Still, we want to minimize the risk that nearly half of users will write or otherwise store their PIN, especially since they would

[2] There was a break in recruiting to ensure sufficient funds would be available for all participants.

Treatment	Wrote assigned secret		Needed reminder		Never learned		Forgot later	
Assigned	57/128	(45%)	7/128	(5%)	~	~	0/110	(0%)
User-Chosen	~	~	0/83	(0%)	~	~	0/76	(0%)
Second-PIN	13/132	(10%)	4/132	(3%)	1/132	(1%)	0/124	(0%)
Mapping	2/164	(1%)	3/164	(2%)	1/164	(1%)	0/151	(0%)
4x20 *Mapping*	5/40	(13%)	0/40	(0%)	4/40	(10%)	0/30	(0%)
6x10 *Mapping*	3/50	(6%)	0/50	(0%)	2/50	(4%)	1/40	(3%)
6x20 *Mapping*	9/52	(17%)	0/52	(0%)	3/52	(6%)	2/43	(5%)
Arrowless	5/80	(6%)	3/80	(4%)	2/80	(3%)	1/70	(1%)
6x20 *Arrowless*	14/54	(26%)	0/54	(0%)	4/54	(7%)	2/48	(4%)
Instructionless	0/42	(0%)	0/43	(0%)	10/43	(23%)	0/28	(0%)

Table 3. The proportions of participants who reported writing down their assigned secret, requested and opened a PIN reminder, and of those who failed to login with their assigned secret during the follow-up. For the *needed reminder* column, note that our reminders contained users' *chosen* PIN for all treatments except *Assigned.* For the *forgot later* column, not that we exclude from our analysis those participants who never demonstrated learning their assigned secret (those in the *never learned* column).

likely carry a written reminder on them at the same time they were carrying their mobile device.

We had hypothesized that participants in the *Mapping* treatment would be less likely to write down their assigned secret than those in the *Assigned* and *Second-PIN* treatments. The proportion of participants who wrote or stored their assigned secret in the *Assigned* treatment (57/128, 45%) was significantly higher than the *Mapping* treatment, supporting Hypothesis 2a (2/164 1%): H2a FET p<0.0001. The proportion of those in the *Second-PIN* treatment (13/132, 10%) who wrote their assigned secret was also significantly higher than those in the *Mapping* treatment, supporting Hypothesis 2b: H2b FET p=0.0008.

5.3. Login and learning speed

We had hypothesized (naïvely, in retrospect) that participants in the *Mapping* group would spend less total time in learning – time learning their PINs – than those in the *Second-PIN* treatment. Whereas PIN-entry time for *Mapping* starts when the keypad appears and ends when the PIN is validated, the time for *Second-PIN* continues until the second PIN is validated. Recall that learning time is the sum of these PIN-entry times up to, but not including, the first login during which a participant enters their assigned secret before the secret or mapping is revealed. We present statistics elucidating the learning time in Table 4. Turning to Hypothesis 3, comparing the learning times for the *Mapping* treatment and the *Second-PIN* treatment does reveal a significant difference: H3 U=4,491.0, p<0.0001. However, the direction of the difference was the opposite of what we had hypothesized!

Participants in the *Second-PIN* treatment required fewer logins to learn their secret, and thus had a lower learning time than those in the *Mapping* treatment despite having to enter twice as many keys (8 vs. 4) per login. We suspect participants in the *Second-PIN* required fewer treatments to learn because their assigned-secret was presented as a single chunk of four digits, whereas participants in the *Mapping* treatment were presented with their assigned-secret one key at a time (only seeing the final PIN as a chunk if they paid attention to the arrows or letters).

The impact of chunking goes beyond the number of logins required to learn the secret and also impacts the PIN-entry time for each login, which we infer from Figure 6. In fact, between the second and 13th logins, participants in the *Second-PIN* group were able to enter their 8 digits in less time, on average, than participants in the *Mapping* group could enter four digits! Again, *chunking* likely plays a role. We had employed a single delay before revealing the entire chunk of four digits to participants in the *Second-PIN* treatment, whereas we had employed four delays, one before revealing the positions of digits before *each* key of the PIN, for the *Mapping* treatment.

We suspect that participants in the *Mapping* treatment were also slowed down by their need to perform visual searching. Until they learned the positions of the digits of their PIN, they would have to perform four visual searches per login: one for each key. In contrast, participants in the *Second-PIN* treatment would find their four-digit secret displayed at an easy-to-find location (above the keypad) and the keys to enter these digits were at well-known positions—no visual searching was required.

Figure 6 does show that, once participants in the *Mapping* treatment learned their assigned key sequence, login speeds for the *Mapping* treatment decline rapidly and closely approach those of chosen PINs. Since users users often choose PINs with keys in close proximity, such as 1111 or 1212, we thought some fraction of participants might be slowed down by using keys at random positions. Yet, the 5*th* percentile times for the 49*th* login, representing the fastest 1 of every 20 participants, were tiny.

For the *Second-PIN* treatment, the login time per digit that needed to be entered was par with those of other treatments, and so we would expect performance equivalent to the *Assigned* treatment once users could skip their chosen PIN. While we had designed our *Second-PIN* treatment to mirror the approach in our prior paper, in retrospect we worried that this choice may have put it at an unfair disadvantage. For all other treatments, participants enjoyed post-learning login speeds the moment they learned their PIN. For the *Second-PIN* treatment, participants had to continue entering their chosen PIN even after they had learned their assigned PIN. We later addressed this methodological shortcoming, as we will explain in Section 6.

As expected, participants in treatments with more complicated secrets required longer learning periods, were more

Treatment	Logins to learn 50 %ile	 95 %ile	Training time (sec) 50 %ile	 95 %ile	Time for 49th login (sec) 5 %ile	 50 %ile	 95 %ile
Assigned	~	~	~	~	2.16	3.32	10.32
User-Chosen	~	~	~	~	1.82	2.97	7.09
Second-PIN	7.0	15.0	81	228	4.05	6.34	17.00
Mapping	12.0	23.0	172	507	1.98	3.09	7.67
4x20 *Mapping*	16.5	~	264	~	2.47	4.46	44.64
6x10 *Mapping*	16.0	38.4	346	1,579	2.70	4.65	20.28
6x20 *Mapping*	16.0	43.0	382	1,638	3.48	5.72	60.87
Arrowless	14.0	30.7	195	801	2.25	3.71	45.60
6x20 *Arrowless*	17.0	~	405	~	3.44	6.34	47.92
Instructionless	24.0	~	495	~	2.10	3.96	110.87

Table 4. Participants' performance on speed metrics, including (1) the number of logins prior to the first login in which they typed the code without seeing it, (2) the total learning time consumed by those learning logins, and (3) the time for participants' 49th login.

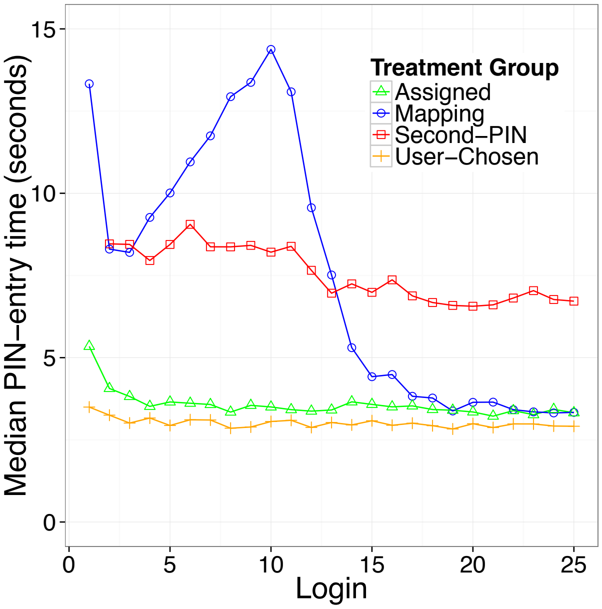


Figure 6. The median time to enter the login PIN(s) for the first through 25th login. The set of participants includes only those who completed the study. For the *Second-PIN* treatment, times include the time to enter both PINs.

Treatment	Pattern		Letters		Both		Did not	
Mapping	129	(80%)	14	(9%)	19	(12%)	2	(1%)
4x20 *Mapping*	18	(50%)	8	(22%)	10	(28%)	4	(11%)
6x10 *Mapping*	39	(81%)	5	(10%)	4	(8%)	2	(4%)
6x20 *Mapping*	22	(43%)	11	(22%)	18	(35%)	1	(2%)
Arrowless	26	(33%)	26	(33%)	28	(35%)	0	(0%)
6x20 *Arrowless*	7	(14%)	31	(62%)	12	(24%)	4	(8%)
Instructionless	31	(89%)	0	(0%)	4	(11%)	7	(20%)

Table 5. We asked participants in *Mapping* and its variants "If you learned how to enter your PIN on the keypad without waiting for digits to appear on the keys, how did you remember which keys to press?"

Treatment	No		Maybe		Yes	
Assigned	17	(13%)	44	(34%)	67	(52%)
User-Chosen	8	(10%)	25	(30%)	50	(60%)
Second-PIN	18	(14%)	43	(33%)	71	(54%)
Mapping	11	(7%)	53	(32%)	100	(61%)
4x20 *Mapping*	1	(3%)	14	(35%)	25	(63%)
6x10 *Mapping*	5	(10%)	13	(26%)	32	(64%)
6x20 *Mapping*	4	(8%)	19	(37%)	29	(56%)
Arrowless	4	(5%)	30	(38%)	46	(58%)
6x20 *Arrowless*	6	(11%)	13	(24%)	35	(65%)
Instructionless	2	(5%)	17	(40%)	23	(55%)

Table 6. We asked participants "If you wanted to keep your phone or tablet secure, would you want to use a PIN like the kind you used to sign into our experiment's website?"

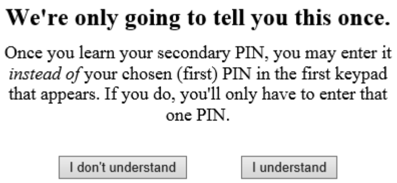
likely to find a way to write or store their assigned secret, and were more likely to forget their assigned secret later. The *Instructionless* treatment had the greatest proportion of participants who never learned, required the most logins for those who did learn, and had the highest learning time—a clear indication that systems should provide *some* guidance.

Participants assigned to use mappings with the large keyboard were more likely to use the letters to help them memorize their secret, as can be seen from Table 5. Reliance on letters grew when we required a longer PIN or removed the arrow affordance. Those assigned the most difficult mapping (6x20 *Arrowless*) were the most likely to memorize the string of letters instead of a pattern of key positions. However, without instructions to inform them that the letters could be used to assist their memories, most participants learned their sequence as a pattern of key positions.

5.4. Participant sentiments

Table 6 summarizes participants' responses to the sentiment question, which asked whether they would want to use a PIN like the kind used in the study. For *Mapping*, 100 of 164 (61%) responded *yes*. Turning the three possible responses into an ordinal sentiment score (*no*=0, *maybe*=1, *yes*=2), participants in *Mapping* responded more positively than those in *Assigned* (52% yes) and those in *Second-PIN* (54% yes) as we posited in Hypotheses 4a and 4b, but the differences did not exceed our significance threshold: H4a: U=9,381.5, p=0.0772 ; H4b: U=9,805.5, p=0.1135

Written explanations in response to the sentiment question reveal that many participants were able to grasp what we were trying to accomplish in creating the mapping-based approach. In the words of one of our pilot participants:



We're only going to tell you this once.

Once you learn your secondary PIN, you may enter it *instead of* your chosen (first) PIN in the first keypad that appears. If you do, you'll only have to enter that one PIN.

Figure 7. We presented this message on participants' first login after their third attention test. If they clicked on the "I don't understand" button we popped up an alert encouraging them to email us, then attempted to open a *mailto:* link to the study email address. None emailed us.

> This was pretty slick. I noticed I wasn't getting my PIN anymore but was still logging in. That's when I saw the pattern I had adapted to. I also realize that, in my head, I was repeating my actual credit card PIN which was not the PIN to get into the system. The numbers (and letters) were completely irrelevant and I thought that was awsome [sic].

6. Rematch: *Second-PIN* (v2) vs. *Mapping*

We conducted a second experiment to address concerns that we may have shortchanged the *Second-PIN* treatment. Users should be able to skip their chosen secret once they have learned their assigned secret. Whereas this design choice did not impact the hypotheses tested in our prior work, it may have put our *Second-PIN* treatment at an unfair disadvantage with respect to learning times and user sentiment.

We tested only two treatments in this experiment. We modified the *Second-PIN* treatment so that participants could enter their assigned secret in the first PIN-entry keypad and bypass the need to enter a second-PIN. For comparison we also included a *Mapping* treatment identical to that in our original experiment. For participants in the *Second-PIN* treatment, we presented the interstitial dialog in Figure 7 to participants just before presenting the PIN-entry keypad on the first login after the completion of their third attention test (which required a minimum of two prior logins).

Since, in the main experiment, 95% of participants in *Mapping* and *Second-PIN* had learned their secret by the 25th login, we shortened the study to 25 logins within four days for a total payment of $4.00.

Using data from our main experiment, we made one change to our calculation of learning time to better reflect the actual time lost to learning during each treatment. We subtracted 3 seconds for each learning login to account for time that would have been spent logging in even if no learning were taken place. We chose 3 seconds as it is approximately the median PIN-entry time for the user-chosen PIN treatment in the main study. This revised learning-period calculation better approximates the time users consumed due to the actual learning, excluding time the user would have had to spend logging in even if no learning were occurring.

We recruited participants for the rematch in bursts between 7:00PM EST on March 5 and 2:00PM EDT on March 9, excluding prospective participants who had participated in the earlier experiment.

We present the updated comparison of median PIN-entry times for these two groups in Table 7 and Figure 8, replicating the analyses that appear as Table 7 and Figure 6 from the main experiment. Since the *Mapping* treatment was unchanged, its results are similar to the same treatment in the previous experiment.

As expected, our improvements to the *Second-PIN* treatment made learning even faster. The median number of logins to learn (3) indicates that most participants were able to enter their assigned PIN from memory on their fourth login. In contrast, during the the main study it was only on the eighth login that we could determine that the majority of participants had learned their second PIN. It appears that, during the main study, participants in the *Second-PIN* treatment were either unable to enter the PIN quickly enough to prove knowledge of it (we revealed the digits for them to copy before they could enter the correct digit) or were less motivated to do so. For the *Mapping* treatment, in both the main and rematch experiments, the majority of participants did not enter their PIN without assistance until their 13th login.

For our rematch, the learning time for participants in the *Second-PIN* treatment was dominated by their first three logins. The first login required so much time, a median (over all participants) of 20.85 seconds, that it does not appear in our graph. The median learning time was only 40 seconds!

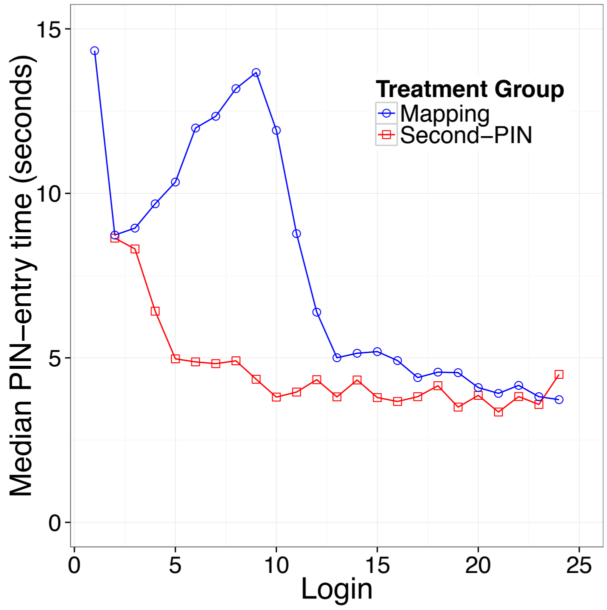


Figure 8. Rematch: median PIN-entry time for first 24 logins. Participants in the *Second-PIN* treatment could skip their chosen PIN and enter only assigned PIN. The first login time for the *Second-PIN* treatment (20.85 seconds) is outside the range of the graph.

In Table 8, we see that 20 of 73 participants in the *Second-PIN* treatment (27%) wrote down their assigned PIN. This proportion is not only greater than the *Mapping* treatment, as expected, but much greater than the *Second-PIN* treatment in the previous experiment. We fear that the cause of this difference, if not pure chance (**posthoc** FET: p=0.0011), was participants who wrote down their second PIN after learning they could use it to skip their first.

In Table 9 we summarize the rematch-study participants' responses to the sentiment question, which asked if partici-

	Logins to learn		Training time (sec)		Time for 24th login (sec)		
Treatment	**50 %ile**	**95 %ile**	**50 %ile**	**95 %ile**	**5 %ile**	**50 %ile**	**95 %ile**
Second-PIN	3.0	12.4	40	145	2.20	4.50	16.25
Mapping	12.0	26.0	117	412	2.51	3.73	33.93

Table 7. Participants' performance on speed metrics (see Table 4) for the rematch experiment.

Treatment	Wrote assigned secret		Needed reminder		Never learned	
Second-PIN	20/73	(27%)	0/73	(0%)	0/73	(0%)
Mapping	3/61	(5%)	1/61	(2%)	6/61	(10%)

Table 8. Rematch: Secret storage and recall. (Fewer *Mapping* participants learned their secret compared to the main experiment as they had half as many learning logins.)

Treatment	No		Maybe		Yes	
Second-PIN	8	(11%)	30	(42%)	34	(47%)
Mapping	7	(11%)	15	(25%)	39	(64%)

Table 9. Rematch: "If you wanted to keep your phone or tablet secure,would you want to use a PIN like the kind you used to sign into our experiment's website?"

pants would want to use the PIN scheme from the study on their mobile device. Surprisingly, we saw a drop in the desirability of the *Second-PIN*. Even had we hypothesized such a drop, we would not have the statistical strength to be able to dismiss the null hypotheses that both treatments inspire equally positive sentiment (**posthoc** U=1,874.0, p=0.1034). While there was insufficient evidence to prove a difference, it was enough for us to worry that our modifications had somehow made the *Second-PIN* treatment more annoying. To double-check, we examined participants' free-response answers and found no evidence to support this concern. The free-responses for *maybe* were consistently positive, suggesting that chance may have given us participants who rounded their scores down.

In fact, the most common concern focused on the trustworthiness of the party which generated the random PIN for the user. Since in most implementations the PIN would be generated by the device that the user is trusting to authenticate her correctly, this seems like an easy concern to overcome for any assigned secret.

7. Concluding discussion

Assigning users a random authentication secret, as opposed to letting them choose one, maximizes the difficulty of guessing (for a given alphabet/length) and prevents users from reusing prior secrets. We set out to test if spaced repetition, which we previously demonstrated for teaching users text passwords strong enough to resist extended brute-force [16], was also workable in the mobile device unlock setting for shorter, PIN-strength secrets.

We designed a new approach using randomly-assigned sequences (*Mapping*) hoping to make learning time as fast as possible which we feared would be a potential drawback of a direct application of our previous design for numeric PINs (*Second-PIN*). We in fact found the opposite, with users able to memorize a random PIN using our previous approach significantly faster than a random sequence using our new approach (particularly after the adjustment we made in the revision tested in Section 6).

However, both methods showed promise for use on mobile devices with very fast learning times. Both approaches also saw a smaller fraction of participants wrote down their assigned secret as the *Assigned* treatment, for which participants were asked to memorize a secret at sign-in time.

Our results do not yield a clear winner between *Second-PIN* and *Mapping* despite the shorter learning time for the former. *Mapping* offers the advantage that fewer users wrote their secret down, which may be attractive to system administrators who impose minimal-authentication requirements on devices used to access their systems (e.g., phones connecting to corporate email must have a PIN). Further, once participants had learned their secrets, login times for assigned secrets approached those for user-chosen secrets—remaining just a few percentage points higher.

With learning times of under a minute, the second-PIN approach requires surprisingly little effort. While a greater proportion of participants in our study reported wanting to use the *Mapping* treatment, we suspect that if prospective users knew which the approach required fewer learning logins and less visual searching, a substantial fraction of those might choose *Second-PIN*.

Acknowledgments

We are indebted to the hard-working and diligent participants from the Mechanical Turk marketplace who took part of the study, identified bugs, and took the time to write detailed explanations to survey questions. The presentation of this research benefited tremendously thanks to feedback from the anonymous reviewers, Simson Garfinkel (our shepherd), and Craig Agricola. Joseph Bonneau is supported by a Secure Usability Fellowship from Simply Secure and the Open Technology Fund.

References

[1] "Automated Password Generator (APG)". *NIST Federal Information Processing Standards Publication*, 1993.

[2] Irfan Altiok, Sebastian Uellenbeck, and Thorsten Holz. Graphneighbors: Hampering shoulder-surfing attacks on smartphones. In *Sicherheit*, pages 25–35, 2014.

[3] Panagiotis Andriotis, Theo Tryfonas, and George Oikonomou. Complexity metrics and user strength perceptions of the pattern-lock graphical authentication method. In *Human Aspects of Information Security, Privacy, and Trust*. Springer, 2014.

[4] AnotherPersona (Turkopticon Alias). Turkopticon review of microsoft research attention and pin study.

https://turkopticon.ucsd.edu/reports?id=A3A0N4OPTWRYPD, February 27, 2015.

[5] Apple, Inc. iOS Security. https://www.apple.com/business/docs/iOS_Security_Guide.pdf, April 2015.

[6] Reinhold G. Arnold. The Diceware Passphrase Home Page. , 2014.

[7] Mikhail J Atallah, Craig J McDonough, Victor Raskin, and Sergei Nirenburg. Natural Language Processing for Information Assurance and Security: An Overview and Implementations. In *Proceedings of the 2000 New Security Paradigms Workshop*. ACM, 2001.

[8] Md Tanvir Islam Aumi and Sven Kratz. AirAuth: Evaluating In-Air Hand Gestures for Authentication. In *Proceedings of the 16th International Conference on Human-Computer Interaction with Mobile Devices & Services*, pages 309–318. ACM, 2014.

[9] Adam J Aviv, Katherine Gibson, Evan Mossop, Matt Blaze, and Jonathan M Smith. Smudge attacks on smartphone touch screens. *WOOT*, 10: 1–7, 2010.

[10] Alan D Baddeley. *Human memory: Theory and practice*. Psychology Press, 1997.

[11] Chandrasekhar Bhagavatula, Blase Ur, Kevin Iacovino, Su Mon Kywe, Lorrie Faith Cranor, and Marios Savvides. Biometric authentication on iphone and android: Usability, perceptions, and influences on adoption. *USEC*, 2015.

[12] Robert Biddle, Sonia Chiasson, and Paul C Van Oorschot. Graphical passwords: Learning from the first twelve years. *ACM Computing Surveys (CSUR)*, 44 (4): 19, 2012.

[13] Jeremiah Blocki, Saranga Komanduri, Lorrie Faith Cranor, and Anupam Datta. Spaced Repetition and Mnemonics Enable Recall of Multiple Strong Passwords. *NDSS*, 2015.

[14] Cheng Bo, Lan Zhang, Xiang-Yang Li, Qiuyuan Huang, and Yu Wang. Silentsense: Silent user identification via touch and movement behavioral biometrics. In *Proceedings of the 19th Annual International Conference on Mobile Computing & Networking*, pages 187–190. ACM, 2013.

[15] Joseph Bonneau. The science of guessing: analyzing an anonymized corpus of 70 million passwords. In *2012 IEEE Symposium on Security and Privacy*, May 2012. URL http://www.jbonneau.com/doc/B12-IEEESP-analyzing_70M_anonymized_passwords.pdf.

[16] Joseph Bonneau and Stuart Schechter. Towards reliable storage of 56-bit secrets in human memory. In *Proceedings of the 23rd USENIX Security Symposium*. USENIX, August 2014. URL http://research.microsoft.com/apps/pubs/default.aspx?id=216723.

[17] Joseph Bonneau and Rubin Xu. Of contraseñas, sysmawt, and mìmă: Character encoding issues for web passwords. In *Web 2.0 Security & Privacy*, May 2012. URL http://www.jbonneau.com/doc/BX12-W2SP-passwords_character_encoding.pdf.

[18] Joseph Bonneau, Sören Preibusch, and Ross Anderson. A birthday present every eleven wallets? The security of customer-chosen banking PINs. In *FC '12: Proceedings of the the 16th International Conference on Financial Cryptography*, March 2012. URL http://www.jbonneau.com/doc/BPA12-FC-banking_pin_security.pdf.

[19] Julie Bunnell, John Podd, Ron Henderson, Renee Napier, and James Kennedy-Moffat. Cognitive, associative and conventional passwords: Recall and guessing rates. *Computers & Security*, 16 (7): 629–641, 1997.

[20] Nicholas J Cepeda, Harold Pashler, Edward Vul, John T Wixted, and Doug Rohrer. Distributed practice in verbal recall tasks: A review and quantitative synthesis. *Psychological Bulletin*, 132 (3): 354, 2006.

[21] Sonia Chiasson, Paul C van Oorschot, and Robert Biddle. Graphical password authentication using cued click points. In *Computer Security–ESORICS 2007*, pages 359–374. Springer, 2007.

[22] Sonia Chiasson, Alain Forget, Robert Biddle, and Paul C van Oorschot. Influencing users towards better passwords: persuasive cued click-points. In *Proceedings of the 22nd British HCI Group Annual Conference on People and Computers: Culture, Creativity, Interaction-Volume 1*, pages 121–130. British Computer Society, 2008.

[23] Nathan L Clarke, Steven M Furnell, Phihp M Rodwell, and Paul L. Reynolds. Acceptance of subscriber authentication methods for mobile telephony devices. *Computers & Security*, 21 (3): 220–228, 2002.

[24] Fergus IM Craik and Robert S Lockhart. Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, 11 (6): 671–684, 1972.

[25] Darren Davis, Fabian Monrose, and Michael K Reiter. On User Choice in Graphical Password Schemes. In *USENIX Security Symposium*, volume 13, pages 11–11, 2004.

[26] Alexander De Luca, Martin Denzel, and Heinrich Hussmann. Look into my eyes!: Can you guess my password? In *Proceedings of the 5th Symposium on Usable Privacy and Security*, page 7. ACM, 2009.

[27] Alexander De Luca, Alina Hang, Frederik Brudy, Christian Lindner, and Heinrich Hussmann. Touch me once and i know it's you!: implicit authentication based on touch screen patterns. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 987–996. ACM, 2012.

[28] Alexander De Luca, Emanuel Von Zezschwitz, Ngo Dieu Huong Nguyen, Max-Emanuel Maurer, Elisa Rubegni, Marcello Paolo Scipioni, and Marc Langheinrich. Back-of-device authentication on smartphones. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 2389–2398. ACM, 2013.

[29] Alexander De Luca, Marian Harbach, Emanuel von Zezschwitz, Max-Emanuel Maurer, Bernhard Ewald Slawik, Heinrich Hussmann, and Matthew Smith. Now you see me, now you don't: protecting smartphone authentication from shoulder surfers. In *Proceedings of the 32nd Annual ACM Conference on Human Factors in Computing Systems*, pages 2937–2946. ACM, 2014.

[30] Rachna Dhamija and Adrian Perrig. Deja vu-a user study: Using images for authentication. In *USENIX*

Security Symposium, volume 9, pages 4–4, 2000.

[31] Hermann Ebbinghaus. *Über das gedächtnis: untersuchungen zur experimentellen psychologie.* Duncker & Humblot, 1885.

[32] Serge Egelman, Sakshi Jain, Rebecca S Portnoff, Kerwell Liao, Sunny Consolvo, and David Wagner. Are you ready to lock? In *Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security*, pages 750–761. ACM, 2014.

[33] Mario Frank, Ralf Biedert, Eugene Ma, Ivan Martinovic, and Dawn Song. Touchalytics: On the applicability of touchscreen input as a behavioral biometric for continuous authentication. *Information Forensics and Security, IEEE Transactions on*, 8 (1): 136–148, 2013.

[34] Morrie Gasser. A random word generator for pronounceable passwords. Technical report, DTIC Document, 1975.

[35] Marian Harbach, Emanuel von Zezschwitz, Andreas Fichtner, Alexander De Luca, and Matthew Smith. It's a hard lock life: A field study of smartphone (un) locking behavior and risk perception. In *Symposium on Usable Privacy and Security (SOUPS)*, 2014.

[36] Eiji Hayashi, Oriana Riva, Karin Strauss, AJ Brush, and Stuart Schechter. Goldilocks and the two mobile devices: going beyond all-or-nothing access to a device's applications. In *Proceedings of the Eighth Symposium on Usable Privacy and Security*, page 2. ACM, 2012.

[37] Grant Ho. Tapdynamics: strengthening user authentication on mobile phones with keystroke dynamics. Technical report, Technical report, Stanford University, 2014.

[38] Jun Ho Huh, Masooda Bashir, Hyoungshick Kim, Konstantin Beznosov, and Rakesh B Bobba. On the memorability of system-generated pins: Can chunking help? 2014.

[39] Markus Jakobsson, Elaine Shi, Philippe Golle, and Richard Chow. Implicit authentication for mobile devices. In *Proceedings of the 4th USENIX Conference on Hot Topics in Security*, pages 9–9. USENIX Association, 2009.

[40] Ian Jermyn, Alain Mayer, Fabian Monrose, Michael K Reiter, Aviel D Rubin, et al. The design and analysis of graphical passwords. In *Proceedings of the 8th USENIX Security Symposium*, volume 8, pages 1–1. Washington DC, 1999.

[41] Sundararaman Jeyaraman and Umut Topkara. Have the cake and eat it too—Infusing usability into text-password based authentication systems. In *Computer Security Applications Conference, 21st Annual*. IEEE, 2005.

[42] Patrick Gage Kelley, Saranga Komanduri, Michelle L Mazurek, Richard Shay, Timothy Vidas, Lujo Bauer, Nicolas Christin, Lorrie Faith Cranor, and Julio Lopez. Guess again (and again and again): Measuring password strength by simulating password-cracking algorithms. In *2012 IEEE Symposium on Security and Privacy*, pages 523–537. IEEE, 2012.

[43] Hassan Khan and Urs Hengartner. Towards application-centric implicit authentication on smartphones. In *Proceedings of the 15th Workshop on Mobile Computing Systems and Applications*, page 10. ACM, 2014.

[44] Saranga Komanduri, Richard Shay, Patrick Gage Kelley, Michelle L Mazurek, Lujo Bauer, Nicolas Christin, Lorrie Faith Cranor, and Serge Egelman. Of passwords and people: measuring the effect of password-composition policies. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 2011.

[45] Manu Kumar, Tal Garfinkel, Dan Boneh, and Terry Winograd. Reducing shoulder-surfing by using gaze-based password entry. In *Proceedings of the 3rd Symposium on Usable Privacy and Security*, pages 13–19. ACM, 2007.

[46] Stanley A. Kurzban. Easily Remembered Passphrases: A Better Approach. *SIGSAC Rev.*, 3 (2-4): 10–21, September 1985. ISSN 0277-920X. doi:10.1145/1058406.1058408. URL http://doi.acm.org/10.1145/1058406.1058408.

[47] Taekyoung Kwon and Sarang Na. Tinylock: Affordable defense against smudge attacks on smartphone pattern lock systems. *Computers & Security*, 42: 137–150, 2014.

[48] Michael D Leonhard and VN Venkatakrishnan. A comparative study of three random password generators. In *IEEE EIT*, 2007.

[49] Zhigong Li, Weili Han, and Wenyuan Xu. A large-scale empirical analysis of chinese web passwords. In *Proc. USENIX Security*, pages 1–16, 2014.

[50] Jiayang Liu, Lin Zhong, Jehan Wickramasuriya, and Venu Vasudevan. User evaluation of lightweight user authentication with a single tri-axis accelerometer. In *Proceedings of the 11th International Conference on Human-Computer Interaction with Mobile Devices and Services*, page 15. ACM, 2009a.

[51] Jiayang Liu, Lin Zhong, Jehan Wickramasuriya, and Venu Vasudevan. uwave: Accelerometer-based personalized gesture recognition and its applications. *Pervasive and Mobile Computing*, 5 (6): 657–675, 2009b.

[52] Alexander De Luca, Alina Hang, Emanuel von Zezschwitz, and Heinrich Hussmann. I Feel Like I'm Taking Selfies All Day! Towards Understanding Biometric Authentication on Smartphones. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 2014.

[53] Ingrid Lunden. Gartner: Device Shipments Break 2.4B Units In 2014, Tablets To Overtake PC Sales In 2015, July 6 2014. URL http://techcrunch.com/2014/07/06/gartner-device-shipments-break-2-4b-units-in-2014-tablets-to-overtake-pc-sales-in-2015/.

[54] Federico Maggi, Alberto Volpatto, Simone Gasparini, Giacomo Boracchi, and Stefano Zanero. A fast eavesdropping attack against touchscreens. In *Information Assurance and Security (IAS), 2011 7th International Conference on*, pages 320–325. IEEE, 2011.

[55] Pascal C. Meunier. Sing-a-Password: Quality Random Password Generation with Mnemonics. 1998.

[56] Robert Morris and Ken Thompson. Password Security: A Case History. *Communications of the ACM*, 22 (11): 594–597, 1979. ISSN 0001-0782. doi:10.1145/359168.359172.

[57] Allan Paivio. Mental imagery in associative learning and memory. *Psychological Review*, 76 (3): 241, 1969.

[58] Shwetak N Patel, Jeffrey S Pierce, and Gregory D Abowd. A gesture-based authentication scheme for untrusted public terminals. In *Proceedings of the 17th Annual ACM Symposium on User Interface Software and Technology*, pages 157–160. ACM, 2004.

[59] Karen Renaud and Antonella De Angeli. Visual passwords: cure-all or snake-oil? *Communications of the ACM*, 52 (12): 135–140, 2009.

[60] Oriana Riva, Chuan Qin, Karin Strauss, and Dimitrios Lymberopoulos. Progressive authentication: Deciding when to authenticate on mobile phones. In *USENIX Security Symposium*, pages 301–316, 2012.

[61] Stefan Schneegass, Frank Steimle, Andreas Bulling, Florian Alt, and Albrecht Schmidt. Smudgesafe: Geometric image transformations for smudge-resistant user authentication. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pages 775–786. ACM, 2014.

[62] Richard Shay, Patrick Gage Kelley, Saranga Komanduri, Michelle L Mazurek, Blase Ur, Timothy Vidas, Lujo Bauer, Nicolas Christin, and Lorrie Faith Cranor. Correct horse battery staple: Exploring the usability of system-assigned passphrases. In *Proceedings of the Eighth Symposium on Usable Privacy and Security*, page 7. ACM, 2012.

[63] Elizabeth Ann Stobert. Memorability of Assigned Random Graphical Passwords. Master's thesis, Carleton University, 2011.

[64] Adam Stubblefield and Dan Simon. Inkblot authentication. *Microsoft Research*, 2004.

[65] Chen Sun, Yang Wang, and Jun Zheng. Dissecting pattern unlock: The effect of pattern strength meter on pattern selection. *Journal of Information Security and Applications*, 19 (4): 308–320, 2014.

[66] Xiaoyuan Suo, Ying Zhu, and G Scott Owen. Graphical passwords: A survey. In *Computer Security Applications Conference, 21st Annual*, pages 10–pp. IEEE, 2005.

[67] Hai Tao and Carlisle Adams. Pass-Go: A Proposal to Improve the Usability of Graphical Passwords. *IJ Network Security*, 7 (2): 273–292, 2008.

[68] Julie Thorpe and Paul C van Oorschot. Human-seeded attacks and exploiting hot-spots in graphical passwords. In *16th USENIX Security Symposium*, pages 103–118, 2007.

[69] Sebastian Uellenbeck, Markus Dürmuth, Christopher Wolf, and Thorsten Holz. Quantifying the security of graphical passwords: The case of android unlock patterns. In *Proceedings of the 2013 ACM SIGSAC Conference on Computer & Communications Security*, pages 161–172. ACM, 2013.

[70] Blase Ur, Patrick Gage Kelley, Saranga Komanduri, Joel Lee, Michael Maass, Michelle L Mazurek, Timothy Passaro, Richard Shay, Timothy Vidas, Lujo Bauer, et al. How does your password measure up? the effect of strength meters on password creation. In *USENIX Security Symposium*, pages 65–80, 2012.

[71] Dirk Van Bruggen, Shu Liu, Mitch Kajzer, Aaron Striegel, Charles R Crowell, and John D'Arcy. Modifying smartphone user locking behavior. In *Proceedings of the Ninth Symposium on Usable Privacy and Security*, page 10. ACM, 2013.

[72] P. C. van Oorschot and Julie Thorpe. Exploiting Predictability in Click-based Graphical Passwords. *Journal of Computer Security*, 19 (4): 669–702, 2011.

[73] Paul C van Oorschot and Julie Thorpe. On predictive models and user-drawn graphical passwords. *ACM Transactions on Information and System Security (TISSEC)*, 10 (4): 5, 2008.

[74] Christopher Varenhorst, MV Kleek, and Larry Rudolph. Passdoodles: A lightweight authentication method. *Research Science Institute*, 2004.

[75] Rafael Veras, Julie Thorpe, and Christopher Collins. Visualizing semantics in passwords: The role of dates. In *Proceedings of the Ninth International Symposium on Visualization for Cyber Security*, pages 88–95. ACM, 2012.

[76] Emanuel Von Zezschwitz, Paul Dunphy, and Alexander De Luca. Patterns in the wild: a field study of the usability of pattern and pin-based authentication on mobile devices. In *Proceedings of the 15th International Conference on Human-Computer Interaction with Mobile Devices and Services*, pages 261–270. ACM, 2013a.

[77] Emanuel Von Zezschwitz, Anton Koslow, Alexander De Luca, and Heinrich Hussmann. Making graphic-based authentication secure against smudge attacks. In *Proceedings of the 2013 international conference on Intelligent user interfaces*, pages 277–286. ACM, 2013b.

[78] Roman Weiss and Alexander De Luca. Passshapes: utilizing stroke based authentication to increase password memorability. In *Proceedings of the 5th Nordic Conference on Human-Computer Interaction*, pages 383–392. ACM, 2008.

[79] Susan Wiedenbeck, Jim Waters, Jean-Camille Birget, Alex Brodskiy, and Nasir Memon. PassPoints: Design and longitudinal evaluation of a graphical password system. *International Journal of Human-Computer Studies*, 63 (1): 102–127, 2005.

[80] Helen M Wood. *The use of passwords for controlled access to computer resources*, volume 500. US Department of Commerce, National Bureau of Standards, 1977.

[81] Jeff Jianxin Yan, Alan F Blackwell, Ross J Anderson, and Alasdair Grant. Password Memorability and Security: Empirical Results. *IEEE Security & Privacy*, 2 (5): 25–31, 2004.

[82] Yang Zhang, Peng Xia, Junzhou Luo, Zhen Ling, Benyuan Liu, and Xinwen Fu. Fingerprint attack against touch-enabled devices. In *Proceedings of the Second ACM Workshop on Security and Privacy in Smartphones and Mobile Devices*, pages 57–68. ACM, 2012.

[83] Ziming Zhao, Gail-Joon Ahn, Jeong-Jin Seo, and Hongxin Hu. On the security of picture gesture authentication. In *USENIX Security*, pages 383–398, 2013.

[84] Nan Zheng, Kun Bai, Hai Huang, and Haining Wang. You are how you touch: User verification on smartphones via tapping behaviors. In *Network Protocols (ICNP), 2014 IEEE 22nd International Conference on*, pages 221–232. IEEE, 2014.

[85] Moshe Zviran and William James Haga. Passwords Security: An Exploratory Study. Technical report, Naval Postgraduate School, 1990.

10. Post-experiment survey

The remainder of this submission contains the survey we presented to participants in place of the 50th attention game, at the end of the main study (but before the follow-up).

Page 1

Congratulations!

You have completed all of your required attention tests. (We're not going to ask you to do the 50th.)

All you need to do now is complete this final survey.

Page 2

Is English your native language?

- Yes (811, 98%)
- No (14 2%)
- I don't understand the question (0, 0%)
- Decline to answer (0, 0%)

What is your gender?

- Female (329, 40%)
- Male (494, 60%)
- Decline to answer (2, 0%)

What is your age?

What is your current occupation?

Page 2

What is the highest level of education you have completed?

- Did not complete high school; High school/GED (7, 1%)
- High school/GED (83, 10%)
- Some college High school/GED (232, 28%)
- Associate's degree; High school/GED (94, 11%)
- Bachelor's degree (319, 39%)
- Master's degree (60, 7%)
- Doctorate degree (4, 0%)
- Law degree (8, 1%)
- Medical degree (7, 1%)
- Trade or other technical school degree (10, 1%)
- Decline to answer (1, 0%)

Page 3

The following question(s) are about how you logged into the attention study using your username and PIN.

During the study, did you enter your PIN using a mouse, touch screen, or some other pointing device? (If you used more than one input method, choose the one you used the most.)

- Mouse (729, 93%)
- Touch Screen (27, 3%)
- Other device (26, 3%)

[*If not in Assigned*]
Was the PIN you chose one you have used before, such as to protect a locker, debit card, credit card, or website?

- Yes (280, 50%)
- No (285, 50%)

[*Unless participant in Second-PIN*]
Did you store your PIN for the study website, such as by writing it down, emailing it to yourself, or adding it to a password manager?
[*If participant in Second-PIN*]
During the course of the study we assigned you a second numeric PIN to enter. Did you store that PIN, such as by writing it down, emailing it to yourself, or adding it to a password manager?

- Yes (148, 18%)
- No (677, 82%)

[*If anwered 'Yes' above*]
Please explain how and where you stored your PIN.

[*If in a mapping treatment*]
During the course of the study, as the delay before the digits of your PIN appeared grew longer, it became faster to sign in by pressing the keys before the digits appeared. In order to do so, did you store that pattern or sequence of letters, such as by writing it down, emailing it to yourself, or adding it to a password manager?

- Yes (44, 9%)
- No (438, 91%)

[*If answered 'Yes' above*]
Please explain where you stored this information, and whether you stored it as a pattern or as a sequence of letters.

[*If in a mapping treatment*]
If you learned how to enter your PIN on the keypad without waiting for digits to appear on the keys, how did you remember which keys to press?

- I remembered the position of each key on which the correct digit would eventually appear, which formed a pattern. (272, 56%)
- I remembered the letter of each key on which the correct digit would eventually appear, which formed a sequence of letters. (95, 20%)
- I remembered both.; (95, 20%)
- I never learned to enter my PIN without waiting for the digits to appear. (20, 4%)

Page 4

As we explained at the start of the study, we are experimenting with a new login system using a PIN.

[*If in a mapping treatment*]
Security researchers have found that computer users often choose predictable PINs, such as those that represent important dates or easy-to-enter patterns. One reason users choose predictable PINs is that it is hard to memorize less-predictable codes without practice.

[*If in a mapping treatment*]
With the PIN system used in this study, you practiced learning a more-secure a random code (the sequence of positions/letters on the keypad) each time you entered your numeric PIN (the digits that you chose when you signed up for the study). Once you learned the positions or letters, the sign-in system could discard the digits of your PIN and never show them again, leaving you with the secure random code that you had memorized through repetition.

[*If in Assigned* or *Second-PIN*]
Security researchers have found that computer users often choose predictable PINs, such as those that represent important dates or easy-to-enter patterns. In this study we assigned you a more secure randomly-generated PIN.

[*If in Second-PIN*]
For the purpose of the following question, assume that you had the option to remove the PIN that you had initial chosen once you had learned the more secure randomly-generated PIN that we assigned you. This option would allow you to login more quickly, using only four digits, instead of eight.

[*If not Second-PIN*] If you wanted to keep your phone or tablet secure, would you want to use a PIN like the kind you used to sign into our experiment's website? (If you also use a fingerprint reader, this would be the code you use when your device needs a stronger proof of your identity.)
[*If in Second-PIN*] If you wanted to keep your phone or tablet secure, would you want to use a randomly-generated second PIN like the kind you learned when signing into our experiment's website? (If you also use a fingerprint reader, this would be the code you use when your device needs a stronger proof of your identity.)

- Yes (478, 58%)
- Maybe (271, 33%)
- No (76, 9%)

Please explain your preference.

Page 5

Last question!

If you encountered any problems during the study, or any bugs in our study website, please let us know about them.

Page 6

You have now completed the entire study. Thank you so much for your time and attention. We will process payment within the next two business days. If your payment does not arrive within that time, please contact us at msrstudy@microsoft.com. (If you forget that address, you can also find it at the bottom of all the web pages on this site.)

You may close this tab at any time.

User ID	Treatment	Explanation
2255	*Mapping*	I simply remembered the pattern.
2400	*Mapping*	I just remembered it
2407	*Mapping*	I just remembered the patter [SIC] based on what my pin was.
2819	*Instructionless*	I just remembered the sequence
2836	*Arrowless*	I memorized the sequence of letters through rote memorization.
2849	4x20 *Mapping*	I just remembered where my numbers appeared

Table 10. Participants who answered *yes* when asked if they wrote down or stored their PINs, but then explained that they had only stored it in their memory.

A. Corrections to multiple-choice responses

We followed many of our multiple-choice questions with follow-up questions that asked participants to explain their multiple-choice response in written form. We used these responses to determine how well participants understood our questions and identify situations in which participants clearly misunderstood a question when answering it.

We discovered that, when we asked participants if they had written their pattern, some reported *yes* but then provided answers that clearly and unambiguously indicated otherwise. We expect this is because participants didn't realize our goal was to teach them the pattern, and interpreted that the question asked them to report storing the key in their own memory and using their memory to enter the PIN before the digits appeared. In presenting our results and performing our analyses, we disregarded a response of *yes* from participants in Table 10, substituting a *no* to reflect their explanation.

We audited responses to questions asking about whether participants in the *Second-PIN* treatment had stored their assigned (second) PIN, and did not find any evidence to suggest that any of those participants had misunderstood the question.

B. Evidence of a priori hypotheses

At 12:06PM eastern time on February 23, we sent the SOUPS program chairs the following base64 encoded SHA256 hash:
xVHlPGs/WkKQZvmAHxhWrwpjy/WCH9oB1GMupwzLx+E=

That hash was generated from the string below. The presence of "\r\n" indicates a carriage return and line break. Any white space, including line breaks produced by the formatting of this document (those not following "\r\n", indicates the presence a single ASCII space character. The numbering of hypotheses 3 and 4 were switched to facilitate exposition.

Hypothesis 1a/1b\r\n
Participants in PATTERN will be less likely to drop out of than those in (a) ASSIGNED and (b) SECOND_PIN\r\n
Statistic: Of participants who reached the sign-in page, the proportion who finish the study\r\n
Test: Fisher's Exact test\r\n
\r\n
Hypothesis 2a/2b\r\n

Participants in PATTERN will be less likely to write down their secret than those in (a) ASSIGNED and (b) SECOND_PIN\r\n

Statistic: Of participants who finished the survey, the proportion who reported writing their PIN or pattern\r\n

Test: Fisher's Exact test\r\n

\r\n

Hypothesis 3a/3b\r\n

Participants in PATTERN will report being more willing to use this authentication system than those in (a) ASSIGNED and (b) SECOND_PIN\r\n

Statistic: Of participants who completed the survey, the their answer to a question about whether they would want to use it was \r\n

scored 'no'=0, 'maybe'=1, 'yes'=2\r\n

Test: Mann Whitney U a.k.a. Wilcoxon\r\n

\r\n

Hypothesis 4: Participants in PATTERN will spend less time learning their secret than those in SECOND_PIN\r\n

Statistic: Of participants who completed the study, the aggregate pin-entry time in seconds from appearance of the PIN to completion, with no login taking more than 60s) from the first login until (but not including) the first session in which the participant entered the code correctly (but no more than the first 49 logins).\r\n

Test: Mann Whitney U a.k.a. Wilcoxon\r\n

Security Practices for Households Bank Customers in the Kingdom of Saudi Arabia

Deena Alghamdi
University of Oxford
Department of Computer Science
Parks Road, OX1 3QD,
United Kingdom
deena.alghamdi@cs.ox.ac.uk

Ivan Flechais
University of Oxford
Department of Computer Science
Parks Road, OX1 3QD,
United Kingdom
ivan.flechais@cs.ox.ac.uk

Marina Jirotka
University of Oxford
Department of Computer Science
Parks Road, OX1 3QD,
United Kingdom
marina.jirotka@cs.ox.ac.uk

ABSTRACT

Banking security is an instance of a socio-technical system, where technology and customers' practices need to work in harmony for the overall system to achieve its intended aims. While the technology of banking security is of interest, our study focuses on exploring the specific practices of household bank customers in the Kingdom of Saudi Arabia (KSA). The findings describe some practices of household customers and reveal some of the reasons behind them. Contrary to banking policy, sharing bank authentication credentials appears to be a common practice for our participants, and a number of different reasons are presented: trust, driving restrictions, the esteem placed in parents, and the 'need to know' this information. On the other hand, some participants consider credentials to be private information and do not share, although other participants view this as a sign of distrust. Implications of such practices on the Saudi banking system are outlined and discussed.

1. INTRODUCTION

A banking system is a socio-technical system which operates on a technical base and is used by people every day [59]. To achieve the aims of a secure and effective banking system, it is necessary for every aspect of a socio-technical system to work in harmony towards these goals, including hardware (PCs, hubs, routers), software (applications, programmes, codes), users and their practices, system procedures and regulations, data flow and structure. Crucially, users' security practices constitute "a vital variable" affecting the effectiveness of a system's security, according to Weirich and Sasse [58]. They found, for example, that colleagues in a workplace will commonly disclose passwords to one another as a sign of trust, to the extent that not participating in this practice can be seen as hiding something, even though this practice offers a clear opportunity for hackers and industrial spies to use social engineering techniques.

Different studies have considered the user as an employee in an organizational context [25] [37] [27] [42] [58]. For instance, Inglesant and Sasse [27] carried out a study among the staff of two organizations to explore the use of passwords in the workplace and found that it affected their productivity and ultimately that of the organisation.

However, the study of users' practices in the household context is a relatively underexplored area. The household context, compared to the organizational one, is interesting for a number of reasons: first, users spend longer hours and conduct a broader variety of different activities at home than they usually do in the workplace. Second, the organizational context typically contains only employees, while the household context may include wider groups of employed and unemployed users such as housewives, househusbands, students, pensioners and others. Third, households interact with a variety of different services – each offering different forms of security components, configuration options, and advice – however the responsibility for security in most households is not held by an experienced security practitioner. While organisations face similar challenges, they can devote greater resources to staffing, formulating security policies, and educating their employees in matters of security.

Two studies of computer users in a household context have explored the impact of users' family relationships on their interactions with computer applications. The VOME project [19], which studied users' privacy and how they used ICT within family settings across the UK. Coles-Kemp and Ashenden held a number of family workshops which recruited pairs of granddaughters and their grandmothers, some of whom were great-grandmothers. Meanwhile, Singh [48] [49] and Singh et al. [52] [50] [51] investigated the financial and banking practices of Australian households by interviewing couples, or individual members of couples in some cases. Despite providing valuable insights into the role of familial relationships in computer users' practices, more detail is needed to understand the different cultural and familial factors that affect security interactions in the home.

Our study focuses on the security practices of household bank customers in the Kingdom of Saudi Arabia (KSA), where modern banking technology has been adopted widely, but where very little work has as yet been done to explore the security practices of household customers. The KSA poses interesting cultural, technical, and security characteristics that have been under explored to date. Kaspersky [29] report that in 2009,

Symposium on Usable Privacy and Security (SOUPS) 2015, July 22-24, 2015, Ottawa, Canada.

Saudi networks suffered the seventh highest incidence of information security attacks in the world. With only 0.007% of the Internet users in the world, Saudis were subjected to 1.81% of such attacks – a significantly higher rate than average. Symantec Intelligence reports rated the KSA as the most spammed country in 2012 and 2013, with a spam rate of 79.0% and 82.7% respectively [53, 54]. And in 2013, the KSA had the highest risk (10.78%) of privacy exposure among its Android users and the third highest volume (7.19%) of malicious Android app downloads [57]. Alarifi et al [5] suggest that this is due to the low levels of information security awareness among computer users in the KSA and according to Alkaabi et al [8] one of the information security awareness issues that needs to be dealt with is credential sharing with co-workers and family members. Algarni [7] observed that Saudi governmental organisations appeared unready to confront and control information security problems, citing the influence of religion and culture, which prevented government representatives from taking certain security approaches such as online monitoring.

Our paper offers two main contributions: i) a brief methodological review and research method design that respects Saudi cultural norms in order to allow us to ii) elicit and explore sensitive security practices and perceptions of household bank customers in KSA.

2. BACKGROUND

To support our analysis of household bank customers' security practices in the KSA, this section provides a brief overview of the history of the KSA, and an introduction to its banking system.

2.1 The Kingdom of Saudi Arabia

The KSA was founded in September 1932 by King Abdul-Aziz Al-Saud. It is one of the largest countries in the Middle East, with an area of over 2,000,000 square kilometres (772,204 square miles). Saudi cultural roots are a mixture of values originating from traditions and Islamic principles where social and religious practices and traditions are thoroughly intertwined. In the KSA, 97% of the population are Muslims [38] and religion plays an important role in Saudi life and politics, as Islam is considered a defining element of Saudi identity [36]. The system of tribes also has a great influence on Saudi life and practices. Before 1932 these tribes were hostile to each other, a situation which changed dramatically when King Abdul-Aziz united them into one country [13] [14]. Another major change came as a result of the oil boom of the 1970s, when the urban population of the KSA jumped from 16% of the total population in 1950 to 49% in 1970 and 80% in 1990 [31], while the nomadic tribal population fell from being the majority early in the twentieth century to between 5% and 25% of the population by 1993 [18]. Tribes still exist today in KSA but with less impact than before.

Women in KSA: Islamic principles stated in the Qur'an and the Prophet's Hadith[1] grants women and men equal but not identical rights in personal, civil, social and political life aspects. None of these principles prevent women from participating in public life but the Qur'an does warn against the mixing of the sexes which could lead to "seduction and the 'evil consequences' that might follow" [11]. In KSA, jurists have interpreted this warning by tightly restricting any type of interaction between a woman and a non-mahram[2] [22], and gender segregation is fundamental to Saudi society [20, 35]. Culturally, gender separation is the rule in almost all aspects of private and public life in the KSA [4] [16]. In higher education for example females study in separate campuses and are taught face-to-face by women, or by men via closed-circuit television (CCTV) [10] which provide one-way video where female students can see their male teacher and two-way audio, from teacher to students and vice versa [9]. According to Deif [21], Saudi Islamic scholars maintain this norm with the social-religious justification that "loose interaction across gender lines is one of the major causes of fornication, which disintegrates society and destroys its moral values and all sense of propriety" (p. 13). Another aspect of this interpretation is that women in the KSA are forbidden from driving and those who defy this ban risk jail; therefore, women must use vehicles driven by a chauffeur or a mahram [60]. Moreover, the idea of walking even for a short distance is culturally quite restricted [4]. Saudi cities rarely have reliable public transport networks and where there are buses, women must use a separate entrance at the back and occupy designated seating [28] . A woman is expected to be accompanied with a mahram when she "ventures into the public sphere" [32].

Households in KSA: A household in the KSA usually consists of a married couple and their unmarried children of either sex; it is against the rules of Islam and culture for an unmarried couple to live together or to have children. Each household is headed by a male, while the role and position of each individual in this type of structure is defined according to sex and age [2]. Some features of households in the KSA, such as size and economic characteristics, have changed over time. The average household size fell from 6.08 persons per house in 1992 to 5.84 in 2010 [6]. In general, the economic level of households in the KSA has improved, especially since the oil boom. Today, the KSA is the top country in the world in terms of density of ultra-wealthy households (18 per 100,000), due to the strong national economy [55].

2.2 Banks in KSA

The history of banks and bank systems in the KSA started in 1926 with a few foreign-based trading houses (including Netherlands Trading Society) and money changers, which provided most of the finance services whose main role was to serve pilgrims and the trading community [47]. After that, foreign banks started gradually entering the market. The Saudi Arabian Monetary Agency (SAMA) was created in October 1952 by the government with the main purpose of achieving a stable monetary mechanism and ensuring currency stability. In 1966 a new Banking Control Law was approved, giving SAMA more supervisory powers [3]. With the approval of the Minister of Finance, the Banking Control Law also permitted SAMA to

[1] The Hadith is a collection of traditions containing sayings of the prophet Muhammad which, with accounts of his daily practice (the Sunna), constitute the major source of guidance for Muslims apart from the Qur'an – Oxford English Dictionary.

[2] A mahram is a woman's husband or a man to whom she cannot be married, either because of a blood relationship, such as her father, brother, grandfather, son, uncle or nephew, or because of a marriage relationship, such as her father-in-law, son-in-law, stepfather or stepson.

recommend institutions for new licenses, issue rules and regulations, and take action against any violators of the Law and therefore banks expanded rapidly covering the entire country [43]. Since the 1980s SAMA continued to introduce new tools and systems to improve and strengthen the Saudi financial markets and to compel all Saudi banks to invest in technology and to improve their back and front office operations. Significant changes were made to modernise the system, starting with the introducing of a national Automated Teller Machine (ATM) system which permitted customers access to their accounts from any machine, followed by the introducing of debit, credit and charge cards and the linking of Saudi banks with the SWIFT payment network, banks also shared the benefits of a point of sales system and an advanced electronic share trading and settlement system, which enabled same-day settlement [43]. Today, bank customers can reach their account and use bank services via different channels: bank branches, online banking, phone banking, ATMs, sales point and online shopping. The security mechanisms can differ for each channel and from bank to bank: for instance some banks use hardware tokens to produce one-time PIN for online authentication while other banks send it to the mobile phone of the account holder. One security mechanism used very widely is an SMS notification where the account holder will receive an SMS for every transaction carried out in their account, e.g. money transfer, deposit, withdrawal, and online access.

SAMA regulations: SAMA has published its regulations for Electronic Banking (clause 3-3 Bank Obligations). One notable extract is that "banks are responsible for providing secure and safe systems and services for their customers unless the customer fails to safeguard their account user number or password and divulges it to a third party" [45]. Interestingly, this appears not to be the case for credit and debit cards, where a client may choose to authorize another to use their card. SAMA's Regulations for the Issuance and Operation of Credit and Charge Cards (clause 7-1) state: "The client (i.e. the cardholder) is not liable for any non-authorized transactions made with his card after it has been reported lost or stolen if the following conditions have been satisfied" and the third condition is: "The client must exercise every care and vigilance in safeguarding his card from loss or theft or unauthorized use." The document defines unauthorized use as "The use of a credit card or debit card by anyone other than the client (card holder) who does not have actual or implied authorization"[44]. SAMA regulations have emphasized more than once that banks must use different channels such as bank websites, promotional publications and others to educate their customers about different regulations for using bank services, protect online credentials (username and password), protect their card and its PIN, using secure passwords and changing them regularly, complaint procedures and others ([45] p. 11,[46] p 6-23). Specifically, the Consumer Protection Principles [46] (clause 11-5) state that the bank should inform the customer officially and clearly about his responsibilities when opening a bank account and the consequences of sharing the account credentials with a third party.

Saudi banks and women: The Saudi banking sector now consists of 12 banks supervised by SAMA. Under the religious and cultural constraints mentioned above, these all have so-called ladies-only branches, access to which is limited to female employees and customers. Before these were established, women had access to bank branches and bank services such as opening accounts and transferring money, but as all of the employees were males this was inconvenient and violated female customers' privacy, because they would normally have to be accompanied by a mahram. The opening of ladies-only branches was thus an important step in the provision of services for women in the KSA. This development was introduced gradually, with the Al Rajhi bank opening its first ladies' branch in 1979 [1], while the Bank of Holland did not do so until 1999 [47]. The participation of women in the Saudi banking sector has continued to grow: women now account for 18% of the banking workforce and women have recently been appointed to more senior positions in several banks [56].

3. METHODOLOGY

The context in which this study is attempting to elicit information about users' practices was methodologically challenging from the outset. Previous work in this area has aimed to elicit security practices from home users as for example Kaye [30] who reports on a Facebook and Twitter survey about password sharing practices. Whilst providing interesting insights into self-reported password sharing, this approach was reported to have experienced problems in recruiting participants due to the sensitive nature of the topic. For our study, given that we were not sure of the issues and topics of interest in the domain, and that the security of banking is indeed a sensitive area, we did not feel a survey was likely to suit our particular research problem. Mathiasen et al. [34] discuss the use of exploratory workshops to discuss in depth a number of concepts and findings.. However this approach requires a certain amount of foreknowledge about the problem domain and we did not feel we had enough information to adopt this method.

The purpose of this research is to elicit and investigate the security practices and perceptions of household bank customers in the Kingdom of Saudi Arabia and consequently, the methodology has to accommodate several features. First, this is a new area of research for which we have little pre-existing knowledge. Second, this investigation aims to gather empirical, qualitative data. Third, the data concerns both the practices and perceptions of these customers.

Based on these requirements, we chose Grounded Theory which is a methodology for *generating* a theory, instead of testing one [26], and a means of understanding a problem that has not yet been explored deeply. Grounded theory originates from the work of Glaser and Strauss [26] and requires the researcher to undertake the iterative process of data collection, coding (assigning meaning to elements of the data), and thematic analysis (relating codes to one another) to provide an overall theory explaining the qualitative data.

Our study was ethically reviewed and approved by the Social Sciences and Humanities Inter-divisional Research Ethics Committee at the University of Oxford. There were two rounds of data collection: an initial telephone interview with six household bank customers (preliminary findings presented in [24]), and a series of four focus groups covering husbands, wives, sons and daughters in the KSA. The first round helped us identify practices and concepts related to the research topic which we wanted to explore in more detail. This informed the direction we took the discussion between the participants in the second round. Each round had to be carefully designed to take into account: the sensitivity of discussing bank credentials, the intimate nature of the household context, and the cultural and religious constraints in the KSA. We present an overview of the research methods used in each round in the following sections.

3.1 First Round of Data Collection

Telephone interviews were conducted with six household bank customers. Telephone interviews were preferred as offering an appropriate method to elicit and discuss customers' practices whilst being sensitive to the practices of Saudi culture. This method was more suitable than face-to-face interviews for both the researcher and bank customers, in particular for male customers as the researcher is a female. First, the gender segregation practiced in the KSA (mentioned previously) extends to the provision of separate areas for males and females. Conducting interviews at home with a stranger is unusual in the Saudi culture even if both participant and researcher are of the same gender. The interviews were semi-structured where the interviewer directs the interviewee to the research topic but does not guide the answers to match their own expectations. All six bank customers were married Saudi citizens, three males and three females. Some had been married for only two years and some for more than ten. Two were married to each other but were interviewed separately. The participants were from various cities across the KSA and all were employed except one, who was the partner of an employee. Each interview lasted 35 to 50 minutes, carried out and transcribed in Arabic. The transcripts were translated into English then analysed using the ATLAS.ti computer-aided qualitative data analysis software (CAQDAS) package.

3.2 Second Round of Data Collection

After the first round of telephone interviews, focus groups were used to explore in more detail the participants' practices and attitudes that emerged from the telephone interviews. Focus group sessions are gatherings, usually of 6-8 participants [17], which encourage interaction amongst the participants to develop opinions and thoughts, as in real-world situations [33]. Four focus groups were conducted with household bank customers: husbands, wives, sons, and daughters. These sessions were segregated by gender, with a female researcher running sessions for groups of female household customers (wives and daughters), and a male moderator running groups of male household customers (husbands and sons). This arrangement was necessary for two reasons: firstly, to conform to the gender segregation that is fundamental to Saudi society; secondly, to explore the impact of different roles within households as factors influencing participants' practices. The daughters' focus group had six members, whose main selection criteria were that each must be unmarried and still living in her parents' house, to reflect the daughter's role. The sons' group comprised five participants, none of who were married and all of whom were still living with their parents, reflecting the son's role. There were six participants in the wives' focus group, four of whom were employed and two were housewives. The main selection criterion was for the participating women to be married, and some participants had children while others did not, to reflect the opinions of wives in both situations. Finally, the husbands' group also had six members, five of them employed and one who had retired. As with the wives, all were married to reflect the husband's role and not all had children.

We encountered significant challenges in running focus groups in the KSA, the first being one of recruitment. In order to recruit sufficient participants it was necessary to invite up to 40 people to join each group via social and professional networks. This difficulty arose mainly because focus groups are not common in the KSA compared to other methods such as questionnaires. The principle and operation of the focus groups therefore had to be described carefully to potential participants in the invitation letter. It was noticed that it was easier to recruit younger participants, for the daughters' and sons' groups, than older ones, for the wives' and husbands' groups. It happened many times that a recruit thought that she might convince one or more of her parents to participate in the study but failed.

Another challenge was finding a place to conduct the sessions. This might not be a problem in other countries, but the researcher spent a great amount of time and effort on this issue. Given the cultural norms arising from the religious and cultural constraints referred to above, it was necessary to find public places to conduct the sessions, as the use of private premises, such as the researcher's home, would not have been acceptable. This was particularly difficult because there are no neighbourhood community centres in KSA, and hiring a meeting room in a hotel would be both expensive and require approval from the City Council which can take up to three months to arrange. We finally resolved this through a closed meeting room in a restaurant, although such facilities are not common in KSA. The room was suitable because it was affordable, accessible to the public but could be closed for the sessions, quiet enough for clear recordings to be made and,no official approval was required. Participants were contacted via WhatsApp (a cross-platform mobile messaging app) to schedule and reschedule the sessions. Each session lasted an hour and a half, and as in the first round, each session was carried out and transcribed in Arabic. The transcripts were then translated into English and analysed using the ATLAS.ti.

4. SHARING CREDENTIALS

Our analysis of security practices highlighted that sharing of authentication credentials for household bank customers in KSA is an important topic and follows two main themes (see Figure 1): **has shared**, where an account holder shares their credentials with family members, and **has not shared**, where the account holder prefers to keep this information secret and not share it with anyone.

The authentication information that was shared most between family members was the card PIN. This was the case for 25 of the participants while only four participants did not share such information – one female participant in the first round and one participant from each group of husbands, daughters and sons.

Only fourteen participants used online banking and five of them shared their online credentials – username/password: three participants in the first round of the data collection and two participants from the sons focus group shared with their brothers only. The two main categories, **has shared** and **has not shared**, can be divided to nine categories (see Figure 1): *Female account holders*, *Male account holders*, *Parents*, *Father's successor*, *Couples*, *Safe*, *Need to know*, *Having two accounts,* and *temporary divorce*. We explore these in more detail below:

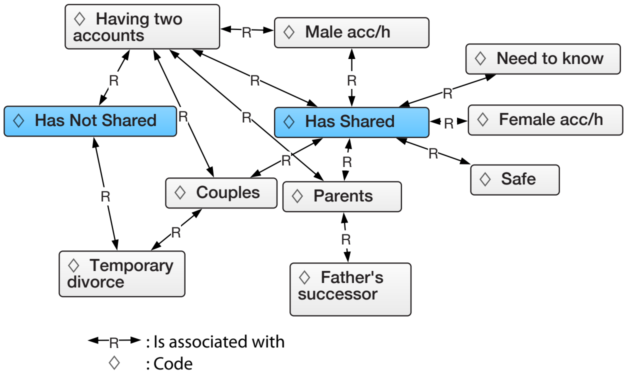


Figure 1: Sharing practices for household bank customers.

4.1 Has Shared

The following are a detailed discussion of the factors affecting credential sharing practices between family members based on familial relationships:

4.1.1 Female account holders

Most of the female participants appeared to see sharing as a *blessing*, because it solves a major problem: withdrawing cash from their account. If a woman needs to withdraw money from her account then she either goes to a bank branch or uses an ATM. In both cases, however, and given the movement limitation mentioned above, she must use vehicles driven by a chauffeur or a mahram. If the vehicle is driven by a chauffeur then she can withdraw money by herself from either a branch or ATM, but if the vehicle is driven by her mahram then she may find that the ATM is on the driver's side; she will then feel obliged to hand the driver her card so that he can operate the ATM, because asking him to pull over so that she could walk to the ATM and withdraw the money in private would be inappropriate and unacceptable. A member of the wives' focus group said*:*

Sometimes I am with my husband in the car and he withdraws the cash for me because the machine is on his side of the car. [P6, F]

An alternative scenario is when she needs to ask her mahram to take her card and bring her the cash when she cannot (e.g. because it is too late) or does not want to accompany him. A participant in the daughters' focus group said:

A girl shares her PIN with her brother to ask him to get cash for her from her account when she can't get out or she doesn't want to, but the brother doesn't need to share such information because he goes out by himself. [P3, F]

In addition, *'laziness'* or *'making life comfortable'* were stated reasons for sharing among females when the female account holder might prefer to stay at home and ask a female member of her family to draw some cash on her behalf. A participating wife stated:

When I am lazy and I don't want to go to make a withdrawal. If for example I am at my family's house and my little sister is going out, I ask her to get me cash. [P12, F]

In this scenario, the wife is taking advantage of the fact that her little sister is already going out accompanied by a chauffeur or a mahram in order to avoid the effort of having to arrange this for herself.

Female participants saw trust as the main reason for sharing; only if trust existed would any other factors be considered, as explained by this focus group member:

If you don't trust your family, would you give them your card? Even if you are lazy or there is a transport problem, you'd say 'No, I'll do it myself, even if I am tired'. If you didn't have this feeling of trust you would overcome your laziness and even the transport problem, you'd force yourself one way or another to go. I think that the first reason is that amount of trust between us, an excellent trust that makes me comfortable, and if I don't want to go it is easy to ask my brother to do it as long as there is no transport. [P15, F]

It is worth mentioning here that if movement limitation did not exist, female participants stated that they would still share for other reasons such as to make life more comfortable, albeit not as much as they did now.

The lifestyle in Saudi Arabia generally forces me to share my credentials with my husband or one of my brothers … I can't drive here and there is no public transport … and even if this situation changed, there would be much less need for sharing, nevertheless I would still do it sometimes, to simplify my life. [P1, F]

4.1.2 Male account holders

"Laziness", or *"to make life more comfortable and easier"* were also reasons given by some male participants for sharing their credentials with others, whether female or male. A member of the husbands' focus group said:

Sharing is a way of supporting each other … a kind of solidarity where we help each other to make our lives easier. [P17, M]

However, it has been noticed that some male account holders **have two accounts** (see section 4.2.1). The man shares credentials for only one of them with his family, allowing them to use it whenever they need to, but his family have no idea about his second account. This practice will be discussed in more details in Section 4.2.1.

4.1.3 Parents

As mentioned above, KSA is a religious country and Islam is the dominant religion. The rules and principles of Islam emphasise the high value of parents which is reflected in Saudi culture as one of the factors influencing participants' practices.

For example, when the participants in focus groups asked about sharing bank credentials with fathers, all of the participants replied that it was the father's right to take his child's money, regardless of gender, and that he or she most likely could not be refused such a request. Thus, a member of the wives' focus group said:

It is hard to refuse such a request even if I want to … I can't say no or I will lose my father … he will say 'I spent all my life supporting you and bringing you up and now you don't want me to have an authorization on your account'. … Even if he takes *all the money ... you and your wealth belong to your father. [P11, F].*

The participant here was quoting a Hadith which states: “You and your wealth belong to your father.” This Hadith was mentioned by many participants, in all four focus groups.

Both daughters and sons are included in this Hadith, but culturally the focus is more on daughters when one of the participants in daughters’ focus group said “*A daughter is always under her father’s wing*”. This describes the relationship between a father and his daughter, asserting that she must always be under his supervision. When a father asks his daughter to share her bank credentials with him or to give him control of her bank account, all the focus group participants (with one notable exception in the daughter’s group – see below for more details) declared that they could not refuse such a request as it would be a sign of disrespect and ingratitude.

Most of the participants felt that any such request would be made in the best interests of the daughter. For example, two wives and two daughters declared that a father could ask to supervise his “extravagant” daughter and monitor her expenditure as a way of “teaching her to not spend all her money”. While other participants said that this arrangement can be a “safeguard” to protect the daughter from others, such as a “greedy husband”. However, when one member of the daughters’ focus group declared that she would refuse such a request because she considered her account and its information to be private, all of the other members in the group disapproved of this declaration despite her asserting many times that her father was an open-minded man who would accept this refusal from his daughter and respect her privacy.

However, participants also identified some cases where a daughter could refuse her father’s request for sharing or delegation and such a refusal would be acceptable, such as where the father had a reputation for using money unwisely or treating his wife badly. A husband said:

She can say no, but will she say it? She has to express the reason why she doesn’t want him to have control over her account. Are the reasons expressible? If her father is an exploitative man, unnatural or maybe a tyrant, or a man who has previous history of exploiting family members’ individual money, then it is acceptable to say no …. In such a case she might be afraid that he will use that delegation badly by getting a loan, or withdrawing money from the account without her permission. [P24, M]

In refusing a father’s request to share details,, boys were seen as different to girls in that a son could refuse such a request and, in this case, the father would accept the refusal and consider it a sign of maturity in his son. This difference seems more cultural than religious, and there was some disagreement among participants. Some said that even boys would not refuse sharing or delegating authority to their father as a sign of respect. Others recognised that there was a difference but believed it to be based on other factors. A member of the husbands’ focus group said:

There is a difference, but it’s based on caring and sympathy, that a father cares more for his daughter than for his son. This means that she might not be able to go to the governmental departments but the boy can come and go, so the father can be delegated control over his daughter’s account to see to her needs. The father is always kind to his children, but the boy can look after himself, as he will carry more responsibility than the girl in life. [P22, M]

Regarding sharing credentials with mothers or delegating authority to them, then an instance of this was raised as a security measure to protect the daughter and her money. A participant in the daughter focus group mentioned that her friend had delegated control of her account to her widowed mother in order to protect her (paternal) inheritance from her husband. The husband had repeatedly tried to convince his wife to give him part of her inheritance so that he could use it to expand his business. To avoid direct conflict with him and to protect her money, the wife passed control of the account to her mother. The husband now recognized that his mother-in-law was supervising her daughter’s account and monitoring every transaction which stopped him from attempting to convince his wife.

Participants were then asked whether a male account holder would be likely to share his account information equally with his wife and his mother. Some replied that a man would trust his mother more than his wife and that this would be reflected in his sharing all of his information with his mother but not necessarily with his wife, because of the high value of mothers in Islam. One said:

The mother has a higher status ... A man would share his credentials more willingly with his mother than his wife. [P10, F]

Conversely, some participants said that while there was no difference between a mother and a wife in the sense that a man would trust them both equally, his wife would be more likely to know his card PIN than his mother would, because the mother would not know how to use an ATM, or because she would not need to know this information as she lived in another house. A member of the husbands’ focus group explained:

There is a difference… I think a husband is more likely to share such information with his wife than his mother… not by preference, but because she is healthier so she can come and go, she can serve me and serve my mother. I can tell her to withdraw money from my account and give it to my mother so I won’t tell my mother to go herself… The trust is there for both of them, but the sharing depends on other reasons. [P29, M]

In addition, when a parent shared his/her bank credentials with his children, the main reason was to help the parent and to make life easier. One mother also admitted that she gave her card to her son not just to help her, but also as a way to control her expenditure.

4.1.4 Father’s successor

In Saudi families, it is common for a son to be raised with the idea that he is his father’s successor and that when the time comes he will be the one who manages the family’s affairs. This son is normally the eldest son, but this is not always the case. In some families he may be the most trusted son, or the closest to his siblings, instead of the eldest. One of the participants explained this in these terms:

It is common in Saudi Arabia… but it isn’t always the eldest brother… It can be any of the brothers. I see it as a characteristic of Saudi society that a family always embraces all its members… and the father is the base. If he is lost, they will cling to any straw, meaning the family will find someone who is dependable and capable of looking after everybody’s interests…

They count on him, so it is assumed to be the eldest son who is raised on that idea and is seen as his father's successor, but it isn't mandatory. Any brother might play this role whether he is old or young…. The eldest might be sick, for example. [P7, F]

As noted above, most participants felt that a woman could not refuse her father in taking control of her bank account, but the situation was seen as quite different in the case of a brother acting as their father's successor, when a sister could refuse her brother's request for delegation. All the participants in the husbands and wives focus groups declared that a sister had the right to refuse her brother's request for delegation while all the participants in the daughters' focus group declared that she can't and there was disagreement between the participants in the sons' focus group. The consequences of this refusal were also discussed and how this might affect her relationship with her brother and with the rest of the family.

When my aunt and her husband died, her eldest son became guardian of his brothers and sisters… They used to have a good relationship, then suddenly he disagreed with one of the sisters and she refused to delegate control of her bank account. Instead, she asked the next brother to be the one who was authorised… Initially, the rest of the family refused and told her that because he was the eldest his word must be obeyed, but she insisted. At first, he was upset and the rest of the family was upset, but afterwards they accepted that and it became normal. [P9, F]

4.1.5 Couples

Focusing on sharing practices in couples, all six participants in the first round of the data collection agreed that sharing card PINs was a sign of trust between the husband and wife. For example, one participant said:

It's all about trust between wife and husband ... and this is more important than just a bank account ... In our culture and with our priorities, this is much more important than materialistic belongings ... and in our marriage, I have given her things more important than a bank account. [P20, M]

For this participant, trust between him and his wife was more important than their bank accounts and he referred to the cultural norms which defined his priorities. It is worth mentioning that according to Saudi cultural norms, most marriages are family arrangements. The groom's mother chooses the girl who she thinks will be most suitable for her son, then, usually after six to 12 months of engagement, they are married. According to the participants, trust and sharing arise during the first months of the relationship which means that couples begin to share sensitive banking information when they have known each other for only a very short period of time.

Another participant declared that he trusted his wife and decided to share such information with her from the fifth month of their engagement, even before they were married. It is noteworthy here that couples usually started sharing before having children, although having children increased the propensity to do so. Thus, a female participant said:

Sharing between me and my husband happened before we had kids. But if I want to be honest, I would say that having kids puts a big emphasis on the need for sharing. For instance, when I want to buy something for them and I don't want to go out, I just send them with their father and give him my card to use. [P5, F]

Similarly, another participant stated that she had shared her credentials with her husband from the first month of their marriage:

First of all, mutual trust between us is natural; also, I used to see my father and his wife exchanging their passwords ... and there were no problems. [P8, F]

Thus, this participant saw mutual trust as natural because she was used to seeing a sharing relationship between her father and stepmother. One explanation is that marriage creates the expectation of mutual trust between husband and wife, which in turn creates the expectation that banking information will be shared; making sharing a consequence of marriage.

Regarding online banking, four participants declared that they never shared their passwords with their partners, while the other two said that they did share them. Those who did not share explained that because they used their ATM cards more than online banking, the need to share online passwords had not arisen and each was responsible for his or her own online banking, a situation unaffected by sharing the same device to access online accounts. A participant said:

I need to give him my card to get cash but I do not need him to access my account online ... But if we needed to I wouldn't mind giving it to him. [P4, F]

As to the two participants who did share their online banking passwords, they explained that the partner with more IT experience in each of their couples was responsible for the online banking for his or her spouse. In one case the wife was responsible and in the other it was the husband.

4.1.6 It is safe

The majority of the participants shared their bank credentials and saw no risk in doing so because the account holder is informed by SMS of any operation on the account, such as transfers, withdrawals or online access. Most participants considered that this service would protect the account from any fraudulent transaction by allowing the account holder to monitor the account. A member of the wives' focus group said, for example:

My husband has my card. When I ask him to do something then he does it. He doesn't withdraw until I ask him to do so, as I receive an SMS on my mobile. I didn't do it due to a lack of trust. It was just a good new service.[P2, F]

SMS are also used by some banks to protect their online banking: the account holder receives a one-time PIN by mobile phone when they request access to their account online; again, this service protects bank customers as well as allowing them to monitor access requests to their accounts.

In addition, some participants claimed that sharing a card and its PIN was relatively safe because the card would be kept by the account holder at all times and no one would have access to the account without the card, even if he knew the card PIN a member of the daughters' focus group said:

Why would I stop this sharing? He doesn't have the card with him all the time… and if he has the card, he just makes the withdrawal and brings it back to me. He doesn't take anything.[P13, F]

Moreover, *sharing is safe* because according to the participants permission must be obtained for every transaction.

Even if my wife's card is with me all the time, it is unacceptable to use it without her permission... In Islam even if she is rich the man has not got the right to take from his wife money without her permission. [P26, M]

4.1.7 Need to know

Some participants saw sharing as a necessity in order to have a backup of their credentials given the perceived importance of this information. Many in the daughters', wives' and husbands' focus groups expressed the view that a wife would need to know about her husband's assets and obligations, in order to avoid any unpleasant surprises if he should die suddenly. A participant in the daughter focus group said that her father had gathered the family together before he died and told them everything about his money and his bank accounts, because he used to help certain people with donations and wanted to be sure that this help would not stop after his death. A related reason for sharing this information was to avoid problems with the family of an account holder who had died, as illustrated by the personal experience of a member of the wives' focus group:

Actually, it is important that the wife asks frankly about those things, given what I have learned from the situation I went through after the death of my husband. I didn't know anything, not even his card PIN, and all those things had to be taken care of by his brother and there were a lot of problems. I couldn't do anything and I couldn't get out as I was in mourning, so the wife should know these things. [P14, F]

Participants also mentioned the importance of needing to share the credentials for cases such as the account holder's illness or travel.

Once I had some heart troubles and I got sick. During that time, my family needed to spend money and I trusted my wife who was careful – thank God – so although I stayed at the hospital for more than a month, my financial situation was fine thank God. [P18, M]

My sister shared her credentials with me because she is sick, she can't get out frequently as she suffers of arthritis, sometimes she needs things and she can't do it. For example the last transaction I did for her two weeks ago was to pay her daughter's school fees. [P12, F]

I share my credentials with my wife because sometimes I travel abroad so I couldn't be there if she needed money to transfer to someone or to pay for electricity or something of the sort. [P23, M]

4.2 Has Not Shared

Not sharing practices between family members can be divided into:

4.2.1 Having two accounts

Some male account holders may have more than one account: one to share with family members, and another private or secret account. This practice was mentioned by two participants in the wives focus group, two participants in the first round, and a participant in the sons' focus group. A participant in the wives' focus group claimed that this was done for control rather than distrust:

The problem is not lack of trust between them, but for instance he might want to build a new house or expand his business and if his wife knew how much he had in the bank she'd ask him to buy new furniture for their home or something else… and in fact that's what we actually do.[P6, F]

Having a secret account was seen as quite common and many wives accepted it, although not all did so. A member of the wives' focus group said:

Normal? I don't see it as normal. I do not consider it normal to have two accounts, because Sharia says that he has to pay for my expenses. This is ... a religious obligation ... and I don't permit anything except what our religion permits. [P9, F]

But another member of the same group thought that it was a matter of privacy for her husband, who had the right to keep his balance secret:

If he insists then it is his personal freedom. I shouldn't interfere. If he insists this is his affair, he can say: 'You have your needs and I fulfil them and you don't have to know my stuff. You can't interfere'. [P3, F]

4.2.2 Temporary divorce

Four participants felt that withholding banking information from a spouse would be a sign of distrust:

If she does not like to share such information with me, then she might have a trust problem ... definitely there is a trust problem. [P16, M]

It's natural for couples to share everything ... Otherwise, there will definitely be a trust problem. [P7, F]

Interestingly, one participant admitted distrusting her husband, stating that they used to share banking information until she felt that she could no longer trust him, after they had been temporarily divorced.[3] Although the couple had decided to re-unite she could not trust her husband again, so she decided not to share sensitive information with him. He, however, still shared his credentials with her:

I used to trust him and tell him everything about my account and how much I had. But after the first divorce, I became afraid; then after the second divorce, I got really afraid. Therefore, I think that divorce is the main reason why I am hiding everything, not just my money. … I know his card password; he always tells me everything about his account and he is really honest with me about money and all his financial stuff. I know that even if he gave some money to his parents, he would tell me about it. Therefore, you can say that he trusts me more than I trust him. [P13, F]

5. SECURITY ANALYSIS

5.1 Summary statistics

- In total, 29 household bank customers participated in this study, 15 being female and 14 male.
- All participants had bank accounts and some had more than one, but it was noted that the use of joint accounts and savings accounts was rare. None of the participants had a

[3] In Islam, couples may obtain a temporary divorce for three months. If they are able to fix their problems during this period, they can come back to each other and continue their married life. If not, this temporary divorce becomes permanent. This process can be used only twice during the life of a marriage.

savings account, although some said that they might use their current accounts for savings. Similarly, joint accounts were not popular: only one couple in the study claimed to use one.

- Twelve (85.7%) of the male customers shared their cards PIN with family members and two did not.
- Thirteen (86.7%) female customers shared their cards PIN with family members and two did not, one also shared her card PIN with friends.
- Only three (20%) female participants used online banking one of them only shared her online credentials – username/password.
- Eleven (78.6%) male participants used online banking four of them shared their online credentials – username/password- with their family.
- Thirteen (86.7%) female and all male participants reported using sales points.
- Only two of each gender (13.3% female, 14.3% male) said that they used online shopping.

5.2 Credential Sharing

Is it safe? 86% of the participants in this study share their bank credentials for various reasons. One of these reasons is that they believe that sharing is safe, and also that they use different methods to ensure the security of their accounts (mostly through SMS notifications). However, the possibility of misusing the shared credentials (e.g. someone uses this information without permission to gain access to another person's account, or to use the money in that account) still exists. By discussing this issue, all the participants in the first round of data collection declared their belief that their own partners, siblings or parents would never do this, for reasons of religion and culture. First, the participants described their partners, siblings and parents as religious believers for whom doing such things would be against their religion. Thus, a participant was sure that her husband

... would not get any money out of my accounts without telling me first, and also because I know this from past experience ... how he dealt with his father's inheritance when he was responsible for it. He divided it equally between himself and his siblings. He had a religious deterrent. [P5, F]

Secondly, they considered such practices alien to Saudi culture. A participant, for example, said that he did not expect his wife or siblings to steal money from his account in any circumstances, because

this is unusual here in our culture . . . I am sure this does not happen here. [P21, M]

In contrast to this widely mentioned belief, one of the participants reported that she had twice suffered misuse of sharing by her husband:

He used the money in my account without my permission twice because he thought I wouldn't mind ... The first time he took less than two thousands riyals ($540), and around ten thousand ($2700) the second time. After the first time I complained in an indirect way and I thought this was enough. But the second time I got really upset and discussed this with him clearly and he stopped. [P11, F]

Despite these two incidences of misuse of sharing by her husband this participant stated that she still shared her card PIN with him. When asked why, she replied that the misuse had happened four years ago and she did not think it would happen again.

Another case appeared when a participant in the daughters' focus group mentioned what happened to her sister when she delegated her husband over her account and he misused this by taking a loan of half million riyals ($133,300) by using her account and because of that she is now paying the instalments for five years.

Accountability and misuse: Despite the belief that sharing credentials is safe, and that family members will not abuse the trust placed in them, there are clear risks associated with this widespread practice.

First, the banking regulations presented in section 2.2 clearly state that "banks are responsible for providing secure and safe systems and services for their customers ***unless the customer fails to safeguard their account user number or password and divulges it to a third party***" [45] (emphasis added). This can mean that banks are free to disclaim any responsibility for cases of fraudulent access to the accounts of people who have shared their credentials – even if the fraud did not happen because of this practice.

Second, accountability for actions undertaken in a given bank account are strongly undermined when different people share the same authentication credentials. In the case of a dispute requiring an investigation into who authorised a given transaction, the habit of sharing credentials can significantly undermine the principle of non-repudiation, allowing an account holder to deny having authorised the transaction.

While the reasons behind this practice have been explored in section 4, mitigating the risks associated with it is a challenge. One of our key findings is that many of the reasons for sharing stem from a sense of pre-existing trust (or a need to appear trustworthy), and also as a means of solving the practical problem of accessing funds (particularly for females). These are different problems which lend themselves to different solutions.

Sharing for trust: While joint accounts do not appear to be common in the KSA, they provide a simple means for different parties to have access to the same account. This does not address the issue, however, of allowing access to a personal account as a demonstration of trust. Another option would be for banks to provide account holders with the means of explicitly delegating access to their personal accounts to other individuals, without having to divulge the authentication credentials. By introducing such a mechanism, the principle of accountability would be easier to achieve, and the possibility of revoking access would also be available to account holders. While such a system might still be undermined by the principle that people share in order to demonstrate trust, it would at least provide a means for trust to be shown without undermining banking accountability.

Access to funds: Convenience and the logistical problem of women gaining access to cash, seems to be another key factor in the need to share credit cards and PINs with others. By doing this, however, the card holder relinquishes authority over the account: leading to the possibility of someone behaving other than intended, or at a later date taking the card and misusing their access. One possible solution to this problem would be to enable account holders to authorise an individual to withdraw a specified sum of money from an ATM on a one-off (or even recurring) basis without requiring either the card or the PIN. A possible high-level design for such a system would be to allow account holders to send an authenticated instruction to the bank to dispense a specific sum of money. The bank would reply with a one-use access code that could be given to the trusted person, who would then use this code in an ATM to withdraw the specified sum of money without having to have the card or know the PIN.

5.3 Future Work

We have identified a number of themes which we would like to highlight for further research.

Bank regulation knowledge: As mentioned in Section 2.2, SAMA regulations state that it is the bank's responsibility to educate customers about the regulations, specifically, the consequences of unauthorized use of an account by sharing of the account credentials with a third party. However, all the participants indicated that their knowledge of bank regulations was both limited and very general. A participant said:

I have four accounts in four different banks and it has not happened at any time that a banker has told me anything about bank regulations or how to protect my credentials. [P25, M]

Some participants mentioned that the only warning given by the banks was when a customer used a card device to register the PIN and was told not to disclose the PIN to bank employees. This lack of awareness about the regulations among bank customers can be considered as one of the reasons behind commonly sharing credentials for 25 (86%) participants especially if this sharing seems safe and solving problems, and without consequences. By focusing more on educating customers about the regulations and the consequences of breaches, it would possibly reduce the chances of sharing.

Online banking use: According to female customers, online banking use was not prevalent among females (only 20%), despite providing convenient access to an account which might be seen as a helping to mitigate the issues of constraints on movement. By asking about the reasons behind this, a female customer claimed that she was determined to learn how to use online banking but she could not, because when she asked her bank for help with this service she was told that she must do it by herself while other participants declared that they tried activating online services but they failed and did not try again.

The Internet was introduced in KSA for public use in 1999 [12] and since then Saudi women have used it for socialising, decreasing social pressures, running businesses, all while remaining anonymous in the home with no obligation to meet men in person [23, 39]. Statistics taken from the Asbar Centre for Studies, Research and Communication [15] indicate that 55% of Internet users are men and 45% women, however, the proportion of female users is predicted to increase as the Internet can be seen as a solution for those who are isolated and/or lacking mobility [40, 41].

As a result, IT literacy does not seem to be a likely reason for the lack of online banking use in females. Further exploration of the process for registering and activating online banking services for different Saudi banks has found that this process requires the use of an ATM. As mentioned above, female access to an ATM depends on the availability of vehicle driven by a chauffeur or a mahram, and can be problematic. One possible solution would be to modify the process of activating online banking to use security mechanisms which do not require physical mobility (e.g. telephone authentication, or wholly online authentication).

Sale points: Card payments at sale points was also discussed as a means of alleviating the problem of accessing money for women. Two women did not use this service, explaining that they had experienced problems with the service before: in one case, a transaction was incorrectly duplicated and led to two debits to her account, requiring her to go to her bank to claim a refund. According to her, this was very inconvenient, especially given the constraints on movement.

P15 [F]: (...) as for the using of the card to buy things, those are very rare occasions. It happened twice or three times only that I went to buy something and I didn't have the full amount so I had to use my card to pay. But it is very rare.

Q: why?

P2: Frankly I don't trust the Saudi bank network… I don't like to use the card. It happened before that the transaction was duplicated and to solve this problem I went to the bank branch to claim a refund. So for me this service created more problems.

The second woman stated that a transaction was seemingly successful, but had not in fact resulted in the correct debit, fuelling her distrust of the system. Given the limited evidence surrounding this, it is not clear how this problem arose, or indeed how widespread it is.

P1 [F]: I don't use sale points because I don't trust this service. It happened to my brother in law, who is a cashier in a shop: when a customer bought things and used her card to pay, the transaction finished smoothly and successfully but then he was informed by the bank that this transaction had not completed and the money was not credited from the customer's account. When he contacted the customer to pay for the goods because the transaction failed and the money was not credited from her account, she refused saying it is not her fault and the receipt showed that the transaction was successful and because of that he had to pay for the goods.

Online shopping: According to the statistics, the use of online shopping is very limited between the participants. When asked why they did not shop online, many participants said that it was because they did not have credit cards which could be used for online shopping; their bank cards could be used only in ATMs and at sales points. Saudi banks often routinely issue ATM/sales-point cards to customers who open an account, but provide credit cards only if the customer requests one. Participants also said that they were wary of trying online shopping without guidance or help from the bank. As an alternative, many customers reported using the services of an intermediary with experience in online shopping. The customer would contact such a person by phone or email to request certain goods, then transfer the money to the account of the intermediary, who would purchase the goods online using his credit card and send them on to the customer.

6. CONCLUSION

Bank customers in the KSA give various reasons for sharing their credentials with their family members including mutual trust, simplifying life and laziness, movement limitation for females and the high esteem placed in parents. They perceive little to no harm or risk from sharing, relying on being able to monitor their accounts by receiving an SMS for every transaction and keeping their banking card safe.

We have discussed the ways in which these practices can undermine good security principles, and proposed a number of solutions aimed at working with the factors that underpin these practices, however more needs to be done to further investigate the issues and assess the effectiveness of our proposals.

7. ACKNOWLEDGMENTS

We would like to express our utmost gratitude to Umm-AlQura University (Makkah, KSA) for generously funding this research project and we would also like to thank the participants in this study for their time and contribution.

8. REFERENCES

1. Al Rajhi Bank. *About Us*. 2014 [cited 22/2/2015; Available from: http://www.alrajhibank.com.sa/en/about-us/pages/default.aspx.
2. Al-Garni, M., *The Impact of Family Structure and Family Function Factors on the Deviant Behaviors of High School Students in Makkah City, Saudi Arabia*. 2000, The Ohio State Univesity.
3. Al-Suhaimi, J., *Consolidation, competition, foreign presence and systemic stability in the Saudi banking industry*, in *The banking industry in the emerging market economies: competition, consolidation and systemic stability*. 2001: Basel, Switzerland.
4. Al-Turki, S., *Al-Mar'ah fi al-Saudiyah The woman in Saudi Arabia*. 1987: London: Highlight Publications.
5. Alarifi, A., H. Tootell, and P. Hyland, *A study of information security awareness and practices in saudi arabia*, in *Communications and Information Technology (ICCIT)*. 2012. p. 6–12.
6. AlBawaba, *Shrinking family size to help Saudi housing sector, Al Bawaba (Middle East) Ltd., Saudi Economic Survey*. 2010.
7. Algarni, A., *Policing internet fraud in saudi arabia: expressive gestures or adaptive strategies?* Policing and Society: An International Journal of Research and Policy, 2013. **23**(4): p. 498–515.
8. Alkaabi, A. and C. Maple, *Cultural impact on user authentication systems.* Int. J. Business Continuity and Risk Management, 2013. **4**(4): p. 323-343.
9. Alkhalifa, H., *The State of Distance Education in Saudi Arabia.* elearn, 2009. **2009**(10).
10. AlLily, A., *On line and under veil: Technology-facilitated communication and Saudi female experience within academia.* Technology in Society, 2011. **33**(1--2): p. 119 - 127.
11. Almunajjed, M., *Women in Saudi Arabia Today*. 1997: London: Macmillan.
12. Alsalloum, O., *Factors affecting the Internet adoption in Riyadh city.* Journal of King Saud University: Administrative Sciences 2005. **18**(1): p. 1-14.
13. Anthony, B., *Tribes and Tribulations: Bedouin Losses in the Saudi and Iraqi Struggles over Kuwait's Frontiers.* {British Journal of Middle Eastern Studies, 2005. **32**(2).
14. Anthony, B., *Last Battles of the Bedouin and the Rise of Modem States in Northern Arabia, 1850- 1950*, in *Nomadic Societies in the Middle East and North Africa: Entering the 21st Century*, D. Chatty, Editor. 2006, Brill. p. 49--77.
15. Asbar Centre for Studies Research and Communication, *Estkhdamat Al Internet fi Al Moujtamah al-Saudi 'Internet usage in Saudi society'*. 2007, Asbar Centre for Studies, Research and Communication: Riyadh.
16. Baki, R., *Gender-segregated Education in Saudi Arabia: Its Impact on Social Norms the Saudi Labor Market.* Education Policy Analysis Archives, 2004. **12**.
17. Bloor, M., et al., *Focus groups in social research*. 2001, London: SAGE.
18. Champion, D., *The Paradoxical Kingdom: Saudi Arabia and the Momentum of Reform*. 2003: New York: Columbia University Press.
19. Coles-Kemp, L. and D. Ashenden. *Community-centric engagement: lessons learned from privacy awareness intervention design*. in *Proceedings of BCS HCI 2012 Workshops: Designing Interactive Secure Systems*. 2012.
20. Deave, S., *The contemporary Saudi women*, in *A world of women: anthropological studies of women in the societies of the world*, B.a.c. E, Editor. 1980, New York: Praeger.
21. Deif, F. *Perpetual minors: human rights abuses stemming from male guardianship and sex segregation in Saudi Arabia. London: Human Rights Watch.* 2008; Available from: http://www.hrw.org/sites/default/files/reports/saudiarabia0408webwcover.pdf.
22. Del Castillo, D., *Teaching through an electronic veil.* The Chronicle of Higher Education, 2003. **49**(29).
23. Dholakia, R., N. Dholakia, and N. Kshetri. *Gender and Internet usage*. 2003 8/3/2015]; Available from: file:///C:/Users/lina2267/Downloads/09e41510a5f11af9d9000000.pdf.
24. Flechais, I., M. Jirotka, and D. Alghamdi, *In the balance in Saudi Arabia: security, privacy and trust*, in *CHI '13 Extended Abstracts on Human Factors in Computing Systems (CHI EA '13)*. 2013, ACM, New York, NY, USA, 823-828: Paris, France.
25. Friedman, B. and J. Grudin. *Trust and Accountability: Preserving Human Values in Interactional Experience*. in *CHI '98*. 1998. ACM.
26. Glaser, B. and A. Strauss, *The Discovery of Grounded Theory: Strategies for Qualitative Research*. Vol. 1. 1967: Aldine. 271.
27. Inglesant, P. and M.A. Sasse. *The true cost of unusable password policies: password use in the wild.* in *CHI '10*. 2010a. Atlanta, Georgia, USA: ACM.
28. Jerichow, A., *The Saudi File*. 1998: New York: St. Martin's Press.
29. kaspersky security bulletin. 2009; Available from: http://www.securelist.com/en/ analysis/204792101/Kaspersky_Security_Bulletin_2009_Statistics_ 2009.

30. Kaye, J.J., *Self-reported Password Sharing Strategies*, in *the SIGCHI Conference on Human Factors in Computing Systems, CHI '11*. 2011, ACM, New York, NY, USA: Vancouver, BC, Canada. p. 2619--2622.

31. Krimly, R., *The Political Economy of Adjusted Priorities: Declining Oil Revenues and Saudi Fiscal Policies.* Middle East Journal, 1999. **53**(2): p. 254-267.

32. Lipsky, G., *Saudi Arabia: its people, its society, its culture*. 1959: New Haven: Hraf Press.

33. Lunt, P. and S. Livingstone, *Rethinking the focus group in media and communications-research.* Journal of Communications., 1996. **46**(2): p. 79-98.

34. Mathiasen, N.R. and S. Bodker. *Experiencing Security in Interaction Design*. in *SIGCHI Conference on Human Factors in Computing Systems, CHI '11*. 2011. Vancouver, BC, Canada: ACM, New York, NY, USA.

35. Mirza, A., *Students perceived barriers to in-class participation in a distributed and gender segregated educational environment.* The Electronic Journal of Information Systems in Developing Countries, 2008. **35**(7): p. 1--7.

36. Moaddel, M., *THE SAUDI PUBLIC SPEAKS: RELIGION, GENDER, AND POLITICS.* International Journal of Middle East Studies, 2006. **38**(01): p. 79--108.

37. Olson, J.S., J. Grudin, and E. Horvitz. *A Study of Preferences for Sharing and Privacy*. in *CHI EA '05*. 2005. Portland, OR, USA: ACM.

38. Pew Research Center. *MAPPING THE GLOBAL MUSLIM POPULATION*. 2009 1/10/2013]; Available from: http://www.pewforum.org/files/2009/10/Muslimpopulation.pdf.

39. Pharaon, N., *Saudi women and the Muslim state in the twenty-first century.* Sex Roles, 2004. **51**(5/6): p. 349–66.

40. Ridings, C. and D. Gefen, *Virtual community attraction: why people hang out online.* Journal of Computer Mediated Communication, 2004. **10**(1).

41. Rohall, D., S. Cotton, and C. Morgan, *Internet use and the self-concept: linking specific uses to global self-esteem.* Current Research in Social Psychology 2002. **8**(1): p. 1-19.

42. Sasse, M.A. and I. Flechais, *Usable security: Why do we need it? How do we get it?*, in *Security and Usability: Designing Secure Systems that People Can Use*, L.F. Cranor and S. Garfinkel, Editors. 2005, O'Reilly. p. 13--30.

43. Saudi Arabian Monetary Agency, *Development and restructuring of the Saudi banking system*, in *BANK RESTRUCTURING IN PRACTICE*. 1999, Bank for International Settlements Information, Press & Library Services: Basel, Switzerland. p. 183-196.

44. Saudi Arabian Monetary Agency. *Regulations for the Issuance and Operation of Credit and Charge cards - page: 12*. 2008 8/3/2015]; Available from: http://www.sama.gov.sa/RulesRegulation/Rules/Pages/REGULATIONS_FOR_ISSUANCE_AND_OPERATIONS_OF_CREDIT_AND_CHARGE_CARDS.pdf.

45. Saudi Arabian Monetary Agency. *Electronic Banking Rules, page: 14*. 2010 8/3/2015]; Available from: http://www.sama.gov.sa/RulesRegulation/Rules/Pages/E_banking_Rules.pdf.

46. Saudi Arabian Monetary Agency. *Consumer Protection Principles*. 2013 8/3/2015]; Available from: http://www.sama.gov.sa/RulesRegulationsandCirculars/ConsumerProtection/Documents/ConsumerProtection.pdf.

47. Saudi Hollandi Bank. *History*. 2015; Available from: http://shb.com.sa/en/about/AboutSHB/History.aspx.

48. Singh, S., *The Social Dimensions of the Security of Internet banking*. JTAER, 2006. **1**(2): p. 72-78.

49. Singh, S., *Secure shared passwords: The social and cultural centered design of banking*. journal of financial transformation, 2008. **23**: p. 110-114.

50. Singh, S., et al. *Password sharing: implications for security design based on social practice*. in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 2007a. San Jose, California, USA: ACM.

51. Singh, S., et al. *Security Design Based on Social and Cultural Practice: Sharing of Passwords*. in *Proceedings of the HCI International Conference on Usability and Internationalization*. 2007b. Beijing, China: SpringerLink.

52. Singh, S., A. Cabraal, and G. Hermansson. *What is your husband's name?: sociological dimensions of internet banking authentication*. 2006. Sydney, Australia: ACM.

53. SYMCINT. 2012; Available from: http://www.symanteccloud.com/en/gb/mlireport/SYMCINT_ 2012_06_June.pdf.

54. SYMCINT. 2013; Available from: http://www.symanteccloud.com/mlireport/SYMCINT_2013_ 01_January.pdf.

55. The South Asian Times, *Saudi Arabia among Top 10 nations in terms of density of ultra-wealthy households, HT Media Ltd.* 2011.

56. Timewell, S. *Banks start to close the gender divide in Saudi Arabia*. 2014; Available from: http://www.thebanker.com/World/Banks-start-to-close-the-gender-divide-in-Saudi-Arabia.

57. Trendmicro. 2013; Available from: http://www.trendmicro.co.uk/media/misc/zero-days-hit-users-hard-at-the-start-of-the-year-en.pdf.

58. Weirich, D. and M.A. Sasse. *Pretty good persuasion: a first step towards effective password security in the real world*. in *Proceedings of the 2001 workshop on New security paradigms*. 2001. Cloudcroft, New Mexico: ACM.

59. Whitworth, B. and A. Ahmad, *Socio-Technical System Design*, in *The Encyclopedia of Human-Computer Interaction*, M. Soegaard and R.F. Dam, Editors. 2013, Aarhus, Denmark: The Interaction Design Foundation.

60. Yamami, M., *Feminism and Islam: Legal and Literary Perspectives.* 1996: New York: New York University Press.

Too Much Knowledge? Security Beliefs and Protective Behaviors Among United States Internet Users

Rick Wash
School of Journalism
Michigan State University
wash@msu.edu

Emilee Rader
Dept. of Media and Information
Michigan State University
emilee@msu.edu

ABSTRACT

Home computers are frequently the target of malicious attackers because they are usually administered by non-experts. Prior work has found that users who make security decisions about their home computers often possess different mental models of information security threats, and use those mental models to make decisions about security. Using a survey, we asked a large representative sample of United States Internet users about different causal beliefs related to computer security, and about the actions they regularly undertake to protect their computers. We found demographic differences in both beliefs about security and security behaviors that pose challenges for helping users become more informed about security. Many participants reported weakly held beliefs about viruses and hackers, and these were the least likely to say they take protective actions. These results suggest that all security knowledge is not the same, educating users about security is not simply a more-is-better issue, and not all users should receive the same messages.

1. INTRODUCTION

For most people, protecting their home computers from hackers and viruses is rather difficult. They see celebrities having their information stolen by hackers [38]; they don't understand how anti-virus software works [35]; they can't see the benefits of patching, but frequently see downsides [33]; and even when they decide they want to protect themselves, they can't always correctly configure their computers [37].

Despite this difficulty, home computer users have to make many security-relevant decisions every day. They receive links to invalid or suspect websites in their email, and have to decide whether to click on them. They hear about problems with computer viruses, and need to decide whether to purchase and use anti-virus software. The anti-virus scan slows down their computer, and they need to decide whether to postpone it to regain their use of the computer or wait until it finishes.

Wash [35] found that individuals can possess multiple different "folk models" of the threats they are worried about. Each different folk model leads users to make different choices when faced with these everyday computer security decisions. Rather than characterizing users along a continuum of less knowledge to more knowledge as is traditionally done [25], Wash's work suggests that there are a number of different beliefs that each can lead to different behaviors.

To better understand how different types of beliefs about computer security threats can affect the way people make choices to protect their computers, we asked a large nationally representative sample of U.S. Internet users about both their computer security beliefs and the security actions they take. The most common beliefs involve more direct and visible threats, and these beliefs are associated with more positive security decisions. More sophisticated security beliefs that involve more technological knowledge often are associated with fewer precautions. We also find that more educated users and older adults (50+) tend to have these more sophisticated beliefs, where younger people and people with lower levels of education tend to focus more on direct and visible threats.

2. RELATED WORK

2.1 Home Computer Security

A recent Pew survey shows that over 76% of the US population accesses the Internet from their home [39]. Computers in people's homes have changed our society, but they have also imposed new risks; home computers are under constant threat. Additionally, we are now seeing increased use of mobile phones, tablets, and other Internet of Things devices that are connected to the Internet and all of these are potential targets [31].

Despite not being security experts, home users are tasked with administering and making security decisions for their computers and devices. This makes protecting these computers difficult. Home computer users who feel psychological ownership for the computer are more likely to engage in protective measures [3]. One of the major strategies they use is to find ways to delegate the responsibility for security to some external entity, which could be technological (like a firewall), social (another person or IT staff), or institutional (like a bank) [13].

Gross and Rosson [18] studied what security knowledge end users, who were not directly responsible for security but had access to sensitive information, possessed in the context of large organizations. Users' security knowledge was "neither comprehensive nor sufficient" to maintain proper secu-

Symposium on Usable Privacy and Security (SOUPS) 2015, July 22–24, 2015, Ottawa, Canada.

rity, but common security actions such as locking the screen when away were better understood and practiced. Users in both organizational [18] and home settings [33] also frequently conflate security and functionality failures or problems. For example, users refuse to install future security updates because past updates changed critical user interface elements [33].

A wide variety of security advice has been provided to computer users, particularly in large organizations. Researchers in large organizations have investigated the effects of different kinds of training programs and security policies on security outcomes [2, 12]. Wash [35] lists 12 pieces of security advice found from Microsoft, CERT, and US-CERT that are specifically targeted at home computer users. Egelman and Peer [14] identified 30 security behaviors that represent common advice given to home computer users. Hoban et al. [20] examined much of this advice, and found that online educational materials often focus on virus and phishing threats, but rarely mention hackers as a threat. Larose, Rifon, and Enbody [25] recommend emphasizing personal responsibility to encourage users to engage in protection behaviors. However, users often do not follow the advice found in security education materials and persuasive appeals. Herley [19] argues that when non-expert users reject security advice, it is often rational for them to do so. Advice to end users often ignores the costs of their time and effort, and therefore overestimates the net value of security.

All of this research uses small non-representative samples. This makes it impossible to understand how prevalent different folk models or different security behaviors are in society [11]. We seek to measure both mental models of security and security behaviors in a large, representative sample. By understanding what behaviors are common and among whom, we can examine society-wide vulnerabilities due to mental models. We can also better understand what types of beliefs make people particularly vulnerable to security problems.

2.2 Mental Models of Security

An important aspect of the security decisions of home computer users is their existing knowledge about computers and computer security issues. More knowledge about common security issues is frequently found to be correlated with intention to behave securely [25]. However, most studies find that knowledge is not enough, and that additional motivations must be in place for people to make secure decisions [3, 26].

Most research in this area has measured security knowledge on a continuum from *less knowledge* to *more knowledge*. For example, Kumar et al. [24] counts the number of common security measures that the user is aware of, and finds that people familiar with more of these are more likely to engage in security behaviors. Shillair et al. [30] measure knowledge by asking about two forms of malware (spyware and Trojans), and combining answers into a single measure of low versus high knowledge. They did not find that more knowledge increased behavior, but prior knowledge had an important interaction effect for what type of communication was best for increasing security intentions.

However, knowledge about security does not easily fall into a more-is-better continuum. Wash found that home computer users have a variety of different "mental models" of security threats [35], and that these models are often used to make security decisions. Mental models describe how a user thinks about a problem; it is the model in the person's mind of how things work. People use these models to make decisions about the effects of various actions [21, 17] by cognitively simulating the actions and running the model forward in time to examine potential outcomes. Mental models are not the same as knowledge; rather, they usually represent a set of causal*beliefs* that a person possesses that he or she uses to guide decisions and behavior [9]. Security experts differ from non-experts in the mental models that they use. Asghapour et al. [4] conducted a card sorting experiment; participants were instructed to match words with a set of computer security related concepts. They found that experts and non-experts show differences in which analogy (medical, crime, etc.) they felt the concepts were closest to.

Economists often talk about products as being horizontally differentiated or vertically differentiated [34]. Products are differentiated vertically when everyone agrees which product is better than the other (e.g. $100 is better than $10). Products are horizontally differentiated when some people like one product, and other people prefer the second product (e.g. baseball vs. football).

Using this to draw an analogy, the traditional way of measuring security knowledge is vertically differentiated: more knowledge is better for making good security decisions than less knowledge. However, we follow the lead of Wash [35] and treat security knowledge as horizontally differentiated: there are a variety of causal beliefs about computer security, and even simplified or incorrect beliefs can lead to good decisions. We seek to measure a number of different types of beliefs about computer security, and then identify which types are associated with positive security behaviors.

2.2.1 Measuring Mental Models of Security

Many scholars have examined mental models in the anthropology tradition. For example, Kempton [22] used both semi-structured interviews and analysis of log data to study folk models of thermostat technology in an attempt to understand the wasted energy that stems from poor choices in home heating. Wash [35] used similar interviews to study mental models of home computer users. D'Andrade [9] summarizes interview-based approaches to discovering mental models. Interview-based approaches allow examination of the details of an individual's mental model; unfortunately, interview-based approaches are too time-intensive to be used on large samples, and therefore cannot be used to measure prevalence of different beliefs in a population.

A number of psychologists have studied mental models using human-subjects lab experiments. Johnson-Laird has done a long series of studies examining how mental models form, how they represent time, and how they are used to make decisions [21]. He never measures the models directly, but rather provides information to participants and then measures behaviors and decisions that emerge from the models. This technique finds patterns in what mental models look like and how people use them, but does not reveal the details of the specific models.

A promising approach for measuring mental models of security is the card-sorting method used by Asghapour et al. [4]. They made a list of a number of words related to security and risk, and asked participants to sort these words into piles. They then analyzed these piles using dimension reduction techniques to find patterns among the words. They pre-specified which words were associated with each mental model, which limited the ability of the study to discover new models. It is also difficult to measure reliability and prevalence of the discovered models using this method.

2.3 Security Intentions and Behaviors

Measuring actual security decisions and behaviors with a survey is very difficult. Most security decisions depend on the *context* of the decisions [13]. While a person might generally prefer to run regular anti-virus scans, they might forgo a scan if they are in the middle of an important phone call or up against an important deadline. It is difficult to replicate real-world contexts in a survey.

Additionally, most security decisions are *repeated* [35]. Installing anti-virus software may be a one-time decision, but decisions like "should I click on this shady link?" or "what password should I choose for this site?" happen frequently. Surveys cannot usually ask about every instance of a security decisions, both because that would make the survey too long and tedious, and because it is difficult to know about the security decisions ahead of time.

To address these issues, most surveys follow one of two approaches. The first approach is to ask questions about *general behavioral intentions* [11]. Asking about intentions focuses on the future and what the participant wants to do. Intentions are a natural focus when conducting research that involves a manipulation, and you want to measure whether the person's future behavior is likely to have changed. There are also theoretical reasons why intentions are likely to map to behavior [1].

A second approach is to ask questions asking participants to recall how frequently they have undertaken a behavior in the past. This approach can be subject to recall bias, since it depends on the memory and honesty of the participant. However, it focuses directly on actual behavior rather than relying on the theoretical link between intention and behavior [11]. In security in particular, participants often find it difficult to enact their intentions due to lack of skill [13] or confusing interfaces [37], and therefore past frequency might have a stronger connection to actual behavior than intention.

A number of recent surveys have used different questions to ask about general intentions toward computer security decisions. Larose et al. [25] ask their participants about their intentions to do eight specific security behaviors on a 7-point Likert scale, and then average the results as an overall measure of preventative intentions. Anderson and Agarwal [3] use two sets of questions about general protective intentions. One set asks general intentions (e.g. "I am likely to take security measures") about the participant's home computer, and the other set asks similar questions about protecting "the Internet." Egelman and Peer recently developed a new security behavior intentions scale that focuses on four common security behaviors: device securement, password generation, proactive awareness, and software updating [14].

With the exception of the scale created by Egelman and Peer (which was not yet published at the time we conducted this study), all of the measures we found used a unidimensional more-is-better measure of intention to make secure decisions. We wanted a measure that could capture more horizontal differentiation between security decisions. We seek to understand how different beliefs – different mental models – can lead people to make different types of security-related decisions. Our approach does not assume from the outset that more knowledge, or more sophisticated beliefs, are better for security. Rather, our survey is a measure of the association between patterns of causal beliefs and protective behaviors. Our survey allows us to better understand how different types of security knowledge can lead to *different* security decisions, rather than simply *more or less secure* decisions.

3. SURVEY INSTRUMENT

We sought to develop a survey instrument to help us understand how different types of security knowledge are associated with different types of security behaviors. By administering such a survey to a representative sample of United States Internet users, we can characterize which mental models are most common in the population, and which behaviors are commonly associated with these models.

Most existing security survey instruments measure vertically differentiated security knowledge; they assume that more-is-better and attempt to measure how much knowledge a person has. Egelman and Peer [14] is one exception, though they focus on horizontally-differentiated *behavioral intentions* rather than horizontally differentiated knowledge. However, Wash [35] found that many people differ horizontally in their beliefs, and therefore their knowledge, about computer security and that these differences lead to different behaviors. Because few of the existing methods for measuring security knowledge allow for horizontal differentiation, we developed new measures for this study.

3.1 Questions About Security Beliefs

Mental models are not traditionally measured with a survey [9]. They often include multiple types of knowledge: factual knowledge [35], counter-factual knowledge [22], knowledge about process [21], and understandings of what context is relevant [9, 35]. This presented a challenge for designing effective survey questions. We began by brainstorming a variety of survey question types based on suggestions in Dillman [11] and the findings in Wash [35]. We then conducted eight rounds of think-aloud trials with 3-5 participants each, to gain a better understanding of how our participants understood the questions we had generated. These think-alouds involved a member of the research team sitting together with a potential participant while they were taking the survey, and asking them to "think aloud" while reading and answering the questions. After these initial trials, we decided to focus on straightforward statements of *beliefs*. Beliefs form the fundamental components of a mental model, and appropriate sets of beliefs can indicate which mental model a person possesses. Statements of belief can easily be presented in a survey. We structured the questions to ask the participant to what extent they agree with each statement of belief.

Once we decided on the type of questions, we worked to develop a set of reliable questions. We identified a set of 132 beliefs that can indicate different mental models. Each belief focuses on exactly one salient aspect of one of the mental models that Wash [35] identified, and were drawn from that paper. For example, we asked participants to what extent they agree with the statement "Hackers target rich and important people." Participants who agree with this statement are likely to possess the *Big Fish* mental model, where participants who disagree with this statement are more likely to possess the *Burglar* or *Digital Graffiti Artist* models of hackers. Since Wash's [35] folk models were divided into two large categories — models about viruses and models about hackers — we divided our statements of belief into similar categories. We began with 67 belief questions about viruses and 65 belief questions about hackers. We also focused on beliefs that are independent of specific technologies (e.g. "Using anti-virus is important" rather than "Use McAfee Anti-virus") to help ensure our resulting instrument would remain valid for some time into the future.

During this analysis, we also found that many questions naturally grouped together into sets that represent high-level beliefs about hackers and viruses. For example, most people answered "Hackers target rich and important people" and "Hackers target large businesses" in very similar ways; both of these beliefs suggest a belief that hackers mostly target other people who aren't me. Rather than measuring mental models by pre-specifying an association between belief and model, we decided to directly measure beliefs and identify which beliefs were frequently held in common.

Over 100 questions is too long for a survey instrument. Following normal scale development practices [10], we identified a smaller number of questions accurately captured patterns in the larger dataset. We conducted an initial trial with 149 participants on Mechanical Turk (paying $2 per survey) to identify which questions are internally reliable and which questions cluster into common beliefs. An initial factor analysis is available in Appendix A.

We identified a total of 34 beliefs that we believed were reliable and had high construct validity. Eighteen questions concerned viruses, and these questions clustered into three sets of 6 questions in the factor analysis. Sixteen questions concerned hackers, which clustered into three sets of five questions along with an additional, standalone question. We verified the reliability of these questions with a pair of validation surveys (one for viruses, and one for hackers), each of which involved 200 participants from Mechanical Turk. Each set of questions had high reliability (Cronbach's $alpha > 0.75$), which suggests that participants answered each question in the set similarly to the way they answered the other questions in the set. Each set of questions also loaded onto a single factor in the factor analysis (loading > 0.400).

3.2 Questions About Security Behaviors

We are interested in the decisions that home computer users make and the behaviors they undertake to protect their computers. We decided to measure the frequency of past behavior rather than the intentions for future behavior. We did this because security actions often don't match up with intentions due to contextual or skill reasons. We are more concerned with what actually happened on the computer than what a person wants to happen.

Since Wash [35] found that most people separate threats from viruses and threats from hackers, we divided potential behaviors into two categories: behaviors that can protect against viruses and other malware, and behaviors that can protect computers from active attacks from hackers.

To develop a set of questions about behavior, we followed the same procedure as above: we identified 20 questions and tested them in the same initial survey on Mechanical Turk (N=149). In our trials we found extremely similar results: behaviors to protect against viruses always clustered into the same two factors (software and "be careful"), and behaviors that protect against hackers fell into the same three factors (software, "be careful", and expert behaviors). As such, we don't include our initial factor analyses here; the final factor analysis is in Appendix B.

4. MAIN SURVEY

4.1 Sample

We conducted a large scale survey with a representative sample of the US Internet-using population. Unfortunately, there is no "phone book" for the Internet from which a truly random sample can be drawn [5]. Following the guidance from the American Association for Public Opinion Research, we decided to use sample quotas to get an appropriate variety of participants for our survey that matched as closely as possible the demographics of the target population [5].

Before we did this, though, we needed to understand the demographics of Internet users in the United States. We focused on getting a representative distribution of age and education level; we believe that those two demographic factors are likely to have the largest variation in information security mental models and behaviors. We began by finding the distribution of age and education in the US population by looking at the 2010 US Census [32]. We then found the percentage of the US population with Internet access grouped by those same age and education groupings from the Pew Internet and American Life project [39]. The Pew data provides a percentage *within* a category that uses the Internet. Multiplying those together and rescaling back to 100% then provides an estimate for the percentage of Internet-using US persons that fit into each demographic category. These estimates are in Table 1.

To recruit participants, we contracted with Qualtrics to provide access to panels of Internet-using adults (>18 years old) meeting certain demographic constraints. We provided our demographic quotas to Qualtrics, and asked them to recruit a total of 2000 Internet users in the United States who met our demographic quotas. Our study was approved as minimal risk by our institutions's IRB; we never collected identifiable information about our participants. We paid Qualtrics $5 per participant, of which $1.50-$2.50 was paid to the participants in accordance with the conditions of their panel membership.

Qualtrics recruited participants to take the survey according to the quotas we specified. Two attention check questions were included in the survey, and participants who answered these questions incorrectly were removed before finishing the survey and did not count as part of the quota. Qualtrics ended up recruiting a total of 2006 participants who passed these attention checks. We removed from this sample anyone who finished the survey in less than 2 minutes, anyone who took more than 4 hours to complete the survey, and anyone who answered exactly the same answer for all Likert scale questions. After this cleaning, the final number of valid responses was 1993.

4.2 Demographics

The demographics of our sample largely reflect the demographics of the US Internet-using population, along the lines of age, education, and race. Table 1 contains the detailed demographics of our sample.

The only major deviation from the US Internet-using population in terms of demographics is gender. Internet survey

Men	749	37.6%
Women	1157	58.0%
No Gender Reported	87	4.3%
Democrat	769	38.6%
Independent	618	31.0%
Republican	595	29.9%
Some High School	245	12.3%
High School Grad	984	49.4%
College	562	28.2%
Grad School	202	10.1%
Age 18–29	420	21.1%
Age 30–49	777	39.0%
Age 50–64	552	27.7%
Age 65+	244	12.2%
No Children	796	40.0%
Has Children	1193	59.9%
White	1629	81.7%
American Indian or Alaskan Native	25	1.2%
Asian or Pacific Islander	51	2.5%
Black or African American	161	8.0%
Hispanic or Latino	94	4.7%
Other or Not Specified	33	1.7%

Table 1: Demographics of our sample.

panels are known to be skewed toward women [5], and our sample has a similar skew. It is about 58% women and only 37.6% men, which differs significantly from the population, where men and women use the Internet in approximately equal numbers [39].

To measure political party affiliation, we used the method and wording of the question in Gallup polls. We first ask whether the participant considers herself a Democrat, Independent, or Republican. The vast majority of Americans answer "Independent". Then, only for the people who answer "Independent", we also ask "As of today, do you lean more to the Democratic Party or the Republican Party?" By combining the two answers, we are able to effectively measure political leaning. Gallup polls performed around the time when this survey was run (summer 2014) showed approximately 40% of the US leans Republican and approximately 40% of the US leans Democrat [15]. Our sample slightly under-represents Republicans and over-represents independents.

We set demographic quotas on Age range and Education level for people responding to our survey. As such, those two demographics exactly match our best estimates of the US Internet-using population. We decided to enforce these quotas because we suspected that age and education would be the largest influences on security beliefs. And indeed, as shown below in Table 4, we found the most variation along those two demographics.

The question measuring race was designed to be comparable with data from the Pew Internet and American Life data [39], and the sample racial demographics of our survey approximately match the US Internet-using population, with the exception that we slightly underrepresent Hispanics.

5. SCALES

We first present the scales that we created, along with data that indicate how common each belief is in our sample of US Internet users. In addition to the scales we developed for security beliefs, we also included a number of questions asking how often participants undertake behaviors that can protect them from these threats.

5.1 Factors About Beliefs

Using data from the large representative sample, we conducted an exploratory factor analysis to confirm the presence of the same factors we found during scale development. EFA can also be used for confirmation of factors [6], and is often preferred when there is not a strong theoretical motivation for associating a set of items as factors.

One of the factors about hacker beliefs proved unreliable; it had a low Cronbach's α (< 0.70) and the factor analysis did not identify it as a factor. We dropped that factor, leaving us with three factors about virus beliefs and two factors about beliefs about hackers. Table 2 contains details of the factors. The final factor analyses is available in Appendix B.

Virus Beliefs.

Most people express a strong concern about malicious software. They usually group all malicious software under the term "virus." [35]. Beliefs about how viruses operate are likely to have an impact on the way that people make decisions to protect their computers. For example, Wash [35] found that people who believe that viruses are simply buggy software don't feel like they need to install anti-virus systems because they can simply not choose to download risky software. We identified three sets of beliefs about how viruses operate.

The first major belief about viruses is that **Viruses create visible problems**. This factor includes questions that indicate a belief that viruses infect home computers and cause a variety of problems that are mostly visible to home computer users. Most US Internet users agree with these questions; 91.4% of our participants averaged above a 3.0 on these questions.

A second major belief about viruses is that **You can protect yourself from viruses**. Viruses come from intentional choices like downloads and viewing ads, and either not downloading, or using an anti-virus software to scan downloads can prevent them. Only about 22.4% of the US Internet using population would agree with this belief, though 4.5% strongly agree with this (mean > 4.0).

The third major belief about viruses is that **Viruses are caught on the Internet**; often, there is little that can be done (other than possibly avoiding the shady parts of the Internet) to prevent them. Clicking on advertisements, downloading files, watching pornography, or simply visiting the wrong webpages can all cause you to catch a virus. Approximately 63.0% of the US Internet-using population share this belief.

Hacker Beliefs.

Many people also express a strong concern about hackers [35, 20]. "Hacker" is often a catch-all term for bad people who operate via computers and all of the associated concerns. Different beliefs about hackers have been found to influence protective behaviors. For example, Wash [35] found that people who believe hackers only steal information en

	Alpha	*Mean*	*SD*
Belief: Viruses create visible problems	**0.76**	**4.00**	**0.62**
A virus causes computers to crash		4.13	0.87
A virus causes annoying problems		4.45	0.77
A virus erases important files on the computer		3.86	0.89
A virus steals personal and/or financial information		3.92	0.93
Being aware of what websites I go to will help me avoid getting a virus		3.62	0.91
Belief: You can protect yourself from viruses	**0.80**	**2.59**	**0.82**
You can't get a virus if you keep your anti-virus software up to date		2.65	1.13
Anti-virus software always detects viruses		2.72	1.12
The only way to get a virus is by downloading something		2.38	1.11
You can't get a virus if you never download things from the Internet		2.35	1.11
Blocking pop-ups makes it very difficult to get a virus		2.87	1.00
Belief: Viruses are caught on the Internet	**0.75**	**3.40**	**0.75**
Clicking on advertisements will give you a virus		3.22	0.93
Downloads from the Internet will give you a virus		3.47	0.96
Merely visiting the wrong webpages will give you a virus		3.52	1.01
Watching pornography on the Internet will give you a virus		3.39	1.06
Behavior: Use Security Software	**0.90**	**4.32**	**0.86**
Check anti-virus software to make sure it is up to date		4.21	0.99
Regularly scan the computer with anti-virus software		4.23	1.02
Use anti-virus software		4.48	0.93
Use security software such as firewall		4.34	0.99
Behavior: Be Careful on the Internet	**0.82**	**4.36**	**0.70**
Avoid downloading anything without knowing what exactly is being downloaded		4.39	0.87
Be aware of what websites you visit		4.37	0.83
Avoid clicking on email attachments from people you do not know		4.53	0.84
Block pop-ups		4.17	0.95

	Alpha	*Mean*	*SD*
Belief: Hackers target home users	**0.83**	**3.84**	**0.69**
A hacker watches what you are doing on your computer		3.86	0.88
A hacker intentionally puts viruses on the computer		4.00	0.91
A hacker makes a record of everything on the computer		3.64	0.95
Hackers target home computer users		3.90	0.85
A hacker installs monitoring software on the computer		3.82	0.87
Belief: Hackers target others	**0.85**	**3.58**	**0.78**
Hackers target rich and important people		3.31	1.08
Hackers target large businesses		3.81	0.97
Hackers target the upper class		3.18	1.06
Hackers target banks		3.76	0.94
Hackers target large databases		3.85	.92
Behavior: Expert Security Settings	**0.82**	**3.32**	**1.08**
Disable scripting on emails		3.14	1.38
Disable scripting on websites		3.13	1.33
Back up your information on an external hard-drive, network, or server		3.41	1.32
Update patches regularly		3.58	1.31
Behavior: Be Careful on the Internet	**0.76**	**4.20**	**0.73**
Be careful downloading software from the Internet		4.34	0.88
Use good passwords (good passwords include uppercase and lowercase letters, numbers, and symbols)		4.31	0.87
Avoid clicking on attachments		4.03	0.98
Always sign out of accounts when you are done with that website		4.11	1.09
Behavior: Use Security Software	**0.71**	**4.23**	**0.90**
Use some pre-existing security software such as firewall		4.26	1.06
Scan your computer regularly with anti-virus software		4.21	0.98

Table 2: Questions and Scales. All belief items use a 'Strongly Disagree' to 'Strongly Agree' 5-point Likert scale, converted for analysis to numbers 1–5. All behavior items were phrased "How often do you do the following security precautions to avoid getting a virus? / avoid being hacked?" and use a 'Never', 'Rarely', 'Sometimes', 'Often', 'Always' scale, also converted to numbers 1-5. Virus-related questions are in the left column; hacker-related questions are on the right. The names of the factors that we assigned to each group of questions are bolded.

mass from large websites often don't take actions to protect their personal computers. From our survey, we identified two distinct sets of beliefs about hackers.

The first belief about hackers is that **Hackers target home computer users**; they break into home computers, monitor everything you do on your computer, and install viruses on your computer. 84.5% of the US Internet-using population would agree or strongly agree with this belief.

The second belief is that **Hackers target others**, mostly rich and important individuals and banks. This is a belief that hackers intentionally choose targets, and those targets are often other people with more money or power. About 71.3% of the US Internet-using population would agree with this belief.

5.2 Factors about Behavior

We were also interested in understanding what kinds of security behaviors people undertake to protect themselves. We asked how often the participant would do specific security-related behaviors that he or she could take *for the specific purpose of avoiding a virus/hacker*. We found that these actions almost always clustered into two major categories, with most people answering questions in each category very similarly.

The first cluster was behaviors that place trust in **using security software** (trust-in-software): anti-virus, firewall, and security products. Most users claimed to do this both to protect against viruses and to protect against hackers. 67.4% of the US Internet-using population would state that they use security software to protect against viruses at least "Often" (mean > 4.0 on a 1-5 scale). 58.1% would claim at least "Often" to protect against hackers.

The second cluster of behaviors place trust in oneself; they involved things that are frequently described as **be careful on the Internet** (trust-in-self): use good passwords, don't click on unknown things, block popups, and sign out of accounts when done. 69.1% of the US Internet-using population stated that they do these actions at least "Often" to protect against viruses, and 59.0% claim at least "Often" to protect against hackers.

Both of these clusters are likely influenced by *social desirability bias*: participants believe that it is socially desirable to be seen as doing these actions they are told are important, so they report doing it more often than they actually do [28]. Still, the *variation* in responses – exactly who reported doing these rarely – provides useful correlations with beliefs.

We found a third cluster of behaviors that some people used to protect against hackers. These behaviors are more advanced, **expert security settings**: disabling scripting on web pages, updating software patches, and backing up information. Many fewer US Internet-users do these behaviors, with only 24.2% of people reporting doing these "Often" or "Always". And due to the previously mentioned social desirability bias, the true number is likely to be lower.

6. RESULTS

6.1 The Relationship Between Belief and Behavior

We ran a series of regressions to better understand the relationship between the beliefs that a person has and the self-reported behaviors that they undertake to protect themselves. Table 3 contains the detailed results.

Protecting Against Viruses.

There is a relationship between the beliefs that people possess and the actions that they state they take to protect themselves. People who believe that viruses cause visible problems report taking both trust-in-software and trust-in-self actions more often. This makes sense and is good; people who see viruses as causing problems for personal computers report trying to protect themselves (reading across the second row in Table 3).

People who believe that you can protect yourself from viruses by avoiding downloads and running anti-virus software actually report lower levels of use of both trust-in-software and trust-in-self actions (row 3). The effect size of this negative relationship is smaller but still statistically significant. This is interesting, and suggests that believing that you can protect yourself actually leads to more risky behavior.

Finally, people who believe that viruses are caught simply by browsing the Internet did not show any correlation, positive or negative, with actions to protect themselves (row 4).

Protecting Against Hackers.

We found two clearly distinct sets of beliefs about hackers: that hackers target home computers, and that hackers target others. These two beliefs are not mutually exclusive – they have a 0.60 Pearson correlation – but they represent different worries about what hackers might do. This can be seen by looking at how they correlate with behaviors.

If a person believes that hackers target home computers, then they take positive actions to protect their computers. There is a positive and statistically significant relationship between this belief and all three type of actions – trust-in-software actions, trust-in-self actions, and expert actions (row 5). This makes sense; more concern about being attacked leads to more effort to protect themselves.

On the other hand, we found no relationship between a belief that hackers target others and any actions to protect computers (row 6). Our estimates are both very small and not statistically significant. The fact that this isn't negative suggests that participants don't necessarily feel safer on their computers. But rather, this belief is largely unrelated to the security precautions that people undertake when using their computers.

6.2 How Beliefs and Behaviors Vary: Demographics

Since we have a representative sample of US Internet users, we can compare beliefs about viruses and hackers across demographic groups. To do this, we calculated g_ψ, an effect size measure for each comparison of demographic groups. g_ψ is a generalization of Hedge's g designed to be used in situations where there are more than two groups to be compared. g_ψ uses as the standardizer the estimated standard deviation of the whole population; this way, all g_ψ estimates are in the same units and can be compared with each other [23].

A potentially more traditional approach to this is to directly compare means of groups, and then conduct a statistical hypothesis test for each comparison. The hypothesis tests normally accomplish two goals: they account for vari-

		V. Software	*V. Careful*	*H. Expert*	*H. Careful*	*H. Software*
1	(Intercept)	2.35***	2.62***	1.13***	2.24***	2.05***
2	Virus: Visible Problems	0.27***	0.32***	0.08	0.23***	0.27***
3	Virus: Can Protect Yourself	-0.09***	-0.10***	0.14***	-0.05*	-0.11***
4	Virus: Caught on Internet	0.06.	0.03	0.07	0.01	0.03
5	Hacker: Target Home Users	0.12**	0.06.	0.22***	0.20***	0.17***
6	Hacker: Target Others	-0.01	0.01	0.01	-0.00	0.03
7	Woman	-0.07.	0.02	-0.14**	0.01	-0.10*
8	Independent	0.05	0.01	0.14*	0.08*	0.11*
9	Republican	0.01	-0.03	0.03	0.04	0.07
10	Age 30-49	0.29***	0.16***	0.15*	0.20***	0.35***
11	Age 50-64	0.49***	0.24***	0.14.	0.32***	0.54***
12	Age 65+	0.55***	0.37***	0.15	0.28***	0.61***
13	HS Grad	0.26***	0.28***	0.34***	0.20***	0.30***
14	College	0.28***	0.22***	0.35***	0.20***	0.35***
15	Grad School	0.19*	0.25***	0.40***	0.15*	0.27**
16	Has Children	-0.04	-0.03	-0.03	-0.05	-0.11**

Table 3: Regression Results. Each column is a regression, with the dependent variable being the title of the column. The intercept represents the baseline category: man, republican, age 18-29 who didn't complete high school and doesn't have children.

ance in the underlying measurements, and they provide an indirect measure of effect size (larger effects lead to lower p-values). However, in this situation, hypothesis tests are problematic. Uncorrected, they have a problem with false positives (too many tests are determined to be statistically significant). However, using a multiple comparisons correction like Bonferroni dramatically reduces the power of the tests, which also leads to improper interpretation of results and a bias towards overestimates [16, 7]. Instead of relying on an indirect estimate of effect size, we chose to report g_ψ, which directly estimates the size of the difference. Since g_ψ normalizes by standard deviation, it also properly takes into account variance in the underlying measurements, and is more directly interpretable as a measure of the size of a difference in a population. No correction is necessary for g_ψ since there is no acceptance/rejection decision.

Tables 4 and 5 contain these results for beliefs and behaviors. For the purposes of this paper, we take an effect size larger than 0.10 to be small but worth comment, and an effect larger than 0.30 to be moderate to large [8, 23].

There appear to be almost no differences in beliefs about either viruses or hackers between Men and Women. Also, there are relatively few, and mostly small differences between people across the political spectrum. The most interesting comparison here is between Republicans and Democrats; Republicans tend to be 0.16 standard deviations lower on the belief that you can protect yourself from Viruses by using antivirus software and not downloading files.

There are some noticeable differences between people with different amounts of education. People with only high school educations generally report higher agreement with all beliefs we found, and also report that they engage in more behaviors described as being careful on the Internet (use good passwords, don't click on unknown things, etc.). But this difference is particularly large for two beliefs. People without a college education are much more likely to agree with statements that indicate that you can catch viruses by casually browsing the Internet than people with college educations. Also, people who have attended grad school are much less likely to believe that hackers target home computer users. This suggests that greater education is associated with beliefs that they are less vulnerable online.

There are some differences across age cohorts. The largest difference is for the belief that casually browsing the Internet can cause you to catch a virus. Adults 50 years old and older are much less likely to agree with this belief than younger adults. Also, adults age 30–49 are the most likely to agree that viruses cause visible problems on home computers; adults 65 and older are very unlikely to agree with that belief. Regarding behaviors, older age cohorts are more likely to report that they engage in careful behaviors to protect themselves from viruses and hackers.

There are small differences in beliefs that emerge between people who have children and people who don't. People with children are more likely to believe in threats to their own computers: that viruses cause visible problems on home computers, and that hackers target home computers.

Finally, though white Americans are the most populous racial group in the US, they have very different beliefs than other races. White American Internet users are less likely to believe viruses can be caught on the Internet, and that viruses can be protected against. However, they are most likely to report that they do behaviors that place trust in security software (anti-virus, firewall, and security products).

6.3 Grouping Participants

To better understand what kinds of beliefs happened together, we clustered participants using K-Means clustering. This clustering technique partitions participants into K groups where each cluster has a mean, or prototype, and

	Virus: Visible Problems	*Virus: Protect Yourself*	*Virus: Caught on Internet*	*Hacker: Target Home*	*Hacker: Target Others*
Woman - Man	0.02	-0.03	-0.01	0.01	-0.07
Independent - Democrat	-0.14	-0.11	-0.06	-0.09	-0.12
Republican - Democrat	-0.08	-0.16	-0.04	-0.04	-0.09
Republican - Independent	0.05	-0.06	0.01	0.05	0.02
High School Grad - Some High School	0.07	-0.16	0.06	0.05	0.06
College - High School Grad	-0.15	-0.18	-0.21	-0.10	-0.05
Grad School - College	-0.05	-0.16	-0.24	**-0.33**	-0.10
Age 30-49 - 18-29	0.19	-0.10	0.06	0.16	0.03
Age 50-64 - 30-49	-0.04	-0.17	-0.24	0.03	0.11
Age 65 or over - 50-64	-0.17	-0.01	-0.25	-0.14	-0.14
Has Children - No Children	0.10	-0.04	0.03	0.11	-0.01
American Indian or Alaska Native - White	**0.41**	**0.56**	**0.57**	0.23	0.26
Asian or Pacific Islander - White	0.08	**0.81**	**0.37**	-0.02	0.26
Black or African American - White	0.14	**0.71**	0.28	0.18	0.12
Hispanic or Latino - White	-0.00	**0.54**	0.15	0.11	-0.05

Table 4: Comparing Beliefs Across Demographic Groups. Each value is g_ψ, an estimate of the effect size of the difference, in units of standard deviation of the whole variable. g_ψ is a generalization of Hedge's g. Positive values indicate that the group on the left agrees with the belief more than the group on the right. For example, in the row "College - High School Grad", the effect size for "Virus: Caught on Internet" is -0.21. This means that High School Grads agree more that you can catch a virus simply by browsing the Internet than College graduates do. An effect size larger than 0.10 is small but worth comment, and an effect larger than 0.30 is moderate to large.

	Virus: Software	*Virus: Be Careful*	*Hacker: Expert*	*Hacker: Be Careful*	*Hacker: Software*
Woman - Man	-0.12	0.01	-0.15	-0.02	-0.16
Independent - Democrat	-0.06	-0.09	0.08	0.02	-0.01
Republican - Democrat	0.07	-0.02	0.04	0.06	0.14
Republican - Independent	0.13	0.07	-0.03	0.04	0.15
High School Grad - Some High School	0.28	**0.39**	0.25	0.23	**0.30**
College - High School Grad	0.10	0.01	-0.01	0.05	0.16
Grad School - College	-0.06	0.05	-0.05	-0.10	-0.07
Age 30-49 - 18-29	**0.38**	0.29	0.16	**0.33**	**0.44**
Age 50-64 - 30-49	0.22	0.09	-0.05	0.14	0.22
Age 65 or over - 50-64	0.05	0.13	0.04	-0.04	0.09
Has Children - No Children	0.09	0.05	0.01	0.03	0.02
American Indian or Alaska Native - White	-0.21	0.01	0.29	0.08	-0.20
Asian or Pacific Islander - White	-0.02	-0.09	-0.07	**-0.40**	-0.26
Black or African American - White	-0.22	-0.20	0.04	-0.08	**-0.37**
Hispanic or Latino - White	-0.11	-0.16	0.07	-0.07	-0.16

Table 5: Comparing Security Behaviors Across Demographic Groups. Each value is g_ψ, an estimate of the effect size of the difference, in units of standard deviation of the whole variable. g_ψ is a generalization of Hedge's g. Positive values indicate that the group on the left engages in the behavior more frequently than the group on the right. An effect size larger than 0.10 is small but worth comment, and an effect larger than 0.30 is moderate to large.

Cluster	*Virus: Visible Problems*	*Virus: Protect Yourself*	*Virus: Caught on Internet*	*Group Size*	*Virus Behavior: Software*	*Virus Behavior: Careful*
1	4.32	2.07	3.77	680	4.50	4.53
2	4.34	3.71	3.99	406	4.39	4.40
3	3.56	2.45	2.80	790	4.14	4.23

Table 6: K-Means Clustering of Participants based on their answers to questions about virus beliefs, with $K = 3$. The three beliefs (the left three columns) were clustered, and then means calculated for behaviors for each cluster (the right two columns). $K = 3$ was determined by examining a plot of variance explained and choosing the elbow.

Cluster	*Hacker: Target Home*	*Hacker: Target Others*	*Group Size*	*Hacker Behavior: Expert*	*Hacker Behavior: Careful*	*Hacker Behavior: Software*
1	3.95	3.73	792	3.26	4.22	4.30
2	4.71	4.66	368	3.72	4.51	4.53
3	2.23	1.73	43	3.48	4.10	4.22
4	3.35	2.94	673	3.15	4.03	4.01

Table 7: K-Means Clustering of Participants based on their answers about hacker beliefs, with $K = 4$. The two beliefs (the left two columns) were clustered, and then means calculated for behaviors for each cluster (the right three columns). $K = 4$ was determined by examining a plot of variance explained and choosing the elbow.

each participant is grouped into the cluster with the most similar mean across all the variables [27]. This method allows us to find common patterns of beliefs that might be non-linear in nature.

Clustering By Virus Beliefs.

To begin, we clustered all participant according to their answers to the questions about virus beliefs. This clustering technique allows us to characterize "prototype" individuals for each cluster by examining the mean value for each measure. We first need to decide how many clusters to find. By examining a plot of variance explained by number of clusters (not shown) [27], we decided that we should look for $K = 3$ clusters of participants; this would explain over 50% of the variance in virus beliefs. Table 6 contains these results.

Clusters 1 and 2 are similar in many respects; individuals in both clusters strongly agree that viruses cause visible problems on home computers, and strongly agree that viruses can be caught by browsing the Internet. However, individuals in these two clusters disagree about whether you can protect yourself by using anti-virus and not downloading files. Cluster 1 disagrees with the belief that you can protect yourself from viruses; while Cluster 2 strongly agrees with it.

Cluster 1 has the highest self-reported compliance with both using virus software and being careful about viruses on the Internet. Cluster 2 still complies with all virus protection behaviors, though less so than Cluster 1.

Cluster 3 is different; individuals in this cluster weakly agree that viruses can cause visible problems for home computers, and weakly disagree that viruses can be caught simply from browsing the Internet. Individuals in this cluster also report the lowest use of anti-virus software and the least often behavior of being careful on the Internet. This is also the largest cluster. This suggests that this mental model – believing that you can't randomly catch viruses on the Internet and only slightly believing that viruses cause visible problems on home computers – is associated with the fewest security actions.

Clustering by Hacker Beliefs.

We also clustered participants by their beliefs about hackers. Examining the plot of variance explained by number of clusters, we determined that the optimal number of clusters here would be $K = 4$. Table 7 shows the results of this clustering. Both beliefs tended to vary together, with Cluster 2 having the highest belief for both targeting home computers and targeting others, and Cluster 3 having the lowest belief in both.

Individuals in Clusters 1 and 2 agree with both beliefs about hackers. These people tended to do the most behaviors to protect themselves against hackers. Individuals in Cluster 3 disagree with both beliefs about hackers. And individuals in Cluster 4 are on the fence, neither agreeing nor disagreeing with the beliefs. The people in Cluster 4 actually report the fewest behaviors to protect themselves, and might represent a group that doesn't really think much about hackers.

7. LIMITATIONS

A limitation of this study is that we didn't measure actual actions taken by our participants; rather, we measured what participants were willing to say about their actions. These answers might not match actual actions for at least two reasons: 1) *Social Desirability Bias* [28], or 2) *Imperfect Recall.* Social desirability bias means that participants might intentionally answer incorrectly because they believe they should be taking that action, even if they aren't. They believe it is socially desirable to be seen taking that action. Imperfect recall means that participants might unintentionally answer incorrectly because they do not accurately remember their actions. This is often due to the fact that survey questions ask about general trends (e.g. "Don't click on shady links on the Internet") which are often aggregates of multiple individual events, and some of those individual events might be more salient (not clicking on a link believed to be shady) than others (clicking on a link you didn't realize was shady).

We do not see social desirability bias as a problem in our dataset. There was interesting variation in people's answers to the behavior questions, and those answers varied by belief. We suspect that which actions people see as socially desirable also depend on which folk models people believe. If you don't think hackers attack home computers, then it isn't socially desirable to protect yourself against them. Social desirability bias then works in our favor by emphasizing actions associated with a folk model and de-emphasizing actions that contrast with a folk model. This bias should strengthen our correlational results, but means our exact estimates of how many people undertake a given action might be off.

Imperfect recall is a problem in our data (and any survey). It definitely can add noise to the data; however, our large sample should allow us to distinguish signal from noise. But imperfect recall might not just be random noise; it might be biased in one direction or the other. This means that the absolute level of how much an action is taken might be incorrect, but relative measures — for example, comparing across demographic groups or clustering folk models — should still yield accurate comparisons.

Additionally, many of our questions were framed positively, which could increase the social desirability bias. We did this intentionally; we were trying to capture the horizontal differentiation of mental models expressed in Wash [35]. Mental models are often incomplete, and are not transitive; and they are sometimes self-contradictory. Even if a mental model includes a belief about a positive statement, it does not necessarily follow that the model also includes a disbelief in the equivalent negative statement. We generated questions based on statements of beliefs from Wash's [35] findings, and inverting these statements to make them negative would change their meaning. Still, this framing could possibly be why the means of some of the scales are higher than expected. However, since we are drawing comparisons between groups rather than examining absolute responses, any bias due to positivity should be approximately equivalent across groups.

In order to get more accurate measures, we intend to directly measure security actions in future work to avoid these problems with self-reported measures.

8. DISCUSSION AND CONCLUSIONS

Accurate knowledge about computer security is very hard for everyday computer users to attain, because their decisions can be hard to execute correctly [37], may not lead to correct behaviors [33], and the outcomes of their behaviors are often not visible [36]. But, using the Internet means these users still have to act. In a sense, all Internet users have experience with making computer security decisions, because they make them so often. They just rarely know for sure if they are making the "correct" ones.

Most people struggle to learn from past experiences how to protect themselves from computer security threats [36]. For example, if a user delegates security protection to antivirus, how can he or she be sure the software is doing its job correctly? Users must trust that their actions, such as clicking on one link while not clicking on another, produce positive security outcomes that are difficult to see or verify. These beliefs about actions and outcomes form a mental model about causal relationships [17]: "If I do X (use antivirus), Y outcome will result (my computer will be protected)". People incorporate lessons and analogies like "don't go to the shady parts of the Internet or you'll catch a virus" into beliefs about what causes problems [35, 29]. Many beliefs can exist side-by-side [21], are called upon when mental models relevant to the decision to be made are activated, and are associated with different behaviors. By focusing on variation in beliefs instead of amount of knowledge, we identify relationships between what everyday computer users are thinking and doing. These relationships suggest new ways to help users make better decisions beyond simply providing more knowledge.

Consistent with literature in other domains, we found a number of causal beliefs that are associated with self-reported security behaviors. Additionally, in a representative sample of the United States Internet-using population, we found that there are demographic differences in both beliefs about security, and security behaviors.

Less educated people are more likely to believe computers can catch viruses by casually browsing the Internet, but at the same time least likely to believe it is possible protect their computers from viruses and hackers. People with less than a high school degree are also least likely to report taking any kind of protective actions related to viruses or hackers. People with lower levels of general education are vulnerable because they they feel helpless, like there is nothing they can do to protect themselves.

Older people are much less likely to agree that casually browsing the Internet can give you a virus, and people with more years of education are less likely to believe hackers target home computers. Older people, and people with a high school education or greater report taking more protective actions. These people believe they can protect themselves, but often don't think that they are a target.

These results suggest an interesting relationship between demographics, beliefs and behavior: younger and less educated Internet users are vulnerable in different ways than older and more educated users. This vulnerability likely arises because of differences in their beliefs.

Differences in beliefs make communicating with and educating users about security challenging. Emphasizing vulnerability and using scare tactics is unlikely to help younger or less education users, since they often don't believe there is anything they can do about it. On the other hand, that may work for older adults, where teaching protective measures won't work because they don't believe they are a target.

An important characteristic of mental models is that many different, but related, causal beliefs can be held by a single person at the same time [35]. Our survey showed that people who strongly believe that viruses cause visible computer problems also strongly agree that viruses can be caught by browsing the Internet. But, some of these people believe they can take actions to protect themselves (n=406), while others do not (n=680). In addition, people who agree that hackers target other people instead of themselves also believe that hackers target home computers; these people do the most behaviors to protect themselves. Seeing yourself as a target isn't necessary to undertake protective actions.

The relationship between mental model and behavior is not as straightforward as Wash [35] suggests.

Interestingly, people with weak beliefs about viruses ($n = 790$) and hackers ($n = 673$) also had weak beliefs about how they should protect themselves. They also reported the lowest amount of protection behaviors. It seems like having a strong belief about cause and effect — any cause and effect — may be related to taking protective actions. Interventions intended to influence behavioral outcomes should focus on users whose causal beliefs are weakest.

More-is-better measures of security knowledge do not capture the range of beliefs that real users possess. By characterizing beliefs, we identify groups of users that have different challenges in understanding computer security. This work suggests that different demographic segments of the population are likely to respond differently to persuasive and educational messages, and a one-size-fits-all education approach is inappropriate for computer security.

9. ACKNOWLEDGMENTS

Special thanks to Alex Hinck, Michelle Rizor, and Kami Vaniea for their work on the early versions of the survey. Thanks to everyone in the BITLab for providing feedback. This material is based upon work supported by the National Science Foundation under Grant Nos. CNS-1116544 and CNS-1115926.

10. REFERENCES

[1] I. Ajzen. From Intentions to Actions: A Theory of Planned Behavior. In J. Kuhl and J. Beckmann, editors, *Action Control: From Cognition to Behavior*, pages 11–39. Springer Berlin Heidelberg, Berlin, Heidelberg, Sept. 1985.

[2] E. Albrechtsen and J. Hovden. Improving information security awareness and behaviour through dialogue, participation and collective reflection. An intervention study. *Computers & Security*, 29(4):432–445, June 2010.

[3] C. Anderson and R. Agarwal. Practicing Safe Computing: A Multimethod Empirical Examination of Home Computer User Security Behavioral Intentions. *MIS Quarterly*, 34(3), September 2010.

[4] F. Asgharpour, D. Liu, and L. Camp. Mental models of computer security risks. In *Workshop on the Economics of Information Security (WEIS)*, 2007.

[5] R. Baker, S. Blumberg, J. Brick, M. Couper, M. Courtright, J. Dennis, D. Dillman, M. Frankel, P. Garland, R. Grovers, C. Kennedy, J. Krosnick, and P. Lavrakas. Research synthesis: AAPOR report on online panels. *Public Opinion Quarterly*, 74(4):711–781, Winter 2010.

[6] D. Bandalos and S. Finney. Factor analysis: Exploratory and confirmatory. In G. R. Hancock and R. O. Mueller, editors, *The Reviewer's Guide to Quantitative Methods in the Social Sciences*, chapter 8. Routledge, 2010.

[7] K. S. Button, J. P. A. Ioannidis, C. Mokrysz, B. A. Nosek, J. Flint, E. S. J. Robinson, and M. R. Munafò. Power failure: why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience*, 14(5):365–376, May 2013.

[8] J. Cohen. *Statistical Power Analysis for the Behavioral Sciences.* Lawrence Erlbaum Associates, second edition, 1998.

[9] R. D'Andrade. *The Development of Cognitive Anthropology.* Cambridge University Press, 2005.

[10] R. F. DeVellis. *Scale Development: Theory and Applications.* SAGE Publications, Inc, third edition, June 2011.

[11] D. Dillman. *Internet, Mail, and Mixed-Mode Surveys: The Tailored Design Method.* Wiley, third edition, 2008.

[12] N. F. Doherty, L. Anastasakis, and H. Fulford. The information security policy unpacked: A critical study of the content of university policies. *International Journal of Information Management*, 29(6):449–457, Dec. 2009.

[13] P. Dourish, R. E. Grinter, J. Delgado De La Flor, and M. Joseph. Security in the wild: User strategies for managing security as an everyday, practical problem. *Personal and Ubiquitous Computing*, 8(6):391–401, 2004.

[14] S. Egelman and E. Peer. Scaling the Security Wall. In *ACM Conference on Human Factors in Computing (CHI)*, pages 1–10, Jan. 2015.

[15] Gallup. Party affiliation. `http://www.gallup.com/poll/15370/party-affiliation.aspx`, September 2014.

[16] A. Gelman and D. Weakliem. Of beauty, sex and power. *American Scientistq*, 97, 2009.

[17] E. Goldvarg and P. N. Johnson-Laird. Naive causality : a mental model theory of causal meaning and reasoning. *Cognitive Science*, 25(4):565–610, 2001.

[18] J. Gross and M. Rosson. Looking for Trouble: Understanding End-User Security Management. In *Proceedings of the 2007 Symposium on Computer Human interaction For the Management of information Technology*, pages 30–31, 2007.

[19] C. Herley. So Long , And No Thanks for the Externalities: The Rational Rejection of Security Advice by Users. In *NSPW '09 Proceedings of the 2009 New Security Paradigms Workshop*, 2009.

[20] K. Hoban, E. Rader, R. Wash, and K. Vaniea. Computer security information in stories, news articles, and education documents. In *Poster at SOUPS*, Palo Alto, CA, July 2014.

[21] P. N. Johnson-Laird. Inaugural Article: Mental models and human reasoning. *Proceedings of the National Academy of Sciences*, 2010, Oct. 2010.

[22] W. Kempton. Two Theories of Home Heat Control. *Cognitive Science*, 10(1):75–90, 1986.

[23] R. Kline. *Beyond Significance Testing.* American Psychological Association, Washington DC, 2005.

[24] N. Kumar, K. Mohan, and R. Holowczak. Locking the door but leaving the computer vulnerable: Factors inhibiting home users' adoption of software firewalls. *Decision Support Systems*, 46(1):254–264, Dec. 2008.

[25] R. LaRose, N. J. Rifon, and R. Enbody. Promoting personal responsibility for internet safety. *Communications of the ACM*, 51(3):71–76, Mar. 2008.

[26] D. Lee, R. LaRose, and N. Rifon. Keeping our network safe: a model of online protection behaviour.

Behaviour & Information Technology, 27(5):445–454, Sept. 2008.

[27] S. Lloyd. Least squares quantization in pcm. *IEEE Transactions on Information Theory*, 28(2):129–137, 1982.

[28] D. L. Paulhus. Measurement and control of response bias. In J. P. Robinson, P. R. Shaver, and L. S. Wrightsman, editors, *Measures of personality and social psychological attitudes*, chapter 2, pages 17–59. Academic Press, San Diego, CA, 1991.

[29] E. Rader, R. Wash, and B. Brooks. Stories as informal lessons about security. In *SOUPS '12: Proceedings of the Eighth Symposium on Usable Privacy and Security*. ACM, July 2012.

[30] R. Shillair, S. R. Cotten, H.-Y. S. Tsai, S. Alhabash, R. LaRose, and N. J. Rifon. Online safety begins with you and me: Convincing Internet users to protect themselves. *Computers in Human Behavior*, 48:199–207, July 2015.

[31] Symantec Corporation. Internet Security Threat Report, Volume 18, 2013.

[32] U.S. Census Bureau. Summary of population and housing characteristics, 2010 census of population and housing. `http://www.census.gov/prod/cen2010/cph-1-1.pdf`, January 2013.

[33] K. Vaniea, E. Rader, and R. Wash. Betrayed by updates: How negative experiences affect future security. In *Proceedings of the ACM Conference on Human Factors in Computing (CHI)*, Toronto, Canada, 2014.

[34] H. Varian. *Microeconomic Analysis*. W. W. Norton and Company, 1992.

[35] R. Wash. Folk models of home computer security. In *Proceedings of the Symposium on Usable Privacy and Security (SOUPS)*, Redmond, WA, July 2010.

[36] R. Wash and E. Rader. Influencing mental models of security: a research agenda. In *Proceedings of the New Security Paradigms Workshop (NSPW)*, 2011.

[37] R. Wash, E. Rader, and K. Vaniea. Out of the loop: How automated software updates cause unintended security consequences. In *Proceedings of the Symposium on Usable Privacy and Security (SOUPS)*, Palo Alto, CA, July 2014.

[38] Wikipedia. August 2014 celebrity photo leaks. `http://en.wikipedia.org/wiki/August_2014_celebrity_photo_leaks`, September 21 2014.

[39] K. Zickuhr. Who's not online and why. Pew Research Center's Internet and American Life Project, September 25 2013. `http://www.pewinternet.org/files/old-media//Files/Reports/2013/PIP_Offline%20adults_092513_PDF.pdf`.

APPENDIX

A. INITIAL TRIAL

We conducted an initial survey to evaluate questions about security beliefs. We recruited participants using Amazon's Mechanical Turk. We paid $2 per survey, and required that participants had completed at least 500 HITs and had a 90% approval rate.

Table 8 shows the results of an initial factor analysis of the questions related to beliefs about viruses. From this, we extracted 18 questions representing 3 common factors.

Table 9 shows the results of an initial factor analysis of the questions related to beliefs about hackers. From this, we extracted 16 questions. Questions *hacker.1.3*, *hacker.1.4* and *hacker.1.5* are all very similar, and we decided to keep question *hacker.1.4*. In addition to the top 5 questions for each factor, we also kept question X because we thought it might be interesting.

B. MAIN SURVEY

Tables 10, 11, 12, and 13 contain the final Exploratory Factor Analysis for the full, nationally representative sample. Participants were paid approximately $1.50 via Qualtrics for their participation in the survey.

ID	*Question*	1	2	3
virus.1.1	I closely monitor what I download from the Internet			
virus.1.2	I can tell if a website isn't safe		0.511	
virus.1.3	Downloading things from popular websites is safe			
virus.1.4	I know if I have a virus			
virus.1.5	It is extremely difficult for Macintosh computers to get viruses			
virus.1.6	I can get a virus just from going to a website			
virus.1.7	Anti-virus software always detects viruses			0.761
virus.1.8	The only way to get a virus is by downloading something			0.593
virus.1.9	A website is shady when there are a lot of pop-ups			
virus.1.10	Being aware of what websites I go to will help me avoid getting a virus		0.555	
virus.1.11	Viruses are undetected if no anti-virus software is installed			
virus.1.12	When I download something from the Internet, I probably won't get a virus			
virus.1.13	Limiting my Internet use will help me avoid getting a virus			
virus.2.1	Playing games on the Internet makes it easy to get a virus			
virus.2.2	Using an Apple computer means you can't get a virus			0.540
virus.2.3	Purchased anti-virus software is better than free anti-virus software			
virus.2.4	Not paying attention to cookies can result in getting a virus	0.524		
virus.2.5	Blocking pop-ups makes it very difficult to get a virus			0.529
virus.2.6	You can't get a virus if you never download things from the Internet			
virus.2.7	Being careful with what you click on while browsing the Internet makes it much more difficult to catch a virus			
virus.2.8	You probably won't catch a virus if you do not use the Internet frequently			
virus.2.9	Turning off your computer when you are not using it helps protect against viruses			
virus.2.10	You cannot get a virus if you keep your anti-virus software up to date			0.522
virus.3.1	Strange emails will give you a virus	0.520		
virus.3.2	Downloads from the Internet will give you a virus	0.584		
virus.3.3	Merely visiting the wrong webpages will give you a virus	0.617		
virus.3.4	A virus can be caught and spread automatically without you doing anything			
virus.3.5	Watching pornography on the Internet will give you a virus	0.695		
virus.3.6	Clicking on advertisements will give you a virus	0.602		
virus.3.7	You will get a virus if someone hacks into the your computer and installs a virus			
virus.3.8	You can get a virus if you actively click on a link on the Internet			
virus.3.9	You will catch a virus from randomly searching for things on the Internet			
virus.4.1	Causes computers to crash		0.595	
virus.4.2	Displays images such as skulls and crossbones every time the computer turns on			
virus.4.3	Causes annoying problems		0.654	
virus.4.4	Downloads pornography			
virus.4.5	Erases important files on the computer		0.609	
virus.4.6	Steals personal and/or financial information		0.606	
virus.4.7	Kicks me out of applications that are running		0.525	
	Cronbach's α	0.807	0.796	0.759
	Variance Explained	0.105	0.100	0.087
	Cumulative Variance Explained	0.105	0.205	0.291

Table 8: Exploratory Factor Analysis (using maximum likelihood factor analysis) of virus questions from initial trial of N=149 participants from Mechanical Turk. Loadings are the result of varimax rotation. Loadings < 0.5 were removed. Three factors were chosen based on the elbow of the Scree plot. We focused on the top 6 questions from each of the 3 factors for future analysis. Cronbach's α is for the top 6 questions in each factor.

ID	*Question*	1	2	3
hacker.1.1	A hacker makes a record of everything on the computer	0.663		
hacker.1.2	The hacker intentionally puts viruses on the computer	0.713		
hacker.1.3	A hacker steals personal and financial information	0.686		
hacker.1.4	A hacker sells personal information to other hackers	0.634	0.540	
hacker.1.5	A hacker works with other hackers to steal personal and financial information	0.643	0.545	
hacker.1.6	A hacker breaks stuff on the computer	0.548		
hacker.1.7	A hacker installs monitoring software on the computer	0.707		
hacker.1.8	A hacker watches what you are doing on your computer	0.697		
hacker.1.9	A hacker sells personal and financial information to criminals	0.642	0.528	
hacker.2.1	Hackers choose targets randomly			
hacker.2.2	Hackers target home computer users			
hacker.2.3	Hackers target people with weak computer security			
hacker.2.4	Hackers target large businesses			0.703
hacker.2.5	Hackers target rich and important people			0.820
hacker.2.6	Hackers are choose their victims based on exploiting immediate circumstances			
hacker.2.7	Hackers target banks			0.529
hacker.2.8	Hackers target the upper class			0.740
hacker.2.9	Hackers do not target anyone specifically			
hacker.2.10	Hackers target large databases			0.553
hacker.3.1	Hackers only target really important people; therefore, I do not need to protect myself			
hacker.3.2	Staying away from unfamiliar websites will protect me from hackers			
hacker.3.3	It is important to shut off the computer when it is not in use to avoid being hacked			
hacker.3.4	Only provide personal information to websites you trust to avoid being hacked			
hacker.3.5	Install anti-virus software to keep hackers from breaking in to the computer			
hacker.3.6	Always sign out of accounts and websites when you are done using them to avoid being hacked			
hacker.3.7	Use strong passwords (includes numbers, symbols, and upper and lowercase letters) to prevent hackers from breaking in			
hacker.3.8	Preinstalled firewall prevents hackers from breaking in			
hacker.3.9	There is no way to protect myself from being hacked			
hacker.3.10	Don't check your bank account online to prevent being hacked			
hacker.3.11	Shop in stores instead of online to avoid being hacked			
hacker.4.1	Hackers are college-age technology-savvy students			
hacker.4.2	Anyone can be a hacker			
hacker.4.3	Hackers are lonely college students			
hacker.4.4	Hackers are professional criminals		0.630	
hacker.4.5	Hackers are members of organized crime		0.657	
hacker.4.6	Hackers are a type of criminal		0.726	
hacker.4.7	Hackers work with other criminals		0.866	
hacker.4.8	Hackers have no morals			
hacker.4.9	Hackers are mischevious			
	Cronbach's α	0.863	0.851	0.847
	Variance Explained	0.128	0.118	0.077
	Cumulative Variance Explained	0.128	0.246	0.323

Table 9: Exploratory Factor Analysis (using maximum likelihood factor analysis) of hacker questions from initial trial of N=149 participants from Mechanical Turk. Loadings are the result of varimax rotation. Loadings < 0.5 were removed. Three factors were chosen based on the elbow of the Scree plot. We focused on the top 5 questions from each of the 3 factors for future analysis. Cronbach's α is for the top 5 questions in each factor.

ID	*Question*	1	2	3
virus.1	Watching pornography on the Internet will give you a virus			0.504
virus.2	Merely visiting the wrong webpages will give you a virus			0.539
virus.3	Clicking on advertisements will give you a virus			0.649
virus.4	Downloads from the Internet will give you a virus			0.562
virus.5	Not paying attention to cookies can result in getting a virus			
virus.6	Strange emails will give you a virus	0.476		0.454
virus.7	Anti-virus software always detects viruses		0.722	
virus.8	The only way to get a virus is by downloading something		0.663	
virus.9	Using an Apple computer means you can't get a virus		0.530	
virus.10	Blocking pop-ups makes it very difficult to get a virus		0.556	
virus.11	You can't get a virus if you keep your anti-virus software up to date		0.756	
virus.12	You can't get a virus if you never download things from the Internet		0.626	
virus.13	A virus causes annoying problems	0.618		
virus.14	A virus causes computers to crash	0.687		
virus.15	A virus erases important files on the computer	0.577		
virus.16	A virus steals personal and/or financial information	0.544		
virus.17	Being aware of what websites I go to will help me avoid getting a virus	0.482		
virus.18	A virus kicks me out of applications that are running	0.484		
	Variance Explained	0.155	0.151	0.109
	Cumulative Variance Explained	0.155	0.307	0.416

Table 10: Exploratory Factor Analysis (using maximum likelihood factor analysis) of virus questions from the full survey of N=1993 participants sampled via Qualtrics. Loadings are the result of varimax rotation. Loadings < 0.4 were removed. EFA was used in a confirmatory manner, and three factors were chosen based the previous trials. Question 6 loaded on multiple factors and was removed. Question 5 did not load on any factors and was removed. We constructed scales out of up to 5 questions (highest loaded) for each factor.

ID	*Question*	1	2
hacker.1	Hackers work with other criminals	0.504	
hacker.2	A hacker intentionally puts viruses on the computer	0.656	
hacker.3	Hackers are members of organized crime		
hacker.4	Hackers are professional criminals	0.547	
hacker.5	Hackers have no morals	0.599	
hacker.6	Hackers are mischievous	0.545	
hacker.7	A hacker watches what you are doing on your computer	0.680	
hacker.8	A hacker makes a record of everything on the computer	0.628	
hacker.9	A hacker installs monitoring software on the computer	0.605	
hacker.10	A hacker breaks stuff on the computer		
hacker.11	Hackers target home computer users	0.612	
hacker.12	Hackers target rich and important people		0.717
hacker.13	Hackers target large businesses		0.706
hacker.14	Hackers target the upper class		0.687
hacker.15	Hackers target banks		0.617
hacker.16	Hackers target large databases		0.606
	Variance Explained	0.255	0.197
	Cumulative Variance Explained	0.255	0.452

Table 11: Exploratory Factor Analysis (using maximum likelihood factor analysis) of hacker questions from the full survey of N=1993 participants sampled via Qualtrics. Loadings are the result of varimax rotation. Loadings < 0.5 were removed. EFA was used in a confirmatory manner. Three factors were originally extracted, but many indicators suggested a poor fit. Instead, we extracted two factors. We constructed scales out of up to 5 questions (highest loaded) for each factor.

ID	*Question*	1	2
virus.prevent.1	Check anti-virus software to make sure it is up to date	0.816	
virus.prevent.2	Regularly scan the computer with anti-virus software	0.797	
virus.prevent.3	Use anti-virus software	0.751	
virus.prevent.4	Use security software such as firewall	0.637	
virus.prevent.5	Avoid downloading anything without knowing what exactly is being downloaded		0.728
virus.prevent.6	Be aware of what websites you visit		0.726
virus.prevent.7	Avoid clicking on email attachments from people you do not know		0.684
virus.prevent.8	Block pop-ups		0.500
	Variance Explained	0.337	0.288
	Cumulative Variance Explained	0.337	0.625

Table 12: Exploratory Factor Analysis (using maximum likelihood factor analysis) of questions about protection behaviors from viruses from the full survey of N=1993 participants sampled via Qualtrics. Loadings are the result of varimax rotation. Loadings < 0.4 were removed. EFA was used in a confirmatory manner, and two factors were chosen based the previous trials.

ID	*Question*	1	2	3
hacker.prevent.1	Use some pre-existing security software such as firewall			0.613
hacker.prevent.2	Disable scripting on emails	0.865		
hacker.prevent.3	Back up your information on an external hard-drive, network, or server	0.465		
hacker.prevent.4	Scan your computer regularly with anti-virus software			0.631
hacker.prevent.5	Avoid clicking on attachments		0.555	
hacker.prevent.6	Be careful downloading software from the Internet		0.745	
hacker.prevent.7	Disable scripting on websites	0.837		
hacker.prevent.8	Update patches regularly	0.454		
hacker.prevent.9	Always sign out of accounts when you are done with that website		0.454	
hacker.prevent.10	Use good passwords (good passwords include uppercase and lowercase letters, numbers, and symbols)		0.579	
	Variance Explained	0.213	0.187	0.140
	Cumulative Variance Explained	0.213	0.400	0.539

Table 13: Exploratory Factor Analysis (using maximum likelihood factor analysis) of questions about protection behaviors from hackers from the full survey of N=1993 participants sampled via Qualtrics. Loadings are the result of varimax rotation. Loadings < 0.4 were removed. EFA was used in a confirmatory manner, and three factors were chosen based the previous trials.

“...no one can hack my mind”: Comparing Expert and Non-Expert Security Practices

Iulia Ion
Google
iuliaion@google.com

Rob Reeder
Google
rreeder@google.com

Sunny Consolvo
Google
sconsolvo@google.com

ABSTRACT

The state of advice given to people today on how to stay safe online has plenty of room for improvement. Too many things are asked of them, which may be unrealistic, time consuming, or not really worth the effort. To improve the security advice, our community must find out what practices people use and what recommendations, if messaged well, are likely to bring the highest benefit while being realistic to ask of people. In this paper, we present the results of a study which aims to identify which practices people do that they consider most important at protecting their security online. We compare self-reported security practices of non-experts to those of security experts (i.e., participants who reported having five or more years of experience working in computer security). We report on the results of two online surveys—one with 231 security experts and one with 294 MTurk participants—on what the practices and attitudes of each group are. Our findings show a discrepancy between the security practices that experts and non-experts report taking. For instance, while experts most frequently report installing software updates, using two-factor authentication and using a password manager to stay safe online, non-experts report using antivirus software, visiting only known websites, and changing passwords frequently.

1. INTRODUCTION

Frightening stories about cybersecurity incidents abound. The theft of millions of credit card numbers from a retail chain [10], a billion passwords from various websites [25], and a large set of nude celebrity photos [24] are just a few examples of stories that have been in the news lately.

In response to such security incidents, thousands of online articles and blog entries advise users what to do to stay safe online. Advice ranges from choosing a strong password [27] and having good security questions [38] to making email addresses unguessable [7] and entirely disabling photo backups in the cloud [27]. Besides such incident-related articles, many service providers, enterprises, and universities offer tips and training on how to stay safe online [2, 3, 17, 35].

If one hour of time from all US users is worth $2.5 billion [19], carefully considering the most worth-while advice to recommend is imperative. Even if users accept some responsibility for protecting their data [23, 43] and want to put in some effort [41], we should be thoughtful about what we ask them to do [20] and only offer advice that is effective and realistic to be followed.

Existing literature on giving good advice suggests that for recipients to follow it, the advice should be (a) useful, comprehensible and relevant, (b) effective at addressing the problem, (c) likely to be accomplished by the recipient, and (d) not possess too many limitations and drawbacks [34]. Therefore, to improve the state of security advice, we must assess which actions are most likely to be effective at protecting users, understand what users are likely and willing to do, and identify the potential challenges or inconveniences caused by following the advice. Furthermore, lessons from health advice in outreach interventions suggest that people will not initiate certain actions if they do not believe them to be effective [53]. Therefore, to learn how to best deliver the advice to users, we must also understand how users perceive its effectiveness and limitations.

In preliminary work, we surveyed security experts to identify what advice they would give non-tech-savvy users. The most frequently given pieces of advice were, in order of frequency: (1) keep systems and software up-to-date, (2) use unique passwords, (3) use strong passwords, (4) use two-factor authentication, (5) use antivirus software, and (6) use a password manager. In this paper, we report on results of a study which tries to identify what security advice users currently follow and how their attitudes and practices differ from those of security experts. To this end, we conducted a survey with 294 participants recruited from Amazon’s Mechanical Turk crowdsourcing platform and another with 231 security experts recruited through an online blog. Our results help inform what important security advice users aren’t following.

Our results show that expert participants considered keeping the operating system and applications up-to-date, using strong and unique passwords, turning on two-factor authentication, and using a password manager the most important things they do to stay safe online. Non-expert participants, however, considered using antivirus software, using strong passwords, changing passwords frequently, and visiting only trusted websites to be very effective, but admitted to delaying installation of software updates and expressed some lack of trust in password managers. We found that generally experts’ security practices matched the advice they would give non-tech-savvy users, with a few exceptions. Experts recommended not clicking on links or opening emails from unknown people, yet they reported to do so at a higher rate than non-experts reported. Other security practices that non-experts considered very important, such as visiting only known websites, were not being followed by experts nor were they considered good security advice by experts.

Symposium on Usable Privacy and Security (SOUPS) 2015, July 22–24, 2015, Ottawa, Canada.

Our findings can help inform better security advice that might actually be followed and design campaigns to improve security education. Security practices that experts follow and consider good security advice for non-tech-savvy users, but that non-experts do not yet follow, are good candidates to be recommended to non-tech-savvy users.

2. RELATED WORK

There has been a good deal of past work that investigated specific areas of security-related behavior and others that, like ours, have considered security-related practices and behavior generally.

We cover related work on general security-related attitudes and behavior, and work that focuses specifically on the four primary areas we cover later: updates, antivirus software (and more generally, malware protection), account security, and mindfulness.

2.1 Security-related behavior in general

Some past work has provided study and commentary of user attitudes or behaviors toward security in general. Wash interviewed non-expert users to elicit common mental models about security and showed how these various mental models lead to compliance or non-compliance with various forms of common security advice [52]. Herley has commented at length on the overall state of security advice, arguing that users may often fail to follow it for rational reasons [19], and that there are currently so many security-related demands on users that adding more would be counterproductive [20]. Adams and Sasse [4] were amongst the first to show that users often work around security requirements and some security practices that experts recommend. Similarly, Beautement et al. [6] note that users are often knowledgeable about good security practices and are willing to make some efforts toward complying with them, but that there are limits to how much effort they can or will exert. Howe et al. [21] provide an extensive review of work on home users' security-related behavior.

Some prior work has covered the communications angle of security advice, arguing that perhaps users would be more likely to comply with security advice if it were communicated differently, or at least more effectively. Stewart and Lacey [47] argue for more effective ways to communicate security advice. Camp [9] describes a number of common user mental models about security that might be leveraged to better explain security advice to users. Rader et al. [39] show that stories about others' security-related experiences are a common means by which users learn about security practices, so stories may be a good way to communicate advice to users.

2.2 Security-related behavior in specific areas

In this section, we focus on four top areas of security-related behavior: updates, antivirus software, account security, and mindfulness.

2.2.1 Updates

Vaniea et al. [50] identified some of the reasons users often don't install security updates. They found three main reasons why participants in their study did not install updates: participants found security updates often bundled with other undesirable features, they had difficulty assessing the value of an update, and they were confused about why updates were needed. Vaniea et al.'s work follows on that of Khan et al. [28], who showed that users do not consistently update all their systems and application software in a timely manner.

2.2.2 Antivirus software

Levesque et al. [32] gave instrumented laptops to 50 people to show how user behavior is correlated with malware infections. They note that antivirus effectiveness and vulnerability to malware infections are dependent on user behavior in various respects, from whether users install antivirus software, to how users configure their antivirus software, to what websites they visit. They found that 38% of the participants in their study were exposed to malware that antivirus software cleaned, so they demonstrate that the behavior to install and configure antivirus software can actually make users more secure.

2.2.3 Account security

There has been a great deal of work studying users' selection and use of passwords. Kelley et al. analyzed a set of 12,000 passwords collected from sites with different password strength and composition policies [26]. The authors evaluated the resistence of passwords created under different policies to guessing attacks. Shay et al. evaluated eight password composition policies with the help of 8,143 online participants and a password cracking algorithm. The authors found that longer passwords are more usable than those containing a mix of character classes and, in some cases, more secure as well [44]. Other work has investigated the effect of strength meters on password creation [15, 30, 49].

Hayashi and Hong collected 1,500 password typing events in a diary study with 20 participants. The authors collected data on where participants logged in and how frequently they did so from computers they did not own [18]. Similarly, Inglesant and Sasse had 32 people in 2 organizations, a university and a financial company, keep a diary of their password use. They found that these users were motivated to be secure, but struggled to change passwords, to create new passwords, and to comply with password policies [22]. Florencio and Herley conducted a large scale study in which they monitored participants' password habits through specially designed software running on participants' machines [16]. The authors found that users have a set of passwords which they used on multiple sites. Participants sometimes used trial and error to remember which password goes with which website. Das et al. estimated through a user study and by analyzing hundreds of thousands of leaked passwords from different websites that 43 to 51% of users reuse passwords [12]. Furthermore, users apply a few basic transformations to existing passwords before using them on different sites, which makes it possible for an attacker to guess such transformed passwords.

Chiasson et al. evaluated the usability of two proposed password managers [11]. The authors found that users had incomplete or incorrect mental models of the software. Furthermore, users were not convinced that using a password manager would bring them significant security benefits and were reluctant to give up control over their passwords to a piece of software.

2.2.4 Mindfulness

Actions that we categorize as mindfulness, including practices such as visiting only known websites, checking for HTTPS indicators, and email habits, are typically aimed at guarding against phishing, malware, and man-in-the-middle attacks. A number of past works have covered user behavior related to preventing these attacks.

Early work on phishing awareness and prevention includes Dhamija et al. [13], Wu et al. [54], and Egelman et al. [14]. These works showed that participants had difficulty telling phishing sites from their legitimate counterparts, largely because participants looked at the wrong indicators for legitimacy. At the time, browsers usually had hard-to-notice indicators of possible phishing attempts, and non-blocking phishing warnings that were easy to miss or ignore. Sheng et al. [45] studied susceptibility to phishing attacks through an online role-playing-scenario study of 1001 participants. They found that participants would click on around half of phishing links presented in the role-playing scenario.

HTTPS indicators range from lock icons in browsers and the URL shown in browser address bars to full-page, blocking interstitial warnings. Early studies of user behavior related to SSL indicators and warnings includes Sunshine et al. [48], Sotirakopoulos et al. [46], and Schechter et al. [42]. These papers presented results from lab user studies that showed some of the faults of early browser warnings. More recently, Akhawe and Felt [5] presented telemetry data from millions of real-world browsers showing that warnings work for many users in many situations, but that large percentages of Firefox and Chrome users still proceed through SSL warnings. Lin et al. [33] studied how highlighting the domain over other URL elements can help some users better identify what websites they visit.

We add to the existing body of knowledge an analysis of user behavior and attitudes accross all of these areas of security advice. We compare experts and non-experts and identify how the security practices of each group differ.

3. METHODOLOGY

We gathered data through two online surveys: one of security experts and one of non-security-expert Internet users. To help develop our expert survey, we started with a set of interviews. We describe these next.

3.1 Expert Interviews (N=40)

To design the surveys, we first conducted in-person semi-structured interviews of 40 security experts at the 2013 BlackHat, DefCon, and USENIX Security conferences. We defined experts as conference attendees who reported having at least 5 years of experience working in or studying computer security. We started every interview with our *top-3-advice question*:

> *What are the top 3 pieces of advice you would give to*
> *a non-tech-savvy user to protect their security online?*

We asked follow-up questions to clarify responses. Interviews lasted 8 minutes on average. We transcribed all of the interviews and coded the advice we collected. Interview data informed many of the questions we asked in the surveys, as we note below. In this paper, we report on the data we received from the surveys, but we also include a small number of interesting quotes from the interviews to help ilustrate some points.

3.2 Expert Survey (N=231)

Following our preliminary interviews, we conducted a survey with security experts in February to April 2014. The "expert survey" allowed us to gather data from a larger number of experts than we could through interviews, to gather quantitative data about some of the advice we heard in the interviews, and to inquire about participants' security practices. The survey was written and administered using Google Forms.

3.2.1 Expert Survey Participants

Our expert survey data is based on 231 responses from non-compensated volunteer security experts. As with the interviews, we defined a "security expert" as a survey participant who reported having at least 5 years of experience working in or studying computer security. Participants who did not meet the criteria were eliminated from further analysis.

We recruited participants through a post on the Google Online Security Blog [40] and a request to colleagues to spread a link to the survey via their social media accounts. About 80% of participants were recruited via the blog entry and about 20% via social media (the vast majority of survey responses were received in the days following the blog post, which occurred weeks after the social media effort).

Of the 231 participants who met our expert criteria, 4% were female. Ages ranged from 18 to over 65, with 30% in the 25-34 year-old range, 32% in the 35-44 range, and 18% in the 45-54 range. While 47% of participants were from the United States, others were from 25 countries around the world, including Australia, Germany, India, Israel, Japan, South Africa, and the UK. Participants held a vast range of job titles within computer security including CEO, Chief Information Security Officer (CISO), consultant, grad student, IT specialist, network administrator, security researcher, software engineer, and whitehat hacker. 73% of the sample held a Bachelor's degree or higher. In a check-all-that-apply question, 69% reported working in industry, 15% in academia, 13% in self-employment, 11% in government, and 7% in corporate research labs.

3.2.2 Expert Survey Content

The expert survey asked the same open-ended top-3-advice question with which we started the earlier interviews, then asked another open-ended question about what they actually did, the *things-you-do question*:

> *What are the 3 most important things you do to protect*
> *your security online?*

It went on to ask 34 fixed-response questions, 1 branching question, 4 quality-assurance questions, and 8 demographic questions. The 34 fixed-response questions were developed using advice from the interviews and were divided into two sections. The first section asked 14 questions about whether participants followed a set of 14 pieces of advice commonly mentioned in the interviews. An example question in this set was: *Do you use two-factor authentication (e.g., 2-step verification) for at least one of your online accounts?*

The second section asked experts to rate the "goodness" of 20 pieces of advice we heard frequently in the interviews (e.g., *Use two-factor authentication for online accounts*). In rating goodness, experts were asked to consider both how effective the advice was at keeping the user secure and how realistic it was that users could follow it.

The branching question asked whether the participant owned a personal computer and gated 2 of the 14 behavioral questions; in results, when we report a number of expert participants under 231, it's because some answered "No" to the branching question, so they did not encounter these 2 behavioral questions.

Quality-assurance questions had an answer we considered obviously correct, and we eliminated participants who answered more than one incorrectly from further analysis. An example is: *Pay attention when taking online surveys. We appreciate your input. To let us know you're paying attention, select four for this response.* (We allowed one incorrect response because all participants offering only one incorrect response to a quality-assurance question provided otherwise thoughtful answers to open-ended questions. We piloted the survey with security experts from our organization.

3.2.3 Limitations

Our recruiting methods may have produced a sample with some bias relative to the overall population of security experts, but our sample was large and diverse, so it likely represents a substantial portion of the security expert community. Since most participants came from the Google blog, some readers of the blog may be favorably predisposed toward Google and its products.

3.3 Non-expert Survey (N=294)

To get the non-security-expert perspective on security behaviors, we conducted another survey with non-experts whom we recruited via Amazon Mechanical Turk (MTurk). Like the expert survey, the non-expert survey was written and administered using Google Forms. The "non-expert survey content" was nearly identical to the expert survey content, with a few exceptions noted below.

3.3.1 Non-expert Survey Participants

Non-expert survey participants responded to our task description on MTurk, calling for participation in a Google study about Internet use. Participants were compensated $1 each for completing the survey. We required that MTurk participants be located in the United States, have a task approval rate of 95% or better, and have completed at least 500 tasks.

According to responses to demographic questions in our survey, our non-expert sample was 40% female. Ages ranged from 18 to over 65, with 50% of the sample in the 25-34 age range, and 19% each in the 18-24 and 35-44 age ranges. Participants held a wide range of occupations including artist, cashier, farmer, homemaker, sales, and youth advisor. Educational range was wide, with 47% of participants holding a Bachelor's degree or higher.

3.3.2 Non-expert Survey Content

The non-expert survey started by asking the things-you-do question. The non-expert survey also asked 54 fixed-response questions, the same branching and quality-assurance questions from the expert survey, and 5 demographic questions. We piloted the survey with 20 participants (whose data is excluded from our analysis) from Mechanical Turk.

To assess if poor advice adoption among non-experts stems from a lack of understanding of the security benefits that the advice brings or from other factors altogether, we asked non-experts questions on the perceived effectiveness of different pieces of advice and their likelihood to follow the advice.

We eliminated 6 non-expert participants who answered one quality-assurance question incorrectly from further analysis.

3.3.3 Mechanical Turk as a Recruiting Platform

Our non-expert survey used MTurk, which is sometimes called into question as a recruiting platform for studies. Concerns include whether demographics of its participants are biased relative to the Internet-using population at large, and whether remote participants will provide quality data. MTurk has already been used in prior usable security research including [15, 43], and in other usability work has been found to yield quality results and populations more diverse than typical university samples [8, 29, 37]. Although MTurk is becoming a generally accepted platform for recruiting user study participants, we included quality-assurance questions to filter out any MTurk participants who may be have been answering all required questions as quickly as possible and providing junk data in the process. Only 6 participants were filtered out for providing incorrect quality-control responses.

3.4 Coding open-ended responses

Two raters coded the open-ended responses. The raters read the responses and consulted to develop a codebook of distinct pieces of advice, then assigned codes to each open-ended response to the things-you-do question. The raters achieved a Cohen's κ, a measure of inter-rater reliability, of 0.77—a value generally considered substantial agreement [31, 51].

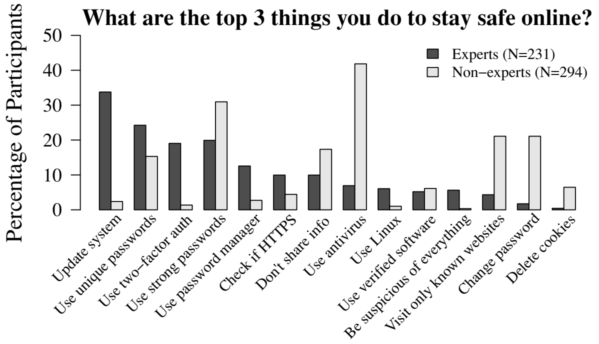


Figure 1: Security measures mentioned by at least 5% of each group. While most experts said they keep their system updated and use two-factor authentication to stay safe online, non-experts emphasized using antivirus software and using strong passwords.

4. RESULTS

Figure 1 shows all security measures that were mentioned by at least 5% of experts or by 5% of non-experts in response to the open-ended things-you-do question. The most common things-you-do responses from each group varied, with only one practice, using strong passwords, in common within each group's top 5 responses. While most experts said they install software updates (35%), use unique passwords (25%), use two-factor authentication (20%), use strong passwords (19%), and use a password manager (12%), non-experts mentioned using antivirus software (42%), using strong passwords (31%), changing passwords frequently (21%), visiting only known websites (21%), and not sharing personal information (17%).

Note that we've chosen to visualize and discuss only security measures mentioned by at least 5% of experts or non-experts as a matter of convenience for presenting our results and due to space constraints; in fact, we collected a long list of security measures that were each mentioned by only a few respondents.

Figure 2 depicts the percentage difference between the groups. The practices mentioned least by non-experts relative to experts were: (1) keep your system up-to-date (31%), (2) use two-factor authentication (18%), and (3) use a password manager (10%).

The security practices mentioned by experts are consistent with experts' rating of different pieces of advice, when we asked them to rank how *good* these are on a 5-point Likert scale. As shown in Figure 3, most experts considered installing OS (65%) and application (55%) updates, using unique (49%) and strong (48%) passwords, using a password manager (48%), and using two-factor authentication (47%) *very good* advice (the highest Likert-scale rating). Other advice that was not frequently mentioned by experts in the top three things they do, but ranked high in this multiple choice question of the advice they'd consider good, included turning on automatic updates (72%), being suspicious of links (60%), not entering passwords on links in emails (60%), and not opening email attachments from unknown people (55%).

In the following, we present and compare expert and non-expert practices and attitudes. We focus on the security practices most mentioned by experts and non-experts. We group these into soft-

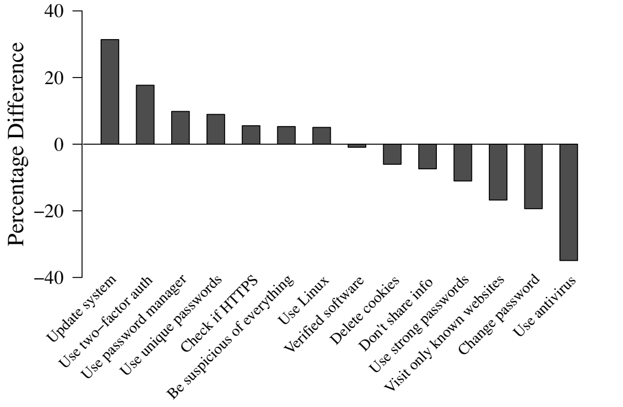


Figure 2: Percentage difference of experts and non-experts mentioning these security practices when asked what are the top three things they do to stay safe online. Security measures with a positive percentage difference were mentioned more by experts than non-experts; those with a negative percentage difference were mentioned more by non-experts.

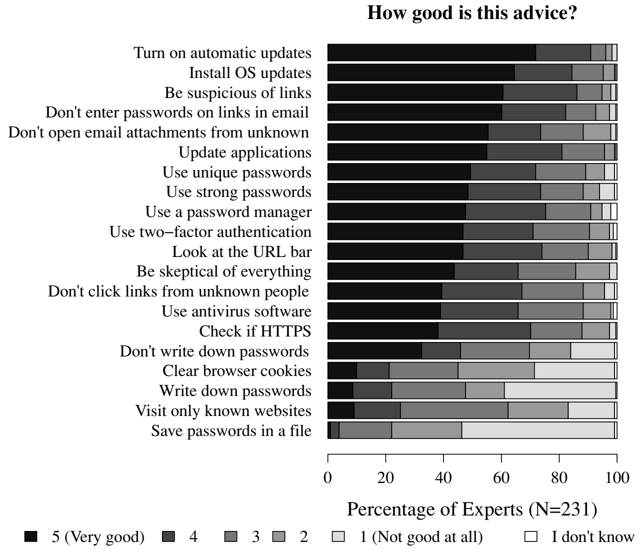


Figure 3: Advice considered "good" (i.e., both in terms of effective and realistic) by experts.

ware updates, antivirus software, password management, and mindfulness. In presenting our results, we draw upon participants' responses to both open-ended and fixed-response questions in the survey.

We refer to expert participants as E1, E2,... E231, and non-experts as N1, N2,... N294. We focus on data collected in the online surveys, but also include some quotes from the expert interviews. We explicitly state when a quote was collected during the interviews. The p values we report refer to Chi-Squared tests and are corrected for multiple tests using the Holm-Bonferroni method. We applied the Holm-Bonferroni correction in R for all the tests we conducted. R adjusts each p-value, rather than reducing alpha (though both techniques are equivalent), so we stuck with alpha=0.05 for determining statistical significance.

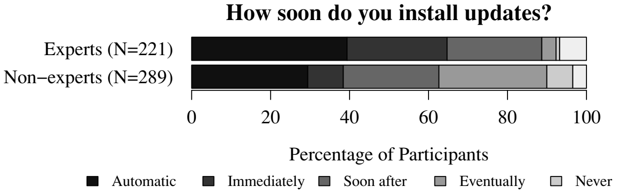


Figure 4: More experts than non-experts reported installing software updates in a timely manner.

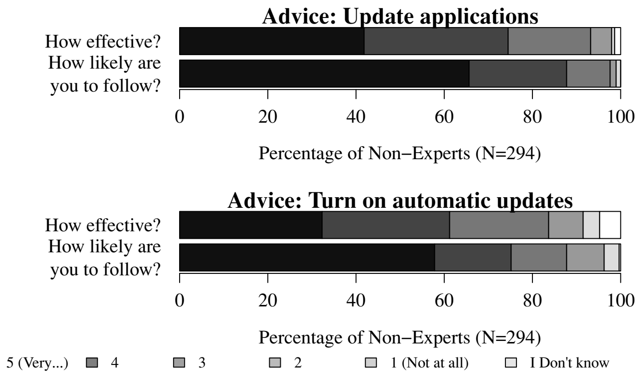


Figure 5: Most non-experts considered the advice to install software updates not effective as a security measure, but said they would be likely to follow it if they heard it was effective.

4.1 Install Software Updates

When asked for the top three things they do to stay safe online, the most common reponse from experts was installing software updates. For instance, E128 said: *"Update all the software and firmware to fix any possible vulnerability."* Furthermore, E78 also said: *"Patch, patch, patch."* Installing updates was also the security measure with the highest percentage difference between experts and non-experts; it was mentioned by 35% of experts, but only by 2% of non-experts. In addition, 2% of experts said they turn on automatic updates—an action that no non-expert mentioned.

To investigate whether the difference in the number of experts and non-experts mentioning updates is also reflected in reported behavior, not just attitudes, we asked both groups in a multiple-choice question how soon after they discover a new version of their operating system is available they install it. Consistent with the previous finding, experts reported installing updates in a more timely manner than non-experts. As shown in Figure 4, 39% of experts—but only 29% of non-experts—answered "Updates are automatically installed." In addition, 25% of experts versus 9% of non-experts said that updates are installed "Immediately." The differences are statistically significant, as summarized in Table 2. Note, however, that these questions did not differentiate between major OS releases and patches. The exact question is included with the survey instrument in the appendix.

We found a similar result for the advice to "Update applications", which only 42% of non-experts considered very effective, yet 66% said they were very likely to follow it (see Figure 5). Table 2 summarizes non-experts rating of effectiveness and likelihood to follow for this and other pieces of advice.

Reported Behavior (Experts & Non-Experts)	Chi-Square Result
How soon do you install updates?	$\chi^2(4, N_e = 221, N_n = 289) = 75.78, p < 0.001$
Do you use antivirus software?	$\chi^2(1, N_e = 221, N_n = 289) = 31.44, p < 0.001$
Do you use two-factor authentication?	$\chi^2(1, N_e = 231, N_n = 294) = 23.37, p < 0.001$
Do you remember you passwords?	$\chi^2(3, N_e = 231, N_n = 294) = 94.68, p < 0.001$
Do you write down your passwords?	$\chi^2(3, N_e = 231, N_n = 294) = 24.78, p < 0.001$
Do you save your passwords in a file?	$\chi^2(3, N_e = 231, N_n = 294) = 1.68, p = 1$
Do you use a password manager?	$\chi^2(3, N_e = 231, N_n = 294) = 131.31, p < 0.001$
Do you reuse passwords?	$\chi^2(3, N_e = 231, N_n = 294) = 37.25, p < 0.001$
Do you look at the URL bar?	$\chi^2(3, N_e = 231, N_n = 294) = 56.29,p < 0.001$
Do you check if HTTPS?	$\chi^2(3, N_e = 231, N_n = 294) = 132.62, p < 0.001$
Do you visit websites you have not heard of?	$\chi^2(3, N_e = 231, N_n = 294) = 62.84, p < 0.001$
Do you enter your password on links in emails?	$\chi^2(3, N_e = 231, N_n = 294) = 2.06, p = 1$
Do you open emails from unknown senders?	$\chi^2(3, N_e = 231, N_n = 294) = 99.60, p < 0.001$
Do you click on links from unknown people?	$\chi^2(3, N_e = 231, N_n = 294) = 51.37, p < 0.001$

Table 1: Results of χ^2 tests comparing expert and non-expert reports on their security behavior. p-values are corrected for multiple testing using the Holm-Bonferroni method.

Security Advice	How Effective is this Advice?	How Likely are You to Follow?	Chi-Square Result
Use antivirus	$\mu = 4.57, \sigma = 0.76$	$\mu = 4.67, \sigma = 0.80$	$\chi^2(5, N = 294) = 15.40, p = 0.12$
Install latest OS updates	$\mu = 4.14, \sigma = 0.94$	$\mu = 4.35, \sigma = 1.03$	$\chi^2(5, N = 294) = 32.38, p < 0.001$
Turn on automatic updates	$\mu = 3.82, \sigma = 1.11$	$\mu = 4.18, \sigma = 1.15$	$\chi^2(5, N = 294) = 49.29, p < 0.001$
Update applications	$\mu = 4.12, \sigma = 0.93$	$\mu = 4.5, \sigma = 0.80$	$\chi^2(5, N = 294) = 39.28, p < 0.001$
Clear cookies	$\mu = 3.6, \sigma = 1.22$	$\mu = 4.21, \sigma = 1.15$	$\chi^2(5, N = 294) = 62.48, p < 0.001$
Use unique passwords	$\mu = 4.58, \sigma = 0.78$	$\mu = 4.30, \sigma = 1.02$	$\chi^2(5, N = 294) = 20.39, p = 0.01$
Use strong passwords	$\mu = 4.61, \sigma = 0.80$	$\mu = 4.63, \sigma = 0.77$	$\chi^2(5, N = 294) = 7.10, p = 1$
Don't write down passwords	$\mu = 3.58, \sigma = 1.54$	$\mu = 3.78, \sigma = 1.55$	$\chi^2(5, N = 294) = 14.16, p = 0.17$
Save passwords in a file	$\mu = 1.75, \sigma = 1.08$	$\mu = 2.15, \sigma = 1.45$	$\chi^2(5, N = 294) = 21.01, p = 0.01$
Use a password manager	$\mu = 2.89, \sigma = 1.46$	$\mu = 2.98, \sigma = 1.60$	$\chi^2(5, N = 294) = 20.03, p = 0.02$
Write down passwords	$\mu = 2.31, \sigma = 1.51$	$\mu = 2.61, \sigma = 1.61$	$\chi^2(5, N = 294) = 14.36, p = 0.17$
Check if HTTPS	$\mu = 4.29, \sigma = 0.89$	$\mu = 4.38, \sigma = 0.93$	$\chi^2(5, N = 294) = 12.57, p = 0.30$
Be skeptical of everything	$\mu = 4.43, \sigma = 0.92$	$\mu = 4.38, \sigma = 1.01$	$\chi^2(5, N = 294) = 6.38, p = 1$
Be suspicious of links	$\mu = 4.78, \sigma = 0.51$	$\mu = 4.76, \sigma = 0.64$	$\chi^2(5, N = 294) = 6.55, p = 1$
Visit only known websites	$\mu = 3.93, \sigma = 1.04$	$\mu = 3.61, \sigma = 1.33$	$\chi^2(5, N = 294) = 24.99, p = 0.002$
Use two-factor authentication	$\mu = 4.46, \sigma = 0.74$	$\mu = 4.25, \sigma = 1.02$	$\chi^2(5, N = 294) = 16.89$, p =0.07
Don't click links from unknown people	$\mu = 4.74, \sigma = 0.57$	$\mu = 4.73, \sigma = 0.67$	$\chi^2(4, N = 294) = 3.85, p = 1$
Don't enter passwords on links in email	$\mu = 4.82, \sigma = 0.46$	$\mu = 4.82, \sigma = 0.48$	$\chi^2(3, N = 294) = 3.89, p = 1$
Look at the URL bar	$\mu = 4.68, \sigma = 0.62$	$\mu = 4.66, \sigma = 0.65$	$\chi^2(5, N = 294) = 1.3986, p = 1$
Don't open email attachments	$\mu = 4.82, \sigma = 0.47$	$\mu = 4.80, \sigma = 0.60$	$\chi^2(4, N = 294) = 3.70, p = 1$

Table 2: Results of χ^2 tests comparing non-expert ratings of security advice in terms of effectiveness versus how likely they are to follow it. p-values are corrected for multiple testing using the Holm-Bonferroni method.

Our results suggest that one reason some non-experts don't install updates might be the lack of awareness on how effective updates are. This hypothesis is also supported by additional feedback that participants provided when given the chance to explain their ratings with a question titled "(optional) Please use this space to clarify any of the above." For example, N56 said: *"I don't know if updating software is always safe. What it you download malicious software?"* Even some experts expressed similar concerns. For example, E163 agreed: *"Automatic software updates are not safe in my opinion, since it can be abused to update malicious content."* In contrast, E143 favored automatic to manual updates *"because update dialogs can be spoofed."*

Seven non-experts reported delaying updates out of concern that new versions of software might contain bugs. For example, N80 explained: *"there are often bugs in these updates initially, that must be worked out by the software vendor."* He, therefore, preferred to wait for the next update *"to make sure it is actually a stable release."* For the same reason, N142 did not like automatic updates: *"sometimes the patches [...] are glitchy [...]. I prefer to have control and know what's being installed by applications."* Even some experts expressed similar concerns. When asked how soon he installs updates, E168 said he only installs updates *"after i do the tests on spare machine."* Eight non-experts said they prefer having control over when updates happen, and seven said they do not like auto updates. N278 went as far as to say: *"I hate automatic updates."* Our findings are consistent with those of Vaniea et al. [50], who found in a study with 37 non-expert Windows 7 users that they frequently decided not to install updates after past negative experiences.

We found some controversy among experts on the difficulty of keeping software updated. While E178 considered it *"easy,"* E161 believed software updates are often *"cumbersome."* E28 pointed out that *"every Windows application uses a different update mechanism,"* and E97 confirmed that users find it difficult to deal with updates: *"I help a lot of non tech savvy users. Panic ensues when an update button shows up."*

4.2 Use Antivirus Software

At the other end of the spectrum depicted in Figure 2 is using antivirus software—the security action mentioned by most non-experts relative to experts. Thirty-five percent more non-experts than experts said that running antivirus software on their personal computers is one of the top three things they do to stay safe online. Figure 1 shows that 42% of non-experts and 7% of experts mentioned using antivirus software among the top three things they do to stay safe online. Five percent of non-experts and 1% of experts also reported keeping the antivirus software up-to-date.

This finding is in line with the high rates of adoption of antivirus software reported in a multiple choice question. When asked if they use antivirus software on their personal computers, 85% of non-experts reported doing so—compared to the 63% of experts who said they do. The difference between the two groups is statistically significant ($\chi(1, N_e = 221, N_n = 289) = 31.44, p < 0.001$). One factor explaining this finding might be the fact that experts and non-experts may use different operating systems. We did not ask what operating systems participants used, but a higher percentage of experts than non-experts (6% vs. 1%) named using Linux as one of the top three things they do to stay safe online. Several experts did mention that the need to run antivirus software is operating system dependent.

Further data we collected also confirms that non-experts consider using antivirus software very effective at protecting their security. We asked non-experts to rate on a 5-point Likert scale *how effective* they consider the security advice *Use antivirus software*. As shown in Figure 7, 69% rated it "5 (Very effective)" and 21% rated it "4 (Effective)". When asked *how likely* they would be to follow this advice if they heard that using antivirus software was effective, 80% of non-experts considered themselves "5 (Very likely)" to do so and another 12% said they were "4 (Likely)." Non-experts' high appraisal for antivirus software is also reflected in optional feedback comments that some non-experts provided. For example, N159 said that *"keeping a good antivirus along with a good malware program [...] is the best way to stay safe.“*

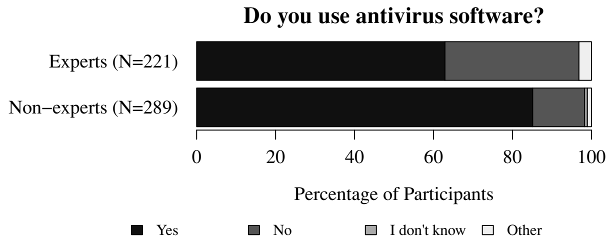


Figure 6: More non-experts reported to use antivirus software on their personal computer than experts.

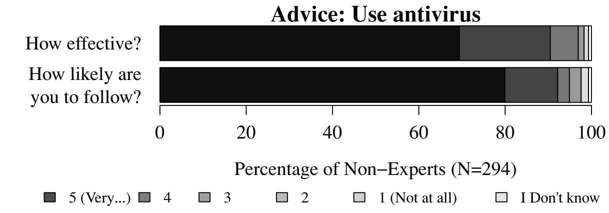


Figure 7: Most non-experts rated the advice 'Use antivirus software' effective or very effective. Similarly, they considered themselves very likely to follow this advice.

The high adoption of antivirus software among non-experts and their high willingness to follow this advice might be due to the good usability of the install-once type of solution that antivirus software offers. Similar to running antivirus, firewalls were also popular among non-experts. Although only 3% of the experts we surveyed mentioned using a firewall as one of the top three things they do, 17% of non-experts mentioned firewalls—often in conjunction with antivirus software. While experts acknowledged the usability of antivirus software, some also cautioned that antivirus is not a bulletproof security solution. For example, E47 believed that *"AV is simple to use, but less effective than installing updates."* Similarly, E116 believed that an antivirus *"is good at detecting everyday/common malware. But nothing that's slightly sophisticated."* E27 also cautioned that an antivirus *"also needs to be kept up-to-date, which is often not the case."*

4.3 Account Security

In the top three things they do, both experts and non-experts spoke frequently of passwords. Using strong and unique passwords were some of the most mentioned strategies by both groups. However, while more experts than non-experts emphasized having *unique* passwords (25% vs. 15%), fewer talked about having *strong* passwords (18% vs. 30%). Similarly, experts mentioned more frequently using a password manager (12% experts vs. 3% non-experts), but spoke less of changing passwords frequently (2% experts vs. 21% non-experts).

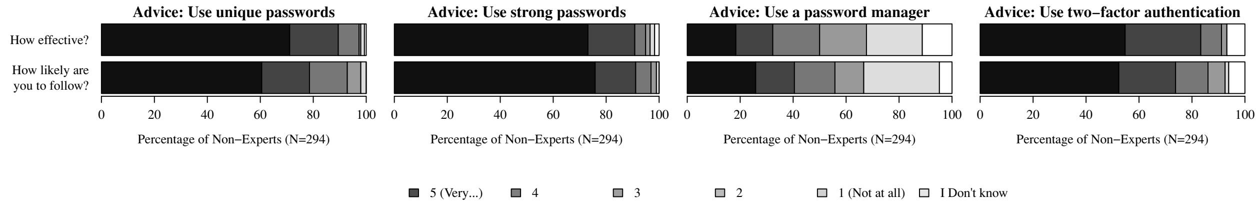


Figure 8: Non-experts considered the advice to use unique and strong passwords very effective, but were less aware of the security benefits of using a password manager or two-factor authentication.

4.3.1 Use a Password Manager

To better understand how the two groups differ in their password management habits, we asked a series of multiple-choice questions about password behavior. While more experts said they *use a password manager* to keep track of their passwords, more non-experts said they *write down* passwords, *remember*, or *reuse* them. As Figure 9 shows, three times more experts than non-experts reported using password managers for at least *some* of their accounts (73% vs. 24%, $\chi(3,\ N_e = 231,\ N_n = 294) = 131.31,\ p < 0.001$). This difference is in line with the fact that four times more experts than non-experts said that using a password manager is one of the most important things they do to stay safe online (13% vs. 3%, see Figure 1). To experts such as E123, *"Password managers change the whole calculus,"* because they make it possible to have both strong and unique passwords.

The low adoption rate of password managers among non-experts might stem from a lack of understanding of it's security benefits. To explore non-experts' attitudes, we asked them to rate in terms of effectiveness on a 5-point Likert scale the advice "Use a password manager." Only 18% considered this advice *very effective*, and another 14% thought it was *effective*. Thirty-nine percent believed that the advice was not effective and 11% said that they did not know. In the optional feedback, seven non-experts explicitly expressed distrust in password managers. For example, N278 said: *"I wouldn't use a password manager even if it helps because I don't trust it."* A reason for this lack of trust was the fear that, if stored or written down, passwords could be leaked. For example, N53 explained, *"I try to remember my passwords because no one can hack my mind."* The fear that software can be hacked is reflected also in N251's comment that password managers should be *"completely trustworthy and impregnable. No other applications seem to be that safe so how can I believe password managers are."* In fact, 2% of non-experts thought that not letting browsers remember their passwords was one of the top things they do.

In addition to perceived lack of effectiveness, other factors such as poor usability might stall adoption of password managers among non-experts. In another Likert scale question, only 40% of non-experts said they would be likely or very likely to follow this advice if they heard it was effective. This percentage is much lower than the 91% who said they would use strong passwords if they heard this security measure was effective, and the nearly 80% who said they would use unique passwords. Some additional comments made by experts might help explain these answers. For example, E71 pointed out that password managers *"tend to be complicated for non-technical users still."* E9 named a specific problem, that when starting to use a password manager *"it is difficult to update existing passwords."* Our results are in line with those of Chiasson et al. [11] who found that the usability of password managers could be improved.

4.3.2 Write Passwords Down

Writing down passwords was seen by some experts as a substitute to using a password manager. E121 believed that *"People understand a paper system very well, and know how to secure it."* Similarly, E79 noted another benefit: *"Malware can't read a piece of paper."* As Figure 9 shows, more non-experts than experts reported to *write down* passwords for at least *some* of their accounts (38% vs. 20%, $\chi(3, N_e = 231, N_n = 294) = 24.78, p < 0.001$). Only one expert said that writing down passwords is fundamentally bad, but several expressed concern for how securely the paper would be stored.

As shown in Figure 9, three times more non-experts than experts said that they *remember all* of their passwords (17% experts vs 52% of non-experts, $p < 0.001$). Furthermore, six times more non-experts say that they *use the same* password for *all or most* of their accounts (19% of non-experts vs. 3% of experts, $\chi(3,\ N_e = 231,\ N_n = 294) = 37.25,\ p < 0.001$). Only 4% of experts and 15% of non-experts said they do not remember any of their passwords. From the additional feedback we received, a couple of techniques for making passwords easier to remember stood out: (a) using an algorithm for creating passwords—mentioned by 19 non-experts and 21 experts—and (b) having different password "levels"—mentioned by 8 experts and 7 non-experts. For example, N277 described the algorithmic approach: *"I use a base password and just have a suffix that is usually unique."* N241 reported using the password levels approach: *"I have three levels of passwords, actually four. I have a password for my bank, Amazon, Paypal accounts."*

Our results are consistent with Florencio et al. [16], who, back in 2007, found that users have a set of passwords which they cycle through, and use trial and error to remember which password they used.

Some participants occasionally forgot passwords, or did not remember which password they used. For example, E163 has a *"few [passwords], so if one doesn't work, I'll try another."* Similarly, N43 tried to *"remember them by trial and error."* Three non-experts mentioned that sometimes they resort to password reset to get back into their accounts. For example, N200 said: *"I don't do anything other than try to remember them and I often have to reset because I've forgotten."*

4.3.3 Change Passwords Frequently

The security action with the highest percentage difference for non-experts compared to experts was changing passwords frequently. 21% of non-experts, but only 2% of experts mentioned changing passwords when asked what are the top three things they do to stay safe online. We did not ask experts to rate how good the advice to change passwords frequently is, but some researchers have questioned the effectiveness of this action. Zhang et al. raised concerns over how effective this action is at protecting against an attacker who has captured an old password [56]. The authors showed that, by knowing the old password and applying simple transformations

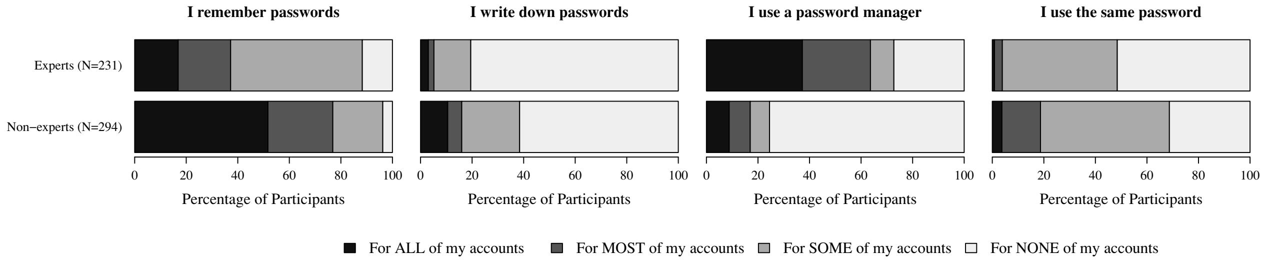


Figure 9: More non-experts than experts reported remembering passwords and using the same password on several accounts, while more experts say they use a password manager.

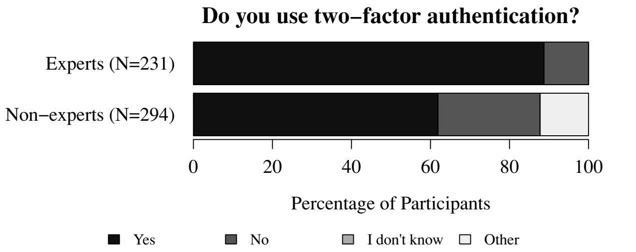


Figure 10: More experts than non-experts reported to use two-factor authentication for at least one of their online accounts.

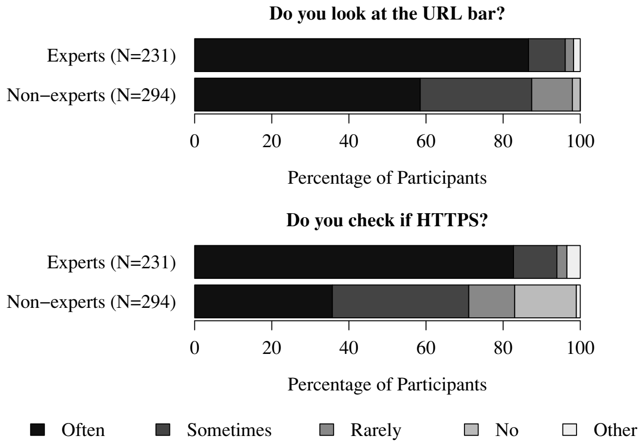


Figure 11: More experts than non-experts reported to look at the URL bar to verify if they are visiting the website they intended to, and to check whether the website uses HTTPS.

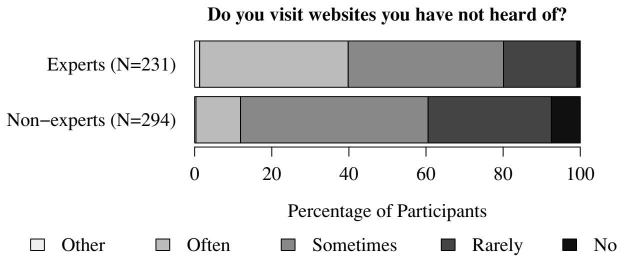


Figure 12: More non-experts than experts reported to only visit websites they have heard of.

to it, an attacker is able to guess the new one 41% of the time for an offline attack, and 17% within five online attempts.

4.3.4 Use Two-Factor Authentication

Another popular security action among experts was to use two-factor authentication. For example, E50 considered two-factor authentication *"hugely important for high-value services (such as Gmail)."* Non-experts rated this advice significantly higher than using a password manager, both in terms of effectiveness (83% vs. 32%, $\chi(4, N = 294) = 177.53$, $p < 0.001$) and likelihood of following the advice (74% vs. 40%, $\chi(4, N = 294) = 107.24$). However, the adoption rates of two-factor authentication among non-experts still lag behind those of experts. When asked if they use two-factor authentication for at least one of their online accounts, more experts than non-experts answered that they do (89% vs. 62%, $p < 0.001$). 12% of non-experts said they don't know if they do, which suggests they may not know what two-factor authentication is.

Ten experts expressed concerns that two-factor authentication is still too difficult for many users or not widely available. For example, E161 said that *"On average most don't yet understand two factor very well,"* and E207 said that two-factor authentication *"will need good instructions on how it works"*. Furthermore, E132 noted that *"using two-factor authentication will likely not be possible or feasible for a lot of sites."*

4.4 Mindfulness

The remaining security actions mentioned by experts and non-experts did not have as stark of a percentage difference as using antivirus, installing software updates, and managing passwords (see Figure 2). In this section, we discuss these remaining items. We focus on visiting only known websites and checking if the websites use HTTPS. We also discuss security advice related to emailing habits, which was not mentioned frequently by either group as a top 3 practice, but was ranked highly when we explicitly asked about it.

4.4.1 Visit Only Known Websites

After using antivirus and changing passwords frequently, the practice most mentioned by non-experts relative to experts (see percentage difference in Figure 2) was visiting only known websites. 21% of non-experts—but only 4% of experts—said they only go to known or reputable websites to stay safe online. In addition, 4% of non-experts said they provide personal information only to trusted websites and 3% said they make purchases only from trusted websites; no expert mentioned these practices.

One might wonder how realistic it is to only visit known websites. The answers to a multiple-choice question we asked seem to suggest that both experts and non-experts sometimes make exceptions to this rule. We asked both groups in a multiple choice question if they visit websites they have not heard of. Figure 12 shows the results. Seven percent of non-experts said they *do not* visit unknown websites—compared to the 21% who mentioned this practice as one of the three most important things they do to stay safe online. Only 1% of experts said they do not visit unknown websites—also lower than the percentage who mentioned this in their top three. Thirty-two percent of experts and 19% of non-experts said they *rarely* visit unknown websites. The difference is statistically significant ($\chi^2(3,\ N_e = 231,\ N_n = 294) = 62.84$, $p < 0.001$). It is unclear how the browsing needs of experts and non-experts differ and how they influence behavior. Perhaps non-experts visit a more limited number of websites because they do not have as high of a need as experts to explore new things and conduct research on the Internet.

When asked to rank the advice "Visit only known websites" on a Likert scale, 76% of non-experts rated it *very effective* or *effective* (see Figure 15). After marking this advice *effective*, N191 commented *"Visiting websites you've heard of doesn't mean they are completely safe, but there is a higher chance of this."* When asked how likely they would be to follow this advice if they heard it was effective, 57% of non-experts said they were *likely* or *very likely* to follow it—lower than the 76% who rated the advice effective. This finding might suggest that 'Visit only known websites' is not always practical. In fact, four non-experts explicitly commented on this. N236 said: *"It would be impossible to only visit websites you know. Why not hide under the bed too?"* For N98, not visiting new websites is *"missing the point of the internet."* Some experts pointed out problems with this advice as well. For example, E134 said: *"Visiting only known websites is great, but paralyzing."* E7 reported another shortcoming: *"Visiting websites you've heard of makes no difference in a modern web full of ads, cross-site requests."*

4.4.2 Check if HTTPS

After software updates and account security (use strong and unique passwords, use a password manager and two-factor authentication), the most mentioned practice by experts when asked about their top three was using HTTPS. As Figure 1 shows, 10% of experts and 4% of non-experts said they check if the website they are visiting uses HTTPS as a top 3 action. In addition, 2% of non-experts said that they do not provide credit card information, and 3% said they don't give credentials or private information, unless the connection is over HTTPS. Looking at the URL bar to check what website they are visiting was mentioned by 3% of experts, but only by one non-expert.

In a multiple-choice question, we asked both groups if they look at the URL bar to verify that they are visiting the intended website. Figure 11 shows the results. 86% of experts and 59% of non-experts said they do so *often*. When asked in a similar question if they check whether the website they are visiting uses HTTPS, 82% of experts and 36% of non-experts said they *often* do. Both differences are statistically significant ($p < 0.001$). Note that for experts, the likelihood of checking for HTTPS and looking at the URL bar to verify the name of the site are equal, whereas non-experts are far more likely to report looking at the URL bar but not checking for HTTPS.

When asked to rate the advice, 75% of non-experts said that checking the URL is *very effective*; 74% said that they would be *very likely* to follow it (see Figure 15). When asked about the advice "Check if the website they're visiting uses HTTPS", 60% rated it *very effective*, but only 50% considered themselves *very likely* to follow it. It is unclear why some participants claim they would look at the URL bar but not check for HTTPS or why some consider checking for HTTPS effective, but would not follow this advice.

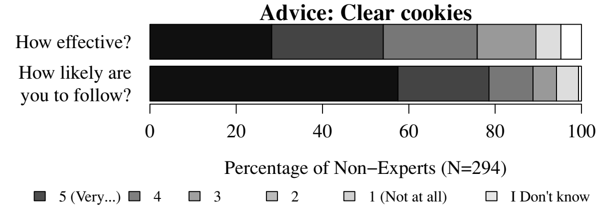


Figure 14: Most non-experts considered themselves very likely to delete cookies if they heard it was an effective security measure.

4.4.3 Clear Browser Cookies

Six percent of non-experts and only one expert said that deleting or restricting cookies is one of the top three things they do to stay safe online. When asked how good the advice "Clear browser cookies" is, 54% of experts rated it not good or not good at all. Only 21% rated it good or very good. E127 said that *"Clearing cookies might be OK to prevent some session hijacking, but the annoyance of logging in again might throw some users off."* E8 specified that *"Clearing cookies is more of a privacy measure."* It is likely that, while experts distinguished between privacy and security measures, non-experts conflated the two areas. In fact, N103 explicitly stated: *"Clearing cookies might not protect you from viruses but 'online security' is a very broad term and I believe privacy is part of online security."*

Figure 14 shows that 78% of non-experts considered themselves likely or very likely to clear browser cookies if they heard this measure helped protect their security online. Fifty-four percent considered deleting cookies an effective security measure. N284 commented *"I forget to clear cookies."* N281 believed that *"If it was an easy solution–like clearing cookies–I'd do it all the time."*

4.4.4 Email Habits

Some email-related security advice that we collected during our interviews was not frequently mentioned by survey participants but was rated highly on Likert scales when explicitly asked about. In the following, we discuss two such pieces of advice: 'Don't enter your password when you click on a link in an email and that link takes you to a website that asks for your password' and 'Don't click on links that people or companies you don't know send you.' However, when explicitly asked about them, these pieces of advice were rated as good advice by experts (see Figure 3)

Similarly, when asked how effective this advice is, 85% of non-experts rated *very effective* the advice 'Don't enter your password when you click on a link in an email and that link takes you to a website that asks for your password.' Eighty-six percent considered themselves themselves very likely to follow this advice. A similarly high number—80%—said that not clicking on links that people they don't know send them is *very effective* advice (see Figure 15). Eighty-two percent said they would be likely to follow this advice if they heard it was effective. In a multiple choice question, both groups reported to generally follow these two pieces of advice (see Figure 13).

It is noteworthy that a significantly higher percentage of experts than non-experts reported to *often* click on links that people they

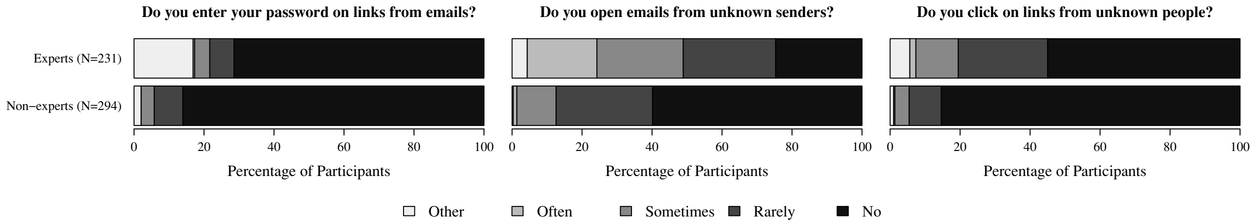


Figure 13: Both experts and non-experts said they follow good email practices. More experts said they sometimes click on links and open attachments received from unknown senders.

don't know send them (38% vs. 12%, $\chi(3, N_e = 231, N_n = 294) = 51.37, p < 0.001$). This result suggest that, although some security advice is being followed more by experts than by non-experts (e.g., installing updates, using a password manager), other advice is perhaps paradoxically being followed more by non-experts. During our interviews, some experts admitted that they do not follow some of the advice they give. For example, after recommending that non-tech-savvy users never open emails from unknown people, an expert admitted: *"I do all the time, [laughter] but I tell my mother not to."* Another expert explained during our interviews: *"I never really found a way of giving more precise advice for people who are not technical on what is really safe and what is not."* A couple other interview participants said that they don't follow their own advice because, unlike non-tech-savvy users, they can distinguish between when it's safe and when not to take certain actions.

Other habits mentioned by experts and non-experts include restricting the amount of personal information they share (10% of experts vs. 17% of non-experts), installing only trusted or verified software (5% of experts vs. 6% of non-experts), using Linux (6% vs. 1%), and deleting cookies (0% vs. 6%).

5. DISCUSSION AND FUTURE WORK

Our results show that experts and non-experts follow different practices to protect their security online. The experts' practices are rated as good advice by experts, while those employed by non-experts received mix ratings from experts. Some non-expert practices were considered "good" by experts (e.g., install antivirus software, use strong passwords); others were not (e.g., delete cookies, visit only known websites).

In the pursuit of better security advice, we should ensure that valuable user time is being spent on the things that would bring them the most benefit. Our results suggest that at least some things that experts do and recommend are not being done by non-experts. In this work, we identified three security practices that experts report to do but non-experts do not: installing updates, using a password manager, and using two-factor authentication.

These three pieces of security advice that we highlight are the security actions that most experts relative to non-experts said they do and consider important. These three security actions were ranked highest by percentage difference in Figure 2. This recommendation is also supported by differences among experts and non-experts in self-reported behavior around these three security actions (Figures 4, 9, and 10). Our results suggest that not just better messaging, but also systems and usability work is necessary to get non-experts to follow these three security practices.

In line with the findings and the recommendations provided by Vaniea et al. [50], our results suggest the need to invest in developing an updates manager that downloads and installs available software updates for all applications—much like mobile application updates on smartphones. Such a centralized updates manager could also access a central repository to check if any problems with available updates have been reported, and, if so, delay the installation. In addition, software developers should separate security updates from those introducing general software features. The update manager could give users the option to install only security updates automatically, while feature updates could be manually reviewed and installed. To that end, a notable area for further research is developing a standard way for applications to communicate what UI changes and security fixes are included in the update.

Another practice employed by experts but not by non-experts was using a password manager to keep track of passwords. Some participants reported cycling through multiple passwords to remember which they used for a given site. Trying various passwords on a website until one works leaves the user vulnerable to a rogue or compromised website. Furthermore, as rainbow tables used to crack password hashes evolve to incorporate password rules used by users, password-creation algorithms that some participants used will most likely not offer real protection against offline attacks, as demonstrated by Das et al. [12]. Password managers can make it feasible to use truly random and unique passwords and help move users away from memorable passwords, which are vulnerable to smart-dictionary attacks [36]. However, non-experts might place a higher emphasis than experts on usability. Previous work has evaluated two different password managers and identified significant usability shortcomings [11]. Perhaps such usability drawbacks are harder to deal with for non-experts than for experts. To that end, more work needs to be done to improve the usability of password managers before recommending them strongly to users. Our results also suggest that users' reluctance to adopt password managers may also be due to an ingrained mental model that passwords should not be stored or written down—advice users have been given for decades. But as threat models are shifting from offline to online attacks and password reuse is becoming an increasing problem, using password managers or writing passwords down in a secure location seems to be a promising solution.

Furthermore, additional work needs to be done to understand why non-experts are not using two-factor authentication. Some of the expert participants in our study offered several reasons, including the fact that this security feature is still too difficult to explain to non-tech-savvy users, that it is not available on all websites, and that it causes significant inconveniences. Our results suggest that more work needs to be done to explain two-factor authentication to users and to make it adequate to use by non-expert users.

A few additional areas for future research stand out based on our findings. Many more non-experts than experts said that, to stay safe online, they only go to trusted websites. This security advice was

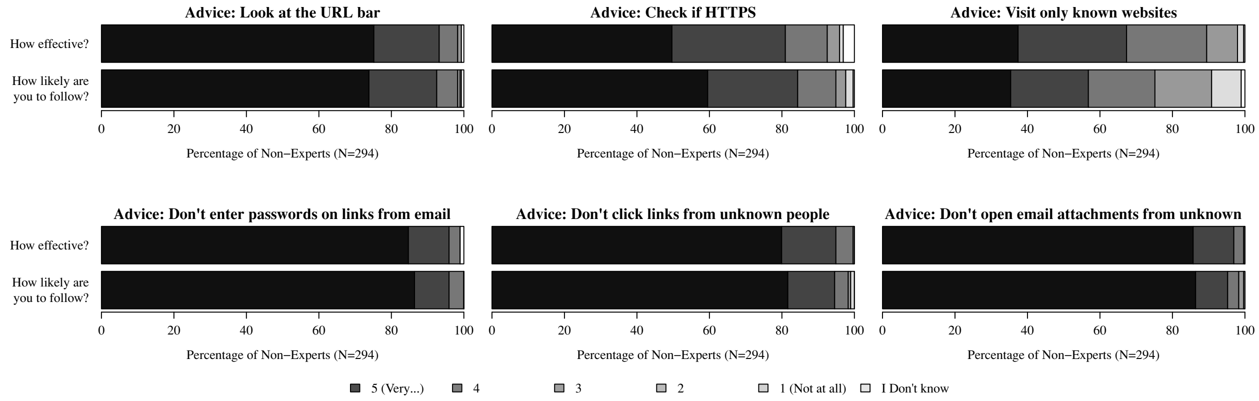


Figure 15: Most non-experts rated email advice very effective and said they are very likely to follow it. They considered looking at the URL bar more effective than checking for HTTPS and visiting only known websites.

not considered "good" by experts. Furthermore, Google's transparency report on safe browsing [1] shows that most malware websites are not malicious attack sites, but compromised sites that are used to spread malware. Therefore, not visiting new websites is not necessarily effective at keeping users safe. A good investment in this area is developing browsers and systems that more effectively warn users when they are about to go to a compromised or known phishing website, something that the Chrome and Firefox browsers already do [5].

Finally, further investigation is needed to understand why, unlike experts, some non-experts claimed they would look at the URL bar but not check for HTTPS. Another study could investigate why some non-experts consider checking for HTTPS effective, but they admit that they would not follow the advice. Our findings seem to indicate that more visible and intuitive HTTPS indicators could help some users better assess if a website uses HTTPS. Better URL indicators have been shown to help some users. Lin et al. found that domain highlighting in the URL helps some (though not all) users judge the legitimacy of a website and avoid phishing attacks [33].

5.1 Limitations

The study presented in this paper is not without its limitations. First, we recruited non-expert participants on the Mechanical Turk platform, which is known to provide a younger and more tech-savvy sample than the general population. All our non-expert participants were from the US; running the study in other countries might lead to different results. Furthermore, all behavioral data that we collected was self-reported and, therefore, unconfirmed. Such data can suffer from several biases, including social desirability, inaccurate recall, and lack of understanding. For example, participants may not be able to accurately remember how often they check if the website they are visiting uses HTTPS or how soon they install available software updates.

We compared expert and non-expert security behavior, but we note that experts by their nature may operate in different computing environments than non-experts, so their reported behavior may be different not because it is objectively better, but simply because it is more suited to the expert environment. For example, it may be the case that experts are more likely to use Unix-based systems while non-experts are more likely to use Windows-based systems, and appropriate security practices (at least some of them, like using antivirus software) may depend upon the operating system in use.

Defining a security "expert" is challenging, and we settled upon a definition that is simple (5+ self-reported years of experience in the area) but, intuitively, suggests strong expertise. However, even experts are not infallible. For example, Yen et al. [55], in a study of malware encounters in a large enterprise, found higher levels of encounters among users with technical job titles (e.g., "engineer") than among those with less technical titles (e.g., "assistant"). While this result could be due to the technical users simply spending more time using computers or perhaps taking greater risks, it shows that even the tech-savvy are prone to security threats. Thus, expert behavior should not necessarily be taken as the right standard for non-tech-savvy users.

6. CONCLUSIONS

Our results find discrepancies between what security practices experts and non-experts follow. While most expert participants install updates, use a password manager, and use two-factor authentication, most non-expert participants use antivirus software, change passwords frequently, and visit only known websites. Non-expert participants reported being reluctant to promptly install software updates, perhaps due to lack of understanding of their effectiveness or bad past experiences caused by software updates. Though using them was considered good advice by experts, password managers were regarded with skepticism by non-experts, who instead preferred to remember passwords, partly because, as one participant said, "no one can hack my mind." Other security advice, however, such as not clicking on links received from unknown people were known and followed by non-experts. More work has to be done on improving the limitations of security practices identified in this work which are used by experts but not by non-experts. Nevertheless, based on our findings, some promising security advice emerges: (1) install software updates, (2) use a password manager, and (3) use two-factor authentication for online accounts.

7. ACKNOWLEDGMENTS

The authors wish to thank the many security experts who volunteered their time to participate in our interviews and respond to our survey. We also wish to thank Eric Grosse, Tim Dierks, and Ulfar Erlingsson at Google for supporting this work and reviewing early paper drafts, and Cormac Herley and Adam Shostack for early discussions that guided our research direction. Finally, we wish to thank the many colleagues, friends, and paper reviewers who contributed valuable feedback. Thanks all!

8. REFERENCES

[1] Google transparency report: Safe browsing. Accessed Feb 2, 2014. `http://www.google.com/transparencyreport/safebrowsing/`.

[2] McAfee security advice center. Accessed Sep 8, 2014. `http://home.mcafee.com/advicecenter/`.

[3] US-CERT: Tips. Accessed Sep 8, 2014. `https://www.us-cert.gov/ncas/tips`.

[4] A. Adams and M. A. Sasse. Users are not the enemy. *Communications of the ACM*, 42(12):40–46, 1999.

[5] D. Akhawe and A. P. Felt. Alice in warningland: A large-scale field study of browser security warning effectiveness. In *Proc. of USENIX Security*, pages 257–272, 2013.

[6] A. Beautement, M. A. Sasse, and M. Wonham. The compliance budget: managing security behaviour in organisations. In *Proc. of NSPW*, pages 47–58. ACM, 2009.

[7] D. Bohn. How to make your email address as hard to guess as your password. *The Verge*, Sep 3, 2014. `http://www.theverge.com/2014/9/3/6100893/how-to-make-your-email-address-as-hard-to-guess-as-your-password`.

[8] M. Buhrmester, T. Kwang, and S. D. Gosling. Amazon's Mechanical Turk a new source of inexpensive, yet high-quality, data? *Perspectives on Psychological Science*, 6(1):3–5, 2011.

[9] L. J. Camp. Mental models of privacy and security. *IEEE Technology & Society*, 2006.

[10] B. X. Chen. Home Depot investigates a possible credit card breach. *The New York Times*, Sep 03, 2014. `http://www.nytimes.com/2014/09/03/technology/home-depot-data-breach.html`.

[11] S. Chiasson, P. C. van Oorschot, and R. Biddle. A usability study and critique of two password managers. In *Prox. of USENIX Security*, volume 6, 2006.

[12] A. Das, J. Bonneau, M. Caesar, N. Borisov, and X. Wang. The tangled web of password reuse. In *Proc. of NDSS*, 2014.

[13] R. Dhamija, J. D. Tygar, and M. Hearst. Why phishing works. In *Proceedings of the SIGCHI conference on Human Factors in computing systems*, pages 581–590. ACM, 2006.

[14] S. Egelman, L. F. Cranor, and J. Hong. You've been warned: an empirical study of the effectiveness of web browser phishing warnings. In *Proc. of CHI*, pages 1065–1074. ACM, 2008.

[15] S. Egelman, A. Sotirakopoulos, I. Muslukhov, K. Beznosov, and C. Herley. Does my password go up to eleven?: the impact of password meters on password selection. In *Proc. of CHI*, pages 2379–2388. ACM, 2013.

[16] D. Florencio and C. Herley. A large-scale study of web password habits. In *Proc. of WWW*, pages 657–666. ACM, 2007.

[17] Google safety center. Accessed Sep 8, 2014. `https://www.google.com/safetycenter/`.

[18] E. Hayashi and J. Hong. A diary study of password usage in daily life. In *Proc. of CHI*, pages 2627–2630. ACM, 2011.

[19] C. Herley. So long, and no thanks for the externalities: the rational rejection of security advice by users. In *Proc. of NSPW*, pages 133–144. ACM, 2009.

[20] C. Herley. More is not the answer. *IEEE Security & Privacy*, 12(1):14–19, 2014.

[21] A. E. Howe, I. Ray, M. Roberts, M. Urbanska, and Z. Byrne. The psychology of security for the home computer user. In *Security and Privacy (SP), 2012 IEEE Symposium on*, pages 209–223. IEEE, 2012.

[22] P. G. Inglesant and M. A. Sasse. The true cost of unusable password policies: password use in the wild. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 383–392. ACM, 2010.

[23] I. Ion, N. Sachdeva, P. Kumaraguru, and S. Čapkun. Home is safer than the cloud!: privacy concerns for consumer cloud storage. In *Proc. of SOUPS*, page 13. ACM, 2011.

[24] M. Isaac. Nude photos of Jennifer Lawrence are latest front in online privacy debate. *The New York Times*, Sep 03, 2014. `http://www.nytimes.com/2014/09/03/technology/trove-of-nude-photos-sparks-debate-over-online-behavior.html`.

[25] M. Isaac. Russian hackers amass over a billion internet passwords. *The New York Times*, Aug 06, 2014. `http://www.nytimes.com/2014/09/03/technology/home-depot-data-breach.html`.

[26] P. G. Kelley, S. Komanduri, M. L. Mazurek, R. Shay, T. Vidas, L. Bauer, N. Christin, L. F. Cranor, and J. Lopez. Guess again (and again and again): Measuring password strength by simulating password-cracking algorithms. In *Proc. of Security and Privacy*, pages 523–537. IEEE, 2012.

[27] S. M. Kelly. How to protect your photos (nude or otherwise) from hackers on iCloud. *Mashable*, Sep 1, 2014. `http://mashable.com/2014/09/01/icloud-nude-photo-hack/`.

[28] M. Khan, Z. Bi, and J. Copeland. Software updates as a security metric: Passive identification of update trends and effect on machine infection. In *MILITARY COMMUNICATIONS CONFERENCE, 2012-MILCOM 2012*, pages 1–6. IEEE, 2012.

[29] A. Kittur, E. H. Chi, and B. Suh. Crowdsourcing user studies with Mechanical Turk. In *Proc. of CHI*, pages 453–456. ACM, 2008.

[30] S. Komanduri, R. Shay, P. G. Kelley, M. L. Mazurek, L. Bauer, N. Christin, L. F. Cranor, and S. Egelman. Of passwords and people: measuring the effect of password-composition policies. In *Proc. of CHI*, pages 2595–2604. ACM, 2011.

[31] J. R. Landis and G. G. Koch. The measurement of observer agreement for categorical data. *Biometrics*, pages 159–174, 1977.

[32] F. L. Levesque, J. Nsiempba, J. M. Fernandez, S. Chiasson, and A. Somayaji. A clinical study of risk factors related to malware infections. In *Proc. of CCS*, pages 97–108. ACM, 2013.

[33] E. Lin, S. Greenberg, E. Trotter, D. Ma, and J. Aycock. Does domain highlighting help people identify phishing sites? In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '11, pages 2075–2084, New York, NY, USA, 2011. ACM.

[34] E. L. MacGeorge, B. Feng, and E. R. Thompson. "Good" and "bad" advice. *Studies in applied interpersonal communication*, page 145, 2008.

[35] Microsoft safety & security center. Accessed Sep 8, 2014. `http://www.microsoft.com/security/default.aspx`.

[36] A. Narayanan and V. Shmatikov. Fast dictionary attacks on passwords using time-space tradeoff. In *Proc. of CCS*, pages 364–372. ACM, 2005.

[37] G. Paolacci, J. Chandler, and P. G. Ipeirotis. Running experiments on Amazon Mechanical Turk. *Judgment and Decision making*, 5(5):411–419, 2010.

[38] A. Peterson. How to game security questions to make yourself safer online. *The Washington Post*, Sep 4, 2014. **http://www.washingtonpost.com/blogs/the-switch/wp/2014/09/04/how-to-game-security-questions-to-make-yourself-safer-online/**.

[39] E. Rader, R. Wash, and B. Brooks. Stories as informal lessons about security. In *Proc. of SOUPS*. ACM, 2012.

[40] R. Reeder. If you could tell a user three things to do to stay safe online, what would they be? *Google Online Security Blog*, March 26, 2014. **http://googleonlinesecurity.blogspot.com/2014/03/if-you-could-tell-user-three-things-to.html**.

[41] M. A. Sasse, C. C. Palmer, M. Jakobsson, S. Consolvo, R. Wash, and L. J. Camp. Helping you protect you. *IEEE Security & Privacy*, 12(1):39–42, 2014.

[42] S. E. Schechter, R. Dhamija, A. Ozment, and I. Fischer. The emperor's new security indicators: An evaluation of website authentication and the effect of role playing on usability studies. In *Proceedings of the 2007 IEEE Symposium on Security and Privacy*, page 51âĂŞ65, Washington, DC, USA, May 2007. IEEE Computer Society.

[43] R. Shay, I. Ion, R. W. Reeder, and S. Consolvo. My religious aunt asked why i was trying to sell her viagra: experiences with account hijacking. In *Proc. of CHI*, pages 2657–2666. ACM, 2014.

[44] R. Shay, S. Komanduri, A. L. Durity, P. S. Huh, M. L. Mazurek, S. M. Segreti, B. Ur, L. Bauer, N. Christin, and L. F. Cranor. Can long passwords be secure and usable? In *Proc. of CHI*, pages 2927–2936. ACM, 2014.

[45] S. Sheng, M. Holbrook, P. Kumaraguru, L. F. Cranor, and J. Downs. Who falls for phish?: a demographic analysis of phishing susceptibility and effectiveness of interventions. In *Proc. of CHI*, pages 373–382. ACM, 2010.

[46] A. Sotirakopoulos, K. Hawkey, and K. Beznosov. On the challenges in usable security lab studies: Lessons learned from replicating a study on ssl warnings. In *Proceedings of the Seventh Symposium on Usable Privacy and Security*, SOUPS '11, pages 3:1–3:18, New York, NY, USA, 2011. ACM.

[47] G. Stewart and D. Lacey. Death by a thousand facts: Criticising the technocratic approach to information security awareness. *Information Management & Computer Security*, 20(1):29–38, 2012.

[48] J. Sunshine, S. Egelman, H. Almuhimedi, N. Atri, and L. F. Cranor. Crying wolf: An empirical study of ssl warning effectiveness. In *USENIX Security Symposium*, pages 399–416, 2009.

[49] B. Ur, P. G. Kelley, S. Komanduri, J. Lee, M. Maass, M. L. Mazurek, T. Passaro, R. Shay, T. Vidas, L. Bauer, et al. How does your password measure up? the effect of strength meters on password creation. In *USENIX Security Symposium*, pages 65–80, 2012.

[50] K. E. Vaniea, E. Rader, and R. Wash. Betrayed by updates: how negative experiences affect future security. In *Proc. of CHI*, pages 2671–2674. ACM, 2014.

[51] A. J. Viera, J. M. Garrett, et al. Understanding interobserver agreement: the kappa statistic. *Family Medicine*, 37(5):360–363, 2005.

[52] R. Wash. Folk models of home computer security. In *Proc. of SOUPS*, pages 1–16. ACM, 2010.

[53] K. Witte. Theory-based interventions and evaluations of outreach efforts. *Research review. Seattle, WA: National Network of Libraries of Medicine Pacific Northwest Region, Outreach Evaluation Resource Centre.*, 9:2007, 1998.

[54] M. Wu, R. C. Miller, and S. L. Garfinkel. Do security toolbars actually prevent phishing attacks? In *Proceedings of the SIGCHI conference on Human Factors in computing systems*, pages 601–610. ACM, 2006.

[55] T.-F. Yen, V. Heorhiadi, A. Oprea, M. K. Reiter, and A. Juels. An epidemiological study of malware encounters in a large enterprise. In *Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security*, pages 1117–1130. ACM, 2014.

[56] Y. Zhang, F. Monrose, and M. K. Reiter. The security of modern password expiration: an algorithmic framework and empirical analysis. In *Proc. of CCS*, pages 176–186. ACM, 2010.

APPENDIX

We include here the questions asked the in Expert and Non-expert Surveys. All multiple-choice questions were single answer only. The questions were identical for the Expert and Non-expert survey, unless otherwise stated.

We mark with "(Experts only)" or "(Non-experts only)" questions that were asked in only one of the surveys.

Survey Instruments

- *(Experts only)* What are the top 3 pieces of advice you would give to a non-tech-savvy user to protect their security online? *(open-ended)*
- What are the 3 most important things you do to protect your security online? *(open-ended)*
- How did you learn about the things you listed above? *(open-ended)*
- Do you use a laptop or desktop computer that you or your family owns (i.e., not provided by school or work)? *(multiple-choice)*
  - Yes
  - No
  - Other
- When did you get that computer? *(multiple-choice)*
  - Less than 1 year ago
  - At least 1 but less than 2 years ago
  - At least 2 but less than 3 years ago
  - At least 3 but less than 5 years ago
  - 5 or more years ago
- How soon after you discover that a new version of your operating system (OS) software is available do you (or somebody else managing your computer) install it? *(multiple-choice)*
  - OS updates are installed automatically
  - Immediately
  - Soon after
  - Eventually
  - OS updates are never installed
  - Other
- Do you use anti-virus software on that computer? *(multiple-choice)*
  - Yes
  - No
  - I don't know
  - Other
- Which anti-virus software do you use? *(open-ended)*
- How do you keep track of your passwords for your online accounts? *(grid question)*
  Answer options: For ALL of my accounts, For MOST of my accounts, For SOME of my accounts, For NONE of my accounts
  - Remember them
  - Write them down on paper
  - Save them in a local file on my computer
  - Have my password manager (e.g., 1Password, LastPass) remember them
  - Use the same password on multiple accounts
- If you use a password manager, which one do you use? *(open-ended)*
- (optional) What other things, if any, do you do to keep track of your passwords? *(open-ended)*
- Do you use two-factor authentication (e.g., 2-Step Verification) for at least one of your online accounts? *(multiple-choice)*
  - Yes
  - No
  - I don't know
  - Other
- Do you look at the URL bar to verify that you are visiting the website you intended to? *(multiple-choice)*
  - Yes, often
  - Yes, sometimes
  - Yes, rarely
  - No
  - I don't know

  - Other

- Google began in January 1996 as a research project. Its initial public oïňĂering took place on August 19, 2004. Did the initial public offering of Google take place in 1996? *(multiple-choice)*
  - Yes
  - No
  - Other
- Do you check if the website you're visiting uses HTTPS? *(multiple-choice)*
  - Yes, often
  - Yes, sometimes
  - Yes, rarely
  - No
  - I don't know
  - Other
- Do you visit websites you have not heard of before? *(multiple-choice)*
  - Yes, often
  - Yes, sometimes
  - Yes, rarely
  - No
  - I don't know
  - Other
- When you click on a link in an email and that link takes you to a website that asks for your password, do you enter it? Do you open emails you receive from people or companies you don't know? *(multiple-choice)*
  - Yes, often
  - Yes, sometimes
  - Yes, rarely
  - No
  - I don't know
  - Other
- Do you click on links that people or companies you don't know send you? *(multiple-choice)*
  - Yes, often
  - Yes, sometimes
  - Yes, rarely
  - No
  - I don't know
  - Other
- *(Experts only)* For each of the following pieces of advice, please rate on a scale from 1 to 5 how good (in terms of both EFFECTIVE at keeping the user secure, as well as REALISTIC that the user can follow it) you think they are at protecting a non-tech-savvy user's security online. *(grid question)*
  Scale: 5 (Very good), 4, 3, 2, 1 (Not at all), I don't know
  - Use anti-virus software
  - Install the latest operating system updates
  - Turn on automatic software updates
  - Update applications to the latest version
  - Clear your Web browser cookies
- *(Experts only)* (optional) Please use this space to clarify any of the above. *(open-ended)*
- *(Non-experts only)* For each of the following pieces of advice, please rate on a scale from 1 to 5 how EFFECTIVE you think the advice would be at protecting your security online, IF YOU FOLLOWED IT. *(grid question)*
  Scale: 5 (Very effective), 4, 3, 2, 1 (Not at all), I don't know
  - Use anti-virus software
  - Install the latest operating system updates
  - Turn on automatic software updates
  - Update applications to the latest version
  - Clear your Web browser cookies

- *(Non-experts only)* (optional) Please use this space to clarify any of the above. *(open-ended)*
- *(Non-experts only)* For each of the following pieces of advice, please rate on a scale from 1 to 5 how LIKELY YOU WOULD BE TO FOLLOW the advice, if you heard it would help protect your security online. *(grid question)*
  Scale: 5 (Very likely), 4, 3, 2, 1 (Not at all), I don't know
  - Use anti-virus software
  - Install the latest operating system updates
  - Turn on automatic software updates
  - Update applications to the latest version
  - Clear your Web browser cookies
- *(Non-experts only)* (optional) Please use this space to clarify any of the above. *(open-ended)*
- *(Experts only)* For each of the following pieces of advice, please rate on a scale from 1 to 5 how good (in terms of both EFFECTIVE at keeping the user secure, as well as REALISTIC that the user can follow it) you think they are at protecting a non-tech-savvy user's security online. *(grid question)*
  Scale: 5 (Very good), 4, 3, 2, 1 (Not at all), I don't know
  - Use different passwords for each account
  - Use passwords that are not easy to guess
  - Don't write down passwords on paper
  - Save your passwords in a local file on their computer
  - Use a password manager (e.g., 1Password, LastPass)
  - Write down passwords on paper
- *(Experts only)* (optional) Please use this space to clarify any of the above. *(open-ended)*
- *(Non-experts only)* For each of the following pieces of advice, please rate on a scale from 1 to 5 how EFFECTIVE you think the advice would be at protecting your security online, IF YOU FOLLOWED IT. *(grid question)*
  Scale: 5 (Very effective), 4, 3, 2, 1 (Not at all), I don't know
  - Use different passwords for each account
  - Use passwords that are not easy to guess
  - Don't write down passwords on paper
  - Save your passwords in a local file on their computer
  - Use a password manager (e.g., 1Password, LastPass)
  - Write down passwords on paper
- *(Non-experts only)* (optional) Please use this space to clarify any of the above. *(open-ended)*
- *(Non-experts only)* For each of the following pieces of advice, please rate on a scale from 1 to 5 how LIKELY YOU WOULD BE TO FOLLOW the advice, if you heard it would help protect your security online. *(grid question)*
  Scale: 5 (Very likely), 4, 3, 2, 1 (Not at all), I don't know
  - Use different passwords for each account
  - Use passwords that are not easy to guess
  - Don't write down passwords on paper
  - Save your passwords in a local file on their computer
  - Use a password manager (e.g., 1Password, LastPass)
  - Write down passwords on paper
- *(Non-experts only)* (optional) Please use this space to clarify any of the above. *(open-ended)*
- *(Experts only)* For each of the following pieces of advice, please rate on a scale from 1 to 5 how good (in terms of both EFFECTIVE at keeping the user secure, as well as REALISTIC that the user can follow it) you think they are at protecting a non-tech-savvy user's security online. *(grid question)*
  Scale: 5 (Very good), 4, 3, 2, 1 (Not at all), I don't know
  - Check if the website you're visiting uses HTTPS
  - Be skeptical of everything when online
  - Be suspicious of links received in emails or messages
  - Visit only websites you've heard of
  - Use two-factor authentication for your online accounts
- *(Experts only)* (optional) Please use this space to clarify any of the above. *(open-ended)*
- *(Non-experts only)* For each of the following pieces of advice, please rate on a scale from 1 to 5 how EFFECTIVE you think the advice would be at protecting your security online, IF YOU FOLLOWED IT. *(grid question)*
  Scale: 5 (Very effective), 4, 3, 2, 1 (Not at all), I don't know

  - Check if the website you're visiting uses HTTPS
  - Be skeptical of everything when online
  - Be suspicious of links received in emails or messages
  - Visit only websites you've heard of
  - Use two-factor authentication for your online accounts
- *(Non-experts only)* (optional) Please use this space to clarify any of the above. *(open-ended)*
- *(Non-experts only)* For each of the following pieces of advice, please rate on a scale from 1 to 5 how LIKELY YOU WOULD BE TO FOLLOW the advice, if you heard it would help protect your security online. *(grid question)*
  Scale: 5 (Very likely), 4, 3, 2, 1 (Not at all), I don't know
  - Check if the website you're visiting uses HTTPS
  - Be skeptical of everything when online
  - Be suspicious of links received in emails or messages
  - Visit only websites you've heard of
  - Use two-factor authentication for your online accounts
- *(Non-experts only)* (optional) Please use this space to clarify any of the above. *(open-ended)*
- *(Experts only)* For each of the following pieces of advice, please rate on a scale from 1 to 5 how good (in terms of both EFFECTIVE at keeping the user secure, as well as REALISTIC that the user can follow it) you think they are at protecting a non-tech-savvy user's security online. *(grid question)*
  Scale: 5 (Very good), 4, 3, 2, 1 (Not at all), I don't know
  - Don't click on links that people or companies you don't know send you
  - Don't enter your password when you click on a link in an email and that link takes you to a website that asks for your password
  - Pay attention when taking online surveys. We appreciate your input. To let us know you're paying attention, select four for this response
  - Look at the URL bar to verify that you are visiting the website you intended to
  - Don't open email attachments from people or companies you don't know
- *(Experts only)* (optional) Please use this space to clarify any of the above. *(open-ended)*
- *(Non-experts only)* For each of the following pieces of advice, please rate on a scale from 1 to 5 how EFFECTIVE you think the advice would be at protecting your security online, IF YOU FOLLOWED IT. *(grid question)*
  Scale: 5 (Very effective), 4, 3, 2, 1 (Not at all), I don't know
  - Don't click on links that people or companies you don't know send you
  - Don't enter your password when you click on a link in an email and that link takes you to a website that asks for your password
  - Pay attention when taking online surveys. We appreciate your input. To let us know you're paying attention, select four for this response
  - Look at the URL bar to verify that you are visiting the website you intended to
  - Don't open email attachments from people or companies you don't know
- *(Non-experts only)* (optional) Please use this space to clarify any of the above. *(open-ended)*
- *(Non-experts only)* For each of the following pieces of advice, please rate on a scale from 1 to 5 how LIKELY YOU WOULD BE TO FOLLOW the advice, if you heard it would help protect your security online. *(grid question)*
  Scale: 5 (Very likely), 4, 3, 2, 1 (Not at all), I don't know
  - Don't click on links that people or companies you don't know send you
  - Don't enter your password when you click on a link in an email and that link takes you to a website that asks for your password
  - Pay attention when taking online surveys. We appreciate your input. To let us know you're paying attention, select four for this response
  - Look at the URL bar to verify that you are visiting the website you intended to
  - Don't open email attachments from people or companies you don't know
- *(Non-experts only)* (optional) Please use this space to clarify any of the above. *(open-ended)*
- What is your gender? *(multiple-choice)*
  - Female
  - Male
  - Transgender
  - I prefer not to answer
  - Other
- What is your age? *(multiple-choice)*
  - 18-24 years old

  - 25-34
  - 35-44
  - 45-54
  - 55-64
  - 65 or older
  - I prefer not to answer
- What is the highest degree or level of school that you have completed? *(multiple-choice)*
  - Professional doctorate (for example, MD, JD, DDS, DVM, LLB)
  - Doctoral degree (for example, PhD, EdD)
  - Masters degree (for example, MS, MBA, MEng, MA, MEd, MSW)
  - Bachelors degree (for example, BS, BA)
  - Associates degree (for example, AS, AA)
  - Some college, no degree
  - Technical/Trade school
  - Regular high school diploma
  - GED or alternative credential
  - Some high school
  - I prefer not to answer
  - Other
- *(Experts only)* How many total years of experience do you have in computer security? *(multiple-choice)*
  - At least 1 but less than 5 years
  - At least 5 but less than 10 years
  - At least 10 but less than 15 years
  - 15 years or more
  - None
- *(Experts only)* What is your current job role? For example, Network Security Engineer, Penetration Tester *(open-ended)*
  - Researcher
  - Principal Architect
  - IT Strategist
  - CEO
  - Manager
  - Security Engineer
  - Engineer
  - Other
- *(Experts only)* Which of the following best characterizes your workplace? *(multiple-choice)*
  - University
  - Corporate research lab
  - Industry
  - Government
  - Self-employed
  - Other
- *(Experts only)* In what country do you work? *(multiple-choice)*
  - Australia
  - Canada
  - Germany
  - India
  - United Kingdom
  - United States
  - Other
- *(Experts only)* In what state do you work? *(open-choice)*
- *(Non-experts only)* Which describes your current employment status? *(multiple-choice)*
  - Employed full-time
  - Employed part-time

  - Self-employed
  - Care-provider
  - Homemaker
  - Retired
  - Student - Undergraduate
  - Student - Masters
  - Student - Doctoral
  - Looking for work / Unemployed
  - Other
- *(Non-experts only)* What is your occupation? *(open-ended)*
- *(Non-experts only)* What is your Mechanical Turk Worker ID? *(open-ended)*
- (Optional) Is there anything else you'd like to add or clarify? *(open-ended)*

A Human Capital Model for Mitigating Security Analyst Burnout

Sathya Chandran Sundaramurthy
Kansas State University
sathya@ksu.edu

Alexandru G. Bardas
Kansas State University
bardasag@ksu.edu

Jacob Case
Kansas State University
jacobcase94@ksu.edu

Xinming Ou
Kansas State University
xou@ksu.edu

Michael Wesch
Kansas State University
mwesch@ksu.edu

John McHugh
RedJack, LLC.
john.mchugh@redjack.com

S. Raj Rajagopalan
Honeywell ACS Labs
siva.rajagopalan@honeywell.com

ABSTRACT

Security Operation Centers (SOCs) are being operated by universities, government agencies, and corporations to defend their enterprise networks in general and in particular to identify malicious behaviors in both networks and hosts. The success of a SOC depends on having the right tools, processes and, most importantly, efficient and effective analysts. One of the worrying issues in recent times has been the consistently high burnout rates of security analysts in SOCs. Burnout results in analysts making poor judgments when analyzing security events as well as frequent personnel turnovers. In spite of high awareness of this problem, little has been known so far about the factors leading to burnout. Various coping strategies employed by SOC management such as career progression do not seem to address the problem but rather deal only with the symptoms. In short, burnout is a manifestation of one or more underlying issues in SOCs that are as of yet unknown. In this work we performed an anthropological study of a corporate SOC over a period of six months and identified concrete factors contributing to the burnout phenomenon. We use *Grounded Theory* to analyze our fieldwork data and propose a model that explains the burnout phenomenon. Our model indicates that burnout is a human capital management problem resulting from the cyclic interaction of a number of human, technical, and managerial factors. Specifically, we identified multiple vicious cycles connecting the factors affecting the morale of the analysts. In this paper we provide detailed descriptions of the various vicious cycles and suggest ways to turn these cycles into virtuous ones. We further validated our results on the fieldnotes from a SOC at a higher education institution. The proposed model is able to successfully capture and explain the burnout symptoms in this other SOC as well.

Symposium on Usable Privacy and Security (SOUPS) 2015, July 22–24, 2015, Ottawa, Canada.

1. INTRODUCTION

With an increase in cyber threats, corporations and government agencies alike are establishing dedicated monitoring stations called security operation centers (SOCs). An organization can decide to build its own SOC or outsource the monitoring to managed operational service providers. The key component of any SOC, in-house or managed, is training and staffing of security analysts. Although tools and processes improve the efficiency of operations, it is the security analysts who make the final decision when analyzing a threat. Hence it is imperative for a SOC to spend adequate resources in developing and maintaining an effective team of security analysts.

In our work we wanted to find answer to an important question — How to maintain a capable and enthusiastic analyst workforce? The problem bears considerable similarity to the human capital model in economics. The Human Capital theory [13], first postulated by Adam Smith, holds that the investment made in education and training of individuals in a society is a resource in itself, more important than capital and natural resources. Security analysts are the human capital of a SOC and proper investment in their continuous improvement is key for efficient operation.

Unfortunately SOCs have been plagued by high analyst turnover due to burnout [10]. Burnout refers to diminished interest in work and is characterized by exhaustion, cynicism and inefficacy [9]. Burnout in SOCs usually results in a high analyst turnover leading to frequent hiring and training of new analysts. A white paper from Hewlett-Packard (HP) [7] points out that the life-time of a security analyst is between 1-3 years. Moreover, the volatile nature also makes it hard for analysts to know each other well, thus affecting team camaraderie, which eventually affects how the entire team responds to security incidents.

In spite of the burnout problem being well recognized, little to nothing is known about the concrete factors that cause the burnout. If the real reasons behind this issue are not identified, we will be only addressing the symptoms and not the actual problem. In order to understand challenges in a SOC environment we first had to find a way to interact with the SOC analysts. SOC analysts typically work under high stress; culturally SOCs are sensitive about talking to outsiders – such as security researchers – about operational issues. To get visibility into operational issues affecting the analysts, the research has to satisfy two requirements: (1) cause minimum interruption (and only when necessary) to the analysts; (2) gain the trust of the entire SOC so that the *real* reasons for burnout are explored.

With the above set of goals we adopted an anthropological approach to study SOC environments. Using an anthropological approach helps us attain the perspective of the analyst on exhaustion and burnout. Security analysts are typically consumed by the routines of their job that they have no time to reflect on the social issues in the SOC. Anthropology also allows the researcher to step in and out of the shoes of an analyst which helps in understanding the complex interactions not attended to by the participants.

A computer science graduate student trained in fieldwork methods by an anthropologist took up a job as a security analyst for six months in a corporate SOC. The corporation is a major information technology (IT) products and services provider headquartered in the United States. The SOC is monitoring the enterprise's network 24x7x365 for security threats. The fieldworker went through the whole new-analyst training process and at the end of it, he was able to do a junior analyst's job. Through the embedding process he earned the trust of the analysts (specific instances that led to building the trust are discussed in further sections) and also was able to simultaneously perform the research with minimum to no interruption to operations.

Daily observations of SOC activities were written down in a digital document for six months. The fieldnotes were analyzed after the fieldwork using a Grounded Theory approach. Through our analysis we found that to mitigate analyst burnout, SOCs have to pay special attention to the interaction of *human capital* with three other factors—*automation, operational efficiency*, and *metrics*. Our analysis yielded a model for human capital management in the SOC and we suggest a number of ways to mitigate analyst burnout, which is the focus of this paper. To the best of our knowledge this is the first study of the burnout problem in a SOC environment using anthropological methods.

2. ANTHROPOLOGICAL APPROACH TO STUDYING SOCS

We started our research, two years ago, with the broader goal of understanding how security analysts do their job and what happens inside a SOC [15]. Before this, our attempts towards this goal were through focused interviews with system administrators and security analysts. This approach was very hard to pursue over time as system administrators and security analysts worked under high pressure and had limited time to talk to the researchers. The other major obstacle was the issue of trust. As security monitoring is considered a sensitive job there is always some hesitation in the minds of the analysts when talking to researchers who are considered "outsiders." After years of failed attempts to truly understand security operations, we discovered that methodologies from anthropology, specifically cultural anthropology, are very relevant to studying this problem.

Anthropology is a discipline where researchers used to spend extended period of time, typically one to three years with an indigenous population. The goal of an anthropologist is to document and make explicit the various cultural aspects of the population as objective as possible. They do this through a research method called *participant observation*, where the researcher becomes one among the members of the society under study. Participant observers go through the same or similar challenges as the members of the group being observed and try to gain an empathetic perspective on the views and practices in that society.

Gaining the trust of the members of the society is a critical aspect for anthropological study. Clifford Geertz in his book Deep Play [6] talks about the experiences of studying cockfights in Bali, Indonesia. He reports that he and his fieldwork partner, both researchers, remained invisible to the villagers until one day they had to run away along with a group of Balinese people from an illegal cockfight when the police arrived to stop the fight. After this specific incident, the researchers were considered as ones among them by the Balinese people. Thus for an outsider to get accepted into a community they have to perform the same activities as the rest of the members. The acceptance leads to establishment of the trust which facilitates knowledge sharing and thus cultural understanding by the researcher.

An anthropological approach to study SOCs means that researchers become analysts and gain the acceptance – hence the trust – of the SOC members, even if it implies spending significant amount of time in the SOC. Towards this goal, our research team consists of an anthropology professor who helped train a graduate student in Computer Science in participant observation methods. One of the key elements of training was about performing reflections on daily observations in the SOC. Without periodic reflections the observations will remain just incidents without their cultural significance being understood. The student also attended a course offered by our anthropologist. This training ensured the student learned the basics of anthropological methods before conducting the fieldwork.

One important aspect of anthropological research is that it helps identify problems that the researchers may not have been aware of. The burnout problem, which includes the phenomenon and the associated causes/effects described in this paper, is one of many problems we discovered in the SOC, *without knowing them or looking for them a priori.* The phenomenon itself was known to SOC operators, but the causes and effects were not clear and there had been no systematic study of this problem in the published literature. There are also other problems we have discovered and analyzed but we will not present those in this paper in order to maintain focus and present a cogent description.

3. ETHICS AND PARTICIPANT SAFETY

The fieldworkers and the SOC analysts who were observed can be considered human subjects. Prior to starting the formal fieldwork, we obtained the appropriate IRB approvals. All participants were asked to sign an "informed consent form" approving of their participation in our research. The consent form explained clearly to the participants the goal of our research, what we expected from them, and how the fieldwork data will be used.

We took efforts to protect the privacy of the participants such as by not using real names of analysts during research discussions and also by not revealing opinions of one analyst to another. We also followed the standard practice in anthropology where the fieldnotes are accessible only to the

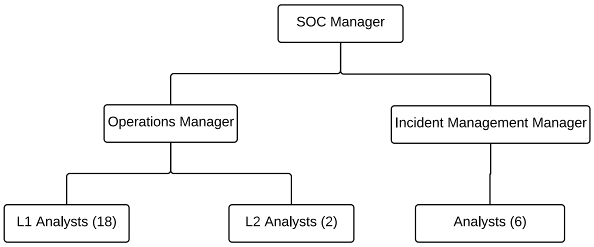


Figure 1: Organizational Chart for the corporate SOC

fieldworker who writes them. We can safely say that our research did not raise any ethical concerns for the participants.

4. FIELDWORK SETUP

Our fieldwork was conducted at a SOC run by and for a major IT products and services provider headquartered in the United States. The mission of the SOC was to monitor the network and hosts therein to identify and mitigate security threats. The network was spread all over the world and there were about 300,000 devices online on any given day. The SOC was the only monitoring station for cyber security for the corporate network spread across the globe.

The organizational layout of the SOC is shown in Figure 1. There were two different teams within the SOC: (1) operations; (2) incident management (IM). The operations team analysts were divided into two categories: Level 1 (L1) and Level 2 (L2). The L1 analysts work four days a week in 10 hour shifts while the L2 analysts work 5 days a week in 8 hour shifts. L1 analysts are the first line of defense monitoring the Security Information Event Management (SIEM) console for any possible (attempted) security breaches. The L2 analysts play more of a managerial and mentoring role for the L1 analysts. Their main job is to provide operational visibility for higher management through metrics and reports. At the time of our fieldwork there were 2 L2 and 18 L1 analysts in the operations team headed by an Operations Manager.

The IM team consisted of six analysts led by a manager. Of the six analysts two were off-site working remotely. The IM team handles incidents that are escalated from the operations team requiring in-depth analysis. The fieldworker spent three months as an L1 analyst in the operations team and the remaining three months as an incident responder in the IM team.

5. SOC DEMOGRAPHY

There were eighteen L1 and two L2 analysts in the operations team of the SOC. 15% (3 out of 20) of the analysts were female and 85% (17 out of 20) were male. 35% (7 out of 20) of the analysts had previously worked in a SOC and for 65% (13 out of 20) of them this was their first SOC experience. Among the analysts who were new to the role of SOC analyst 46% of them (6 out of 13) had worked in a job related to IT services. Of the six analysts in the IM team, one was female and five were male. Two worked previously as consultants performing forensic analysis and two had been system administrators in their previous jobs.

6. GAINING TRUST AND ACCEPTANCE

As described in Section 2 gaining acceptance of the SOC analysts and earning their trust was our initial goal. Trust of L1 analysts was gained by working alongside with them on the SIEM console processing alerts, similar to experiences of Geertz as described in Section 2. However, saturation occurred only after a few weeks of visibility into the SOC operations, as the fieldworker was just following the procedures. The procedures were very static and by following them he was consumed by the routines of an L1 analyst. He then started to identify high-severity threats that could have not been discovered by following the procedures. A number of teams were engaged in solving those high-severity cases through which he obtained the attention of senior analysts and SOC managers – he was not *invisible* anymore. At this stage everyone in the SOC – junior and senior analysts including managers – started to feel comfortable with the fieldworker.

7. DATA COLLECTION

The daily observations of SOC activity were documented in a digital document. The goal was to document every activity in the SOC without any premeditation on what to document. Often, a theme might emerge as the observations were being logged. In those situations, focused interviews were conducted with the analysts and managers to better understand the emerging concept. Taking notes while engaging in a conversation was avoided in order to focus more on the interaction. The details of the communication were transcribed into notes as soon as possible after the conversation. At the end of the six-month fieldwork, we had 85 pages of fieldnotes stored in a word-processor document.

8. GROUNDED THEORY APPROACH TO DATA ANALYSIS

Our goal in analyzing the fieldnotes was to uncover the different cultural aspects in an operational SOC environment. Grounded Theory Method (GT) [14] is a research methodology from the social sciences used to construct a theory from qualitative data such as interview transcripts, images, and video of participants. The outcome of GT-based analysis of data is a model or theory that explains the social connections represented by the data. Since we wanted to understand the burnout problem *through* fieldwork data, GT seemed to be the most appropriate analysis method.

GT analysis requires one to follow the steps of *open*, *axial*, and *selective* coding. In the open coding process labels are assigned to units of observed data. The researcher tries her best to assign codes that are not descriptive but analytical capturing the intent behind the observations. During axial coding the individual open codes are grouped into categories and sub-categories. The goal here is to understand the inter and intra categorical relationships. Finally, in the selective coding process the core category and the main concern of the participants is identified. There are a number of variations of the GT methodology and we used the one proposed by Strauss and Corbin due to its emphasis on theory development.

8.1 Open Coding

The goal of open coding was to assign short labels to fieldwork notes. Unlike other qualitative works where coding is performed by multiple people, the fieldnotes in our work were coded only by the fieldworker. In anthropological research one does not share the fieldnotes with anyone else due to privacy reasons; this is a standard practice in anthropology. Therefore, this was the rationale behind our decision to use the fieldworker as the only coder. The list of codes as they emerged were maintained in another document called the *code book*. When assigning a code to an observation, we first checked against the code book to see if any of the existing codes could be reused. If none of them were relevant to the observation at hand, a new code was generated and the process continued until we coded the entire document. The fieldwork notes made over a period of six months were 85 pages long. Below are a few examples of the codes that emerged through the open coding process.

An IM analyst expressed frustration about his current job:

> "I wanted to work in an environment where there will be continuous learning and I have started to feel that I am not learning anything new in my current job. In fact, I took the current job hoping to analyze malware every day and learn more in that process. I feel that the SOC currently is not doing any real threat detection which in turn is limiting my opportunities for learning. I have decided in my life, to spend a significant amount of time for the improvement of my career. Now I feel bad that my commitment is not paying off."

We assigned the code *lack of growth* to the observation above. This code captures the fact that the analyst felt a lack of intellectual growth, which is a major issue in maintaining a good morale.

The fieldworker and an L1 analyst were discussing an operational scenario:

> "I suggested to the analysts: why not we try to get access to the controller and lookup the data ourselves. One of the analysts said: access to the domain controller is too risky to be given to analysts."

We assigned two codes for this observation, *liability* and *restricted empowerment*. The SOC managers will be responsible if the analysts misuse the credentials and hence they chose to provide only limited privileges on the domain controller.

The open coding process was repeated multiple times. Sometimes there were too many codes which made it hard to proceed and other times the codes were not analytical enough.

8.2 Axial Coding

The goal of axial coding was to group the different codes obtained through the open coding process into categories. The categories emerged simultaneously as we were doing the open coding process. In the initial attempt of the axial coding process, where we coded around 50 percent of the fieldwork notes, we had around 10 categories. This initial attempt, which resulted in 10 categories, did not convey any useful information about the culture of the SOC. We then engaged in a few brainstorming sessions with our anthropologist. After that we repeated our coding process on the entire fieldwork notes that resulted in 4 categories at the end of axial coding. The most important result of the discussions was the identification of the various causal relationships between the four categories.

We followed the guidelines for axial coding proposed by Strauss and Corbin [14] who suggest to look for the following relationship between the codes:

- the phenomenon under study
- the conditions related to that phenomenon (context conditions, intervening structural conditions or causal conditions)
- the actions and interactional strategies directed at managing or handling the phenomenon
- the consequences of the actions/interactions related to the phenomenon

The description of codes and categories as they emerged during the early stage of axial coding process is shown in Table 1. In the end we identified *human capital, automation, operational efficiency, and metrics* as the major high-level categories.

8.3 Selective Coding

The goal of the selective coding process was to identify the *core category* and the *main concern*. The core category emerged out to be *human capital* and the main concern of the participants was the *development and maintenance of human capital*. In other words, the pressing issue in the SOC was to keep the analysts motivated at work. *Theoretical sampling* was performed when we looked for new data from the fieldnotes that supported the core category. The relationship framework between categories obtained as a result of the axial coding process was altered to focus more on the core category–the human capital.

As a result of selective coding we observed the existence of a number of *vicious cycles* connecting the core category with the rest. The final outcome was a model that explains the analyst burnout phenomenon. In the following section we explain the model in details by highlighting the effect of the vicious cycles on analyst morale and provide suggestions to turn them into virtuous ones.

9. A MODEL FOR SOC ANALYST BURNOUT

The grounded theory based analysis of our fieldwork data yielded us a model that explains the burnout of SOC analysts. In summary, the model shows that burnout occurs due to a cyclic interaction of Human Capital with the following three categories:

- Automation
- Operational Efficiency
- Management Metrics

We first describe the notion of Human Capital for the SOC – what it is and the ways it is developed. We then describe the influence of the above three categories on human capital management focusing on specific interactions that cause burnout.

Table 1: Categorization of codes at an early stage of axial coding process

Code	Meaning
Analyst Morale	
Inadequate compensation (perception)	Analyst perceives that she/he is not adequately compensated for their efforts.
Lack of growth	Analyst feels that she/he is not learning on their job.
Detailed procedures	Step by step procedures are too mundane.
Imposement	Analysts are given tasks to do without consultation.
Restricted empowerment	Inadequate privilege or access for an analyst to do their job.
Automation	
Increased workload	High event load is a good incentive for automation.
Lack of reflections	No review of procedures to look for possible automation.
Liability	Fear of responsibility hinders automation.
Operational Efficiency	
Restricted empowerment	Inadequate privilege or access for an analyst to do their job.
Poor intelligence	Incomplete information from sources outside the SOC.
Lack of cooperation	Lower efficiency due to inter-operation issue between teams.
Inadequate context	Low efficiency due to contextual information surrounding an alert.
Lack of clarity	Incomplete understanding of operational processes due to miscommunication.
Teams in silos	Inadequate communication between teams leading to inefficiencies.
Analyst Burnout	
Superficial briefings	Exhausted analysts stop providing detailed operational updates.
Cherry picking	Burned out analysts pick specific events to analyze.
Metrics	
Management visibility	Management uses metrics as a way to gain visibility into SOC operations.
Tools and workflow	Metrics influence the workflow and tools used in the SOC.
Perception	Metrics affect the perception the management has about the usefulness of the SOC.

9.1 Human capital

Human capital, in the context of a SOC, refers to the knowledge, talents, skills, experience, intelligence, training, judgment, and wisdom possessed by individual analysts and the team as a whole. Human capital can also be defined as the collective and individual intellectual stock of a SOC. Proper development and management of the human capital is crucial for the success of SOC operations. Mismanagement of human capital affects the morale of the analysts which in turn reduces operational efficiency. Our model indicates that there are four factors that influence the creation and maintenance of efficient human capital as shown in Figure 2.

- Skills
- Empowerment
- Creativity
- Growth

Next, we describe how the interaction between these factors might either lead to an effective team or an inefficient group of burned-out analysts.

9.1.1 Skills

Security analysts need to possess the right skills to do their job. The skill set of analysts vary depending on a number of factors such as education and prior experience. The dynamic nature of security threats means the analysts have to undergo periodic training. SOCs send their analysts to paid training workshops such as those organized by SANS. The SOC also organized table top exercises for analysts to make sure they can respond to a crisis situation. Analysts were also encouraged to engage in peer training through presentations and hands-on exercises. For example, an L1 analyst who was a SOC analyst before was demonstrating a threat discovery tool. The tool took large volumes of alert information from a SIEM, summarized and provided a graphical interface to help analysts do threat discovery. The IM team also conducted training sessions on tools such as Volatility [5] and Cuckoo [4] for the operations analysts.

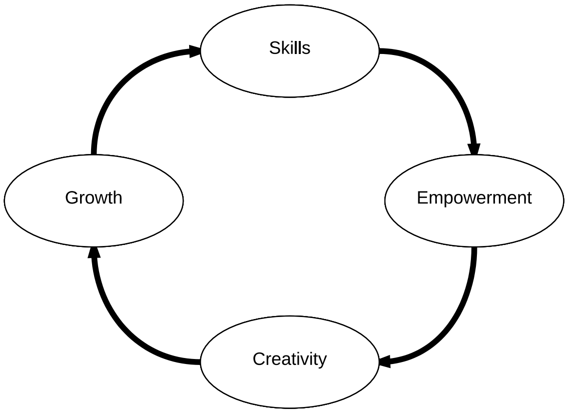


Figure 2: Human Capital Cycle

To summarize, development and continuous improvement of analysts' skill-set is an important aspect of human capital management. If the analysts are not adequately skilled it affects their confidence in dealing with the security alerts. Over time the lack of confidence will manifest itself as frustration, especially when their job demands them to do more than their skills level permits. SOC managers should make sure that analysts receive periodic and adequate training.

9.1.2 Empowerment

Analysts feel that they need to be adequately empowered to perform their job efficiently. An incident observed in the IM team sheds light on the importance of empowerment:

> An IM analyst expected that he be given privileged access on end user machines so that he can perform live monitoring for malicious activity on the suspicious hosts. The analyst was denied this access by the management. This led to frustration as the analyst felt that he was not able to perform his tasks efficiently.

We observed that analysts feel empowered when they were allowed to author new threat detection content or contribute to new tools development. Analysts feel enthusiastic when they see the impact of their effort in one form or the other. An L1 analyst expressed thus:

> "In my previous job as a SOC analyst, access was restricted to the alert console and what we could do was very limited. I like that in this SOC, analysts are asked to give periodic feedback on the tools and procedures. Also, I like the fact that we are encouraged to suggest new threat detection rules to the engineering team."

The skill level of the analysts influences the level of empowerment the management is willing to grant them as indicated by the causality between skills and empowerment in Figure 2. For example, only skilled analysts are trusted to be careful and are provided privileged access to user accounts. We also observed that even if analysts are highly skilled they may not always be empowered.

An IM analyst (highly skilled compared to L1 or L2 analysts) pointed out that his manager was reluctant to provide privileged access to his team due to *liability* reasons:

> "He is afraid that we, IM analysts, have accounts on social networking websites such as LinkedIn and might fall for phishing scams. He is worried that malicious entities might send us targeted emails. If one of us gets compromised, privileged credentials might be exposed and then the whole corporate network will be at risk. He does not want to give us administrator access on user accounts for this reason."

Empowerment plays a major role in boosting the morale of the analysts and SOC managers have to keep in mind this important factor. Highly skilled analysts might feel handicapped in their job when they are not adequately empowered by the management. More research is needed to understand how to provide the right amount of empowerment to the analysts while at the same time minimizing risk for the management.

9.1.3 Creativity

Creativity refers to the ability of analysts to handle an operational scenario that differs significantly from those they have encountered so far. The human capital model in Figure 2 indicates that empowerment directly affects analysts' creativity. If an analyst is adequately empowered, the analyst will be more willing to deviate from the operational norms. Usually, norms are written down procedures which severely inhibit creativity if analysts are not empowered. Lack of creativity will lead to analysts just executing the procedures failing to react appropriately to a novel operational scenario.

Another observation highlights the impact of "lack of creativity" on operations:

> "An analyst encountered an operational scenario where he had to email a member of a business unit to validate an alert but was very hesitant to proceed. After waiting for a while he contacted a senior analyst and asked him for advice on how to proceed. The junior analyst specifically said that he does not know how to proceed as this scenario was not covered by any of the procedures."

On the other hand, members of the IM team were more creative than L1 analysts. We observed that IM analysts were constantly trying to learn new technologies and this behavior was encouraged by their manager. The IM analysts were empowered as they were more skilled than the L1 analysts in the SOC. Thus one can see the causal influence between skills, empowerment, and creativity.

We also observed that the lack of variation in the operational tasks lead to lower creativity levels. The daily alerts received by the SOC are very much alike, which means an analyst has to take the same response steps for each of the received alerts. To summarize, creative development is an important aspect of human capital management. Empowerment plays an important role in ensuring creativity. The SOC management also must make sure that they find ways to engage their analysts in creative activities when the operational tasks get repetitive.

9.1.4 Growth

Growth in the context of the SOC refers to increase in the intellectual capacity of the analysts. Learning on-the-job is one of the dominant ways through which an analyst achieves growth. An analyst, by handling different types of security incidents, learns new skills and improves her knowledge on security analysis. This learning improves her morale as it gives a sense of purpose and accomplishment. As it can be observed in Figure 2, growth is directly influenced by creativity. Mundane daily activities will lead to lower creativity development. Lower creativity means the analyst will use the same set of skills everyday in the job which in turn inhibits intellectual growth. Growth also occurs through learning from role models – an analyst learns from a more experienced one.

We observed that highly empowered analysts sometimes were not satisfied with their growth because of lack of creativity in their job. During a conversation with an IM analyst, who had relatively higher empowerment than an L1 or L2 operational analyst, the analyst expressed his concerns over stagnation of growth:

> "I took this job as an IM analyst because I was excited about analyzing a *variety* of malware everyday but I am not able do it as the SOC is not doing real security monitoring. I also do not have anybody on the team to look up to and learn from. Everyone is less skilled than myself

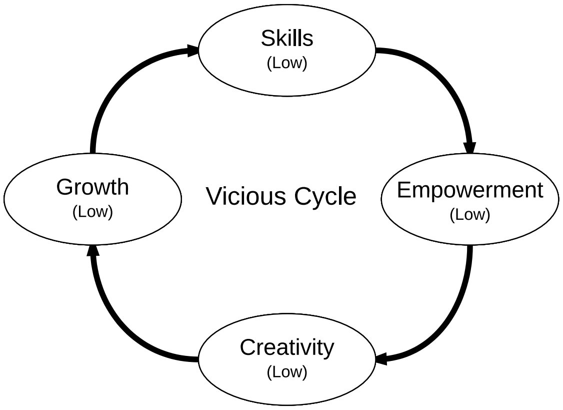


Figure 3: Human Capital Vicious Cycle

> and that means I have to teach them all the time. I love teaching new skills to other analysts but it affects me when I cannot learn from anybody else."

Another possibility through which growth occurs is through career progression.

> An L1 analyst in the SOC was planning to quit his job to pursue an analyst role at another SOC presumably requiring more skills. The managers realized this and offered him a job in the IM team team within the SOC. He accepted the job deciding to stay which was beneficial for the SOC management as they were able to retain an experienced analyst.

If an analyst has outgrown her current role – which may be because they stayed considerably long in the job and reached a saturation point in their learning process – then the manager can reassign her to another position that is more challenging. This will ensure that the learning process never stops ensuring growth and good morale. Unfortunately, this is not a solution that will work all the time since there are only a few positions available at any given time to reassign analysts.

Growth, through any of the suggested means, enhances the skill-set of the analysts.

9.1.5 Burnout trajectory and avoidance

As long as a positive causality among the factors – skills, empowerment, creativity, and growth – exists, the morale of the analysts will remain high. Burnout occurs when a SOC gets stuck in a vicious cycle connecting those factors. For instance, the SOC management hires entry level (not highly skilled) analysts due to budget constraints. These analysts will not be empowered enough as the managers do not trust the abilities of their analysts. This lower empowerment will lead to lower creativity, which will in turn lead to lower growth and skills. Since the skill level of analysts remains the same (very low) this will again lead to low empowerment, creativity, and growth. If this continues eventually the analysts will be burned out as they will start to feel that they are not accomplishing anything in their job–in other words, there is no growth, and the repetitiveness of the job exhausts the analysts. The vicious cycle is illustrated in Figure 3. We call it a vicious cycle because the management wants their analysts' skill set to progressively get better through on-the-job experience, but the negative causality among the four factors deteriorates the human capital of the SOC.

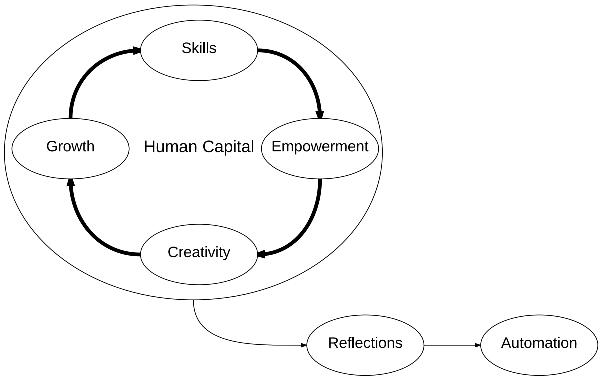


Figure 4: Automation

Although the analysts are less skilled, the management can take some risk and empower the analysts–perhaps gradually. After a few positive cycles in the human capital management cycle (Figure 2) the analysts will gain more skills. The SOC managers will now trust the skilled analysts and empower them with privileges. This will in turn encourage creativity and growth–turning the cycle into a virtuous one. It is possible that even after the cycle is taken in a positive direction the job of an analyst can become too repetitive. One way to deal with the repetitiveness is by providing new opportunities for analysts to stay creative. If the management finds that the analyst has completely outgrown her position, efforts should be taken to find a more challenging role for the person, ensuring positive growth. The bottom line is that to avoid analyst burnout SOC management must be careful not to get caught in the vicious cycle of human capital and be watchful for any signs leading to such trajectory.

9.2 Automation

Automation in a SOC refers to software tools that aid analysts' job and improve operational efficiency. Automation includes complex software such as Security Information Event Management (SIEM) to simple scripts written in Python or Ruby. Software tools are extremely efficient in performing repetitive tasks. Repetitiveness leads to lower creativity as we noted in the discussion on human capital. By automating repetitive tasks, skilled human analyst will have more freedom to engage in more sophisticated investigations.

During the fieldwork we discovered that effective automation takes place only if a process called *reflection* takes place within and among the analysts as shown in Figure 4. Reflection in a SOC is usually done by periodically reviewing the procedures with the goal of identifying operational bottlenecks that can benefit from automation.

Here is an example of reflection leading to automation from our fieldwork:

> Every time an end-user device is identified to be infected with certain classes of malware, the

> standard remediation measure in the SOC was to ask the user to reimage their device with a clean operating system (OS) image. The instructions were written down in an email template and the only data that varied from one user to another was the username and hostname of the device. Other than that, it was the same email and there were hundreds of such emails sent everyday manually. One day an L2 analyst *realized* that this process is cumbersome and wrote down a script that will automatically fetch the username, hostname, and email address of the infected device/user to send a mass mail.

Automation through reflection can only occur if the analysts are *empowered* and *incentivized* to do so. In the example mentioned above the analyst automated the repetitive process due to his own personal interest outside his working hours. He said that the management did not want him to consider automation at the same priority level as report generation.

Reflections are beneficial and practically easier if started earlier. An analyst pointed out the difficulty arising from delaying this process based on his experience:

> "At one point we had procedures written down for everything and analysts were starting to feel like robots performing the same tasks everyday. We did not have any reviews to refine the processes as at one point nobody was even documenting them properly."

To automate complex tasks the analysts have to work with software developers—another form of empowerment. By reflecting on the operational procedures the analysts provide requirements for tools to the developers. The developers develop the tool based on the requirements from the analysts through multiple development iterations. This is called the analyst-developer tool co-creation approach.

We are convinced that this actually works as the fieldworker engaged in a co-creation tool development process at the SOC. The research team of the company developed an algorithm to identify malware from DNS request and response traffic. An initial prototype of the tool was developed by the researchers through collaboration with software developers. The prototype was then deployed in the SOC and the fieldworker was asked to provide feedback on the usability and effectiveness of the tool. The fieldworker and other analysts observed a number of mismatches between the functionality of the tool and the workflow of the SOC. There were weekly meetings during when *actual* workflow requirements of the SOC were conveyed to the researchers. A new version of the tool would then be deployed with the feature requests implemented. Eventually the tool turned out to be very useful for the analysts in their investigations. This process continued even after our fieldwork as the workflow of a SOC is very dynamic. This experience showed that a better way to design tools for SOCs could be to engage the analysts and developers in a co-creation process.

It appears that automation serves two main purposes in enriching the human capital. First of all, analysts can engage in interesting and challenging investigation tasks if the repetitive tasks are automated. We also observe that the co-creation process, which results in automation, provides a platform for the analysts to express their creativity. SOC managers must pay attention to this important effect that automation has on human capital to mitigate burnout.

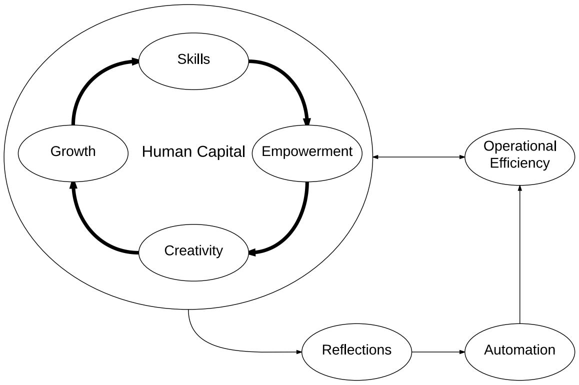


Figure 5: Operational Efficiency

9.3 Operational Efficiency

An efficient SOC will be able to leverage all its resources to detect and respond to threats in a timely manner. Since analysts make all the final decisions during operations, human capital has a direct influence on operational efficiency. We also observed that this relationship is bidirectional in that an efficient SOC positively influences the human capital. We see this two way relationship in Figure 5.

The direct causality from human capital to operational efficiency indicates the obvious fact the highly skilled and creative analysts make operations efficient. Human capital also affects efficiency in operations through automation via reflections. The resulting automation accelerates operations—especially in case of highly repetitive tasks. Here is an example for operational efficiency through automation from our fieldwork. An L1 analyst mentioned the following about creating incident tickets:

> "Case creation takes too much time. Filling in a ticket to locate hosts is the most demanding task. The fields are not fillable as you have to select the entry. There is a script written by an L2 analyst to automate this task. I need to give it a try."

On the other hand, the benefits resulting from operational efficiency in turn create a positive influence on the analysts. Most of the efficiency is achieved through automation by reflections. Reflections provide an opportunity for the analysts to exercise their creativity. This in turn helps in the growth of human capital. An inefficient SOC means reduced automation leading to analysts performing the repetitive tasks manually. This will lead to exhaustion and burnout of analysts eventually. In a vicious cycle, the SOC management could be spending resources on hiring highly skilled analysts who if not empowered to engage in reflections, will lead reduced automation. Operational efficiency suffers due to reduced automation. The inefficiency wears out the analysts as they have to manually perform tasks that could be effectively performed by software. The vicious cycle could be converted to a virtuous one by empowering analysts to facilitate automation through reflections.

9.4 Metrics

A SOC has to periodically measure its efficiency for a number of reasons:

- Measure employee efficiency for bonus considerations
- Measure and tune intrusion detection sensors
- Identify bottlenecks in operational procedures
- Most importantly, provide visibility into the SOC for the upper management

During the fieldwork we tried to understand the influence of metrics on the human capital. We observed that it is very challenging to come to light with good metrics for security operations. It is challenging because either the metrics are too technical making it hard for the management to measure their return-on-investment, or they are too managerial thereby failing to convey exactly the operational activities in the SOC. Some of the metrics that were automatically generated from the SIEM solution are shown in Table 2.

A SOC is an investment from a management perspective and SOC managers have to frequently communicate the benefit the company gets from such an investment. This is vital for the continued support from the upper management as one of the managers mentioned:

> "Corporations are very eager to start an operations center in-house and they fund SOC builders to establish one. After a while they stop seeing value in the SOC, shut it down and move it to managed services. A few years later they realize an in-house SOC is better for them and they redo the process all over. This is hard for guys like me as we spend more than a year establishing the SOC infrastructure and training analysts. A lot of effort goes into it. As SOC managers we need to keep communicating to the management how their investment in the SOC is justified."

Devising appropriate metrics is complicated by the fact that even the higher managers are not sure what shall be the right metrics. A senior manager once responded to an L1 analyst's question on what the higher management perceived as good metrics for SOC operations as thus:

> "I am not sure what the right metrics are and that is something I am working on. But I have some idea on what would be a good metric. If you tell me you processed some thousands of events over a month that does not tell anything interesting to me. But if you tell me a case where you engaged multiple teams–vulnerability management, red team, etc.–and how that resulted in the creation of new detection points–*e.g.* a new AV signature–or how it helped in creating new analytics that will be a good indication of what value I am getting from the SOC. Again, I do not know how you guys can give me the metrics but you have all the data and it is your job to come up with a good way to communicate your success stories."

The pressure for good metrics is relayed down to the SOC managers who in turn hand it down to the analysts. On one occasion, a SOC manager expressed frustration about the lack of good metrics for him to talk to his managers:

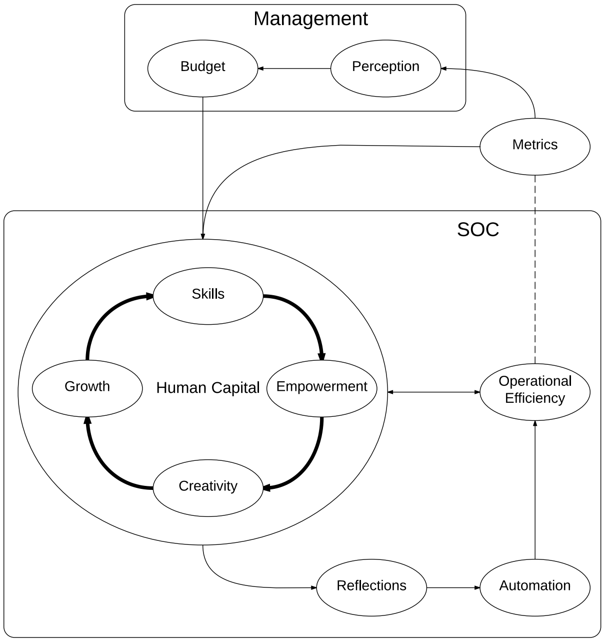


Figure 6: Metrics

> "I need stories to talk to my managers. You analysts are like snipers making a number of hits but I am not getting enough stories to tell my managers. You guys need to generate useful reports so that I can convey the usefulness of the SOC."

Figure 6 also shows that there is a direct causality between metrics and human capital. In the worst case, the metrics will decide the tasks an analyst can perform in the SOC–a form of restricted empowerment. An L1 analyst expressed frustration that supports our claim:

> "We feel that we are not doing security monitoring in the SOC. I think we are just working to generate numbers for higher management. We have raised some ethical concerns with the management regarding this."

The figure also shows the interlink between metrics and the rest of the categories. The dotted line between operational efficiency and metrics is to indicate the fact that metrics act as a communicating channel between SOC operations and the management. There is a possibility for the formation of a vicious cycle even in this context. The demand for metrics from the management might negatively affect the morale of the analysts, that in turn negatively affects operational efficiency. The management's perception of the usefulness of the SOC is driven by the metrics and the lack of good metrics communicating the value of the SOC will lead to reduced funds allocated for the SOC. A reduced budget is usually translated into less training opportunities which will drive the vicious cycle of human capital. More research has to be done on defining meaningful metrics to measure SOC efficiency benefiting the analysts and the management. The analysts benefit from good metrics as promotions and other perks are decided by the numbers conveyed

Table 2: SOC operations metrics

Metric	Measured Quantity	Purpose
Event volume graph	Number of events received per day	Used to anticipate how many analysts will be need to analyze new event sources.
Bulk processed count	Number of events that were bulk processed	Used to identify either improper analysis, or problem with a sensor generating too many events.
Duplicate comments graph	Number of events that were bulk processed	Used to identify either improper analysis, or problem with a sensor generating too many events.
Missed events	Number of events left unprocessed per day	Identify events falling through the cracks, taking more than a day to analyze, and if procedures are not followed.
Average processing time	Average time taken to analyze an event	Estimate how much time it takes for analysts to investigate an event.

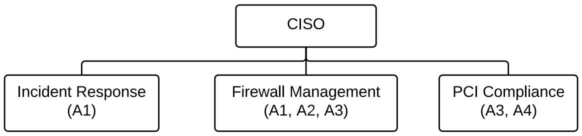


Figure 7: Organizational Chart for the University SOC

by the metrics. The management, on the other hand, will be able to measure better their return on investment on the SOC.

10. MODEL VALIDATION

We have been conducting anthropological fieldwork in a higher education University SOC for over 2 years now. The SOC is relatively smaller compared with the corporate SOC. There are 4 analysts headed by the Chief Information Security Officer (CISO). The SOC is the only security monitoring center for the three campuses of the University. There are 55,000 devices connecting to the network during business hours. The analysts performed operational tasks such as *incident management*, *firewall management*, and *payment card industry (PCI) compliance* as shown in Figure 7. The analysts in this SOC have diverse responsibilities compared to the corporate SOC. This is due to the smaller team size which also means analysts often have to multitask.

10.1 Fieldwork Setup

Five students in Computer Science, graduate and undergraduate, have worked as security analysts over a period of 2 years in the University SOC. They were trained by our anthropologist in participant observation methods and in note taking during fieldwork. The students were embedded as analysts and performed various tasks in the SOC such as incident handling, anti-virus management, host-based antivirus maintenance, and firewall management. In addition to the operational tasks they also built useful tools to increase operational efficiency. As in the corporate SOC we recorded the observations made in a digital document.

10.2 Burnout Symptoms

We did observe signs of analysts burnout, but in the University SOC they exhibited different symptoms. The burnout manifested itself mostly in the form of *frustration*. However, there was not high analyst turnover contrary to the previous SOC. We postulate that this was due to the location of the University. The University is located in a small town and the analysts do not have too many options when switching jobs as was the case in the corporate SOC. In the following sections we present the results of validating our human capital model for analyst burnout on fieldnotes from the University SOC. Some of the factors actually helped improve the morale of the analysts, thus enriching the human capital. During the validation process we found examples for factors that enabled and also helped mitigate burnout.

10.3 Automation

Analysts of this SOC were stuck performing the repetitive tasks everyday. One analyst expressed his frustration:

> "I want to do some interesting analysis on the data we are collecting but I am stuck with processing the same tickets everyday."

Fortunately, the analysts in this SOC were empowered to engage in periodic reflection of operational procedures. The fieldworkers engaged in periodic reflections with the analyst as part of the co-creation process. The result was a tool that automated most of the ticket generation process enabling the analyst to focus on interesting investigations.

In another instance, the fieldworkers developed a tool that automated malware analysis. The University's email provider placed certain restrictions on the type of files that can be attached in an email. This was done to prevent the spread of malware through email. Any message that contained one of those restricted attachment types were forwarded to a special inbox monitored by the SOC. One of the analysts used to manually download the files in the inbox, conduct analysis to eliminate duplicates and false positives, and submit the list of unique file hashes to the anti-virus (AV) vendor. The AV vendor would then determine based on the submitted hashes if there were any new malware unknown to them. Signatures to detect new malware, if any, would then be pushed out to the University's AV subscribers.

All the steps in the above process were performed manually by one analyst. One day the analyst called one of the fieldworkers and asked for help in automating this process. He wrote down the steps he was undertaking in an email after a day of reflection. A tool was then written to automate most of the steps in the process. Later on, the tool was handed over to another team outside the SOC, presumably less skilled than the SOC analysts. The analyst was *very happy* that the task got offloaded to the other team as he now was able to focus on other sophisticated operational tasks.

Many such mundane operational tasks got automated this way, enabling the them to work on more creative tasks. These observations from our fieldnotes validated the causality between *operational efficiency* and *human capital* in Figure 6. Here we see a positive cycle between empowerment, automation, and operational efficiency leading to good morale of the analysts as suggested by our model.

10.4 Empowerment

The major factor that was affecting the human capital in this SOC was the cooperation issue with other information technology (IT) departments within the University. Often times the SOC has to work with other IT teams such as *networking* or *server management* to resolve a security incident. Since security might not be a high priority for other teams the remediation process gets delayed.

This often causes frustration for SOC analysts. For example, one analyst expressed dissatisfaction in ticket management process by stating this:

> "I cannot close my old tickets as we are waiting on these other departments."

In another instance, the university network was experiencing an unusual amount of traffic that was severely slowing down or disrupting vital services. Moreover, an uncommon behavior and load was noticed on several important networking devices. In order to troubleshoot the problem, the analysts needed to interact with the teams that manage the services and also the network equipment vendors. General reactions from the other entities:

> "Don't take my VLAN down. The problem is not here."

It was very challenging for the analysts to identify the source of the problem without temporarily disabling any services. Eventually, one of the analysts was able to pinpoint the misbehaving device. All these events happened during regular business hours and while the higher management was insisting on solving the problem as soon as possible.

During another instance an analyst had to wait on the department that managed servers for resources to deploy the developed tools. There were numerous issues with the *server management* department such as not enough staff and inadequate hardware resources that delayed the tool deployment.

On another occasion, one of the analysts came up with a *creative* way to distribute logs across machines to improve efficiency of log collection. Unfortunately, they were kept waiting for 11 months by the server management department. In this case the lack of inter-departmental support (cooperation) affected the use of the analyst's *creativity.*

In a nutshell, one can classify facilitating support of other departments as a form of empowerment called *cooperation.* The SOC management is in the most capable position to facilitate this. In the examples cited above the analysts required support from other departments in order to perform their jobs effectively. The analysts got *frustrated* due to the lack of support from other departments. This frustration has often led to analysts losing their temper, thus one can see that the causality between empowerment and creativity indicated by the model in Figure 2 is validated.

10.5 Metrics

Similar to the corporate SOC there were metrics in use to measure analysts performance and the usefulness of various tools used in the SOC. We observed that metrics influenced the analysts' view on the perception of their performance. In other words, the more reflective the metrics of the analysts' achievements were, the more confident they were in the management's evaluation.

Defining good operational metrics was again a challenging task for the management, similar to the corporate SOC. For example, the SOC manager wanted the analysts to spend a fixed amount of time on *operational tasks* and the rest on *projects*—tasks that lead to improvement of the SOC infrastructure. The manager also asked each of the analysts to enter the time they spend on each of the two tasks every week. The senior management wanted to know the amount of time the SOC analysts were spending on operations, hence the CISO devised this metric. The rationale behind this, from the management's perspective, was that the purpose of the SOC was to provide services to the users of the University's network. The management wanted the analysts to spend more time on operations for that reason.

The problem though was that there was a difference of opinion between the management and analysts on what constituted an operational task. Analysts were concerned that the metric did not account for the time spent on meetings and other tasks that were neither operational nor project work. The view among the analysts was that one cannot bin the tasks performed by analysts into discrete categories.

There were also attempts to measure the time each analyst spent on creating tickets. This feature was implemented in the ticketing system to track the time taken by an analyst to resolve an incident. The SOC management wanted to use this feature as a way to measure their return on investment on the analysts. The analysts raised *concerns* that this may not be reflective of the actual effort spent in resolving incidents:

> "I sometimes spend a few good hours on an alert and find out that it is a false positive. Does that mean I am not productive? It is just the way incident handling works and this metric does not capture that."

These observations validate the direct causality between metrics and human capital in Figure 6. The metrics did not reflect the effort of the analysts which in turn led to dissatisfaction, driving down their morale.

Another observation we made indicated the effect of metrics on human capital through management perception. As we described earlier this SOC has only 4 analysts but often have too many operational tasks to handle. The CISO mentioned a while ago, a few good times, about hiring a new analyst but that never happened. Meanwhile, we also observed that the analysts were not content with the financial compensation they were receiving. The perception was that they were given more tasks with no perceived increase in their compensation. The continued existence of this concern among the analysts highlighted the fact that the metrics were not reflective of the operational situation. Right metrics could have indicated to the management that the SOC was understaffed and could benefit from recruitment of another analyst. We thus see a validation for the causality between the metrics affecting the morale of the human capital via the management perception and budgeting.

11. LIMITATIONS

Our work has a few limitations. Firstly, the proposed burnout model was validated only on one more SOC thus far. Validating our model requires access to analysts at a number of different SOCs. We were fortunate so far to have been able to conduct our fieldwork at two different locations, a corporate SOC and a University SOC. Our next step in this direction is to try and obtain access to a few more SOCs (industry and government sectors) to validate our findings.

The second limitation of our work is that we cannot conclude that our model is exhaustive. The model is based on our fieldwork at a SOC for 6 months. Although we started to see repetitions in the observations we were making after a few months, one cannot say for sure if new events will or will not occur after we concluded our fieldwork. Despite this limitation the model was able to explain the burnout symptoms in the University SOC without any contradicting observations.

Lastly, the observations were documented and analyzed by the individual fieldworkers at the University SOC, as is the case with any anthropological fieldwork. We tried our best to be objective when documenting the observations. In spite of this we acknowledge the fact that there is a chance that the documented observations might have been affected by the subjectivity of the fieldworkers.

12. RELATED WORK

There have been a number of prior research efforts focused on tool development for analysts [2, 8, 16]. Werlinger et al. [18, 17] studied the effect of human, organizational, and technological factors on analysts through interviews of practitioners besides identifying activities that require communication in IT security management (ITSM). Botta et al. [1] examined the use of cues and norms in ITSM and discussed challenges that undermine their usage. Furthermore, Werlinger et al. [19] studied security analysts engagement in diagnostic work during incident response. In another work, a team of psychologists from George Mason University have been studying computer security incident response teams (CSIRTs) using organizational psychology [3].

Shropshire et al. [11] talks about various factors leading to information technology (IT) employees leaving the field of IT. First, they conduct a survey of studies in the nursing and accounting disciplines. Based on this initial study and survey of other articles on career exodus they identify stress, job insecurity, and burnout as most likely causes for IT analysts leaving the field. To validate their hypothesis, they conduct a survey of IT professionals in a public service organization in the southeastern area of the United States. Their results indicate a strong correlation between intention to leave the IT field and the mentioned three factors. Shuey et al. [12] conducted qualitative semi-structured interviews of 343 IT professionals in 40 small and medium sized companies in four countries to understand worker well-being in the modern economy. Their data was supplemented with quantitative data obtained from 403 employees in those 40 firms. They identify organizational structure, peer pressure, and individual constraints as reasons leading to burnout of IT employees. They found employees to be stressed out either due to working long hours as they thought that was their organization's culture, fear of losing their job when jobs were scarce, or due to transitions in personal life such as starting a new family. While these two papers are closer to our work there are significant distinctions. We study SOCs, a specific type of IT organization with very specific goals compared with IT in general. Moreover, we focus specifically on the burnout problem which is a main precursor to a security analyst leaving his/her job.

13. CONCLUSIONS AND FUTURE WORK

Human security analysts are the most critical components of a SOC followed by tools and procedures. Unfortunately, SOCs have been suffering from the high turnover rates of their analysts resulting form burnout. Frequent turnover leads to increased spending on hiring and training by the management. In this work we try to understand the concrete factors leading to burnout of security analysts. To understand the problem we performed an anthropological study in a corporate SOC for a period of six months. We worked with an anthropologist to train the students in participant observation methods and also in analyzing our fieldwork notes. The fieldnotes were analyzed using a Grounded Theory based approach and the result was a model describing the burnout problem.

To the best of our knowledge this is the first study of the burnout phenomenon in a SOC environment using anthropological methods. We believe that the burnout in SOCs is a human capital management problem. Specifically, burnout is caused due to cyclic interaction of human, technical, and managerial factors. We also note that there exist a number of vicious cycles between those factors leading to burnout. To validate the model, we used the fieldwork notes from a higher education institution SOC. The model was able to successfully explain the reasons for burnout in this other SOC. Throughout the paper we also provide guidelines for SOC management to maintain a high morale among the analysts. To further evaluate the model we are planning to conduct focused interviews of analysts in a few other SOCs.

14. ACKNOWLEDGMENTS

We would like to thank Dr. Jeff Yan of Newcastle University for his useful comments and suggestions.

This research is supported by the National Science Foundation under Grant No. 1314925 and by the Department of Homeland Security (DHS) Science and Technology Directorate, Homeland Security Advanced Research Projects Agency, Cyber Security Division (DHS S&T/HSARPA/CSD), BAA 11-02 and the Air Force Research Laboratory, Information Directorate under agreement number FA8750-12-2-0258. Any opinions, findings and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright notation thereon.

15. REFERENCES

[1] D. Botta, K. Muldner, K. Hawkey, and K. Beznosov. Toward understanding distributed cognition in it security management: the role of cues and norms. *Cognition, Technology & Work*, 13(2):121–134, 2011.

[2] D. Botta, R. Werlinger, A. Gagné, K. Beznosov, L. Iverson, S. Fels, and B. Fisher. Towards understanding it security professionals and their tools. In *Proceedings of the 3rd symposium on Usable privacy and security*, pages 100–111. ACM, 2007.

[3] T. R. Chen, D. B. Shore, S. J. Zaccaro, R. S. Dalal, L. E. Tetrick, and A. K. Gorab. An organizational psychology perspective to examining computer security incident response teams. *IEEE Security & Privacy*, (5):61–67, 2014.

[4] Cuckoo. Cuckoo sandbox. http://www.cuckoosandbox.org/.

[5] V. Foundation. Volatility. http://www.volatilityfoundation.org/.

[6] C. Geertz. Deep play: Notes on the Balinese cockfight. *Daedalus*, 101(1):1–37, 1972.

[7] Hewlett-Packard. Building a successful security operations center. http://h71028.www7.hp.com/enterprise/downloads/software/ESP-BWP014-052809-09.pdf, 2011.

[8] P. Jaferian, D. Botta, F. Raja, K. Hawkey, and K. Beznosov. Guidelines for designing it security management tools. In *Proceedings of the 2nd ACM Symposium on Computer Human interaction For Management of information Technology*, page 7. ACM, 2008.

[9] C. Maslach, W. B. Schaufeli, and M. P. Leiter. *Job Burnout.* Annual Review of Psychology, Vol. 52: 397-422, 2001.

[10] B. Rothke. Building a security operations center. http://www.rsaconference.com/writable/presentations/file_upload/tech-203.pdf, 2012.

[11] J. Shropshire and C. Kadlec. I'm leaving the IT field: The impact of stress, job insecurity, and burnout on it professionals. *International Journal of Information and Communication Technology Research*, 2(1), 2012.

[12] K. M. Shuey, H. Spiegel, J. McMullin, and V. Marshall. The structure of it work and its effect on worker health: job stress and burnout across the life course. *Aging and working in the new economy: changing career structures in small IT firms. Northampton, MA: Edward Elgar Publishing*, pages 163–194, 2010.

[13] A. Smith and J. S. Nicholson. *An Inquiry Into the Nature and Causes of the Wealth of Nations.* T. Nelson and Sons, 1887.

[14] A. Strauss and J. M. Corbin. *Basics of qualitative research: Grounded theory procedures and techniques.* Sage Publications, Inc, 1990.

[15] S. C. Sundaramurthy, J. McHugh, X. S. Ou, S. R. Rajagopalan, and M. Wesch. An anthropological approach to studying csirts. *IEEE Security & Privacy*, (5):52–60, 2014.

[16] N. F. Velasquez and S. P. Weisband. Work practices of system administrators: implications for tool design. In *Proceedings of the 2nd ACM Symposium on Computer Human Interaction for Management of Information Technology*, page 1. ACM, 2008.

[17] R. Werlinger, K. Hawkey, and K. Beznosov. Security practitioners in context: their activities and interactions. In *CHI'08 Extended Abstracts on Human Factors in Computing Systems*, pages 3789–3794. ACM, 2008.

[18] R. Werlinger, K. Hawkey, and K. Beznosov. An integrated view of human, organizational, and technological challenges of it security management. *Information Management & Computer Security*, 17(1):4–19, 2009.

[19] R. Werlinger, K. Muldner, K. Hawkey, and K. Beznosov. Preparation, detection, and analysis: the diagnostic work of it security incident response. *Information Management & Computer Security*, 18(1):26–42, 2010.