Supporting Password Coping Strategies with Persuasive Design
PhD Thesis

Tobias Seitz
tobias.seitz@ifi.lmu.de

October 5, 2017

Contents

Chapter 1

Introduction

1.1 Motivation

1.1.1 Authentication Costs Time

1.1.2 Weak Passwords Cost Money

1.2 Research Objectives

1.2.1 Problem Statement

1.2.2 Better Understanding of Password Usability

1.2.3 Making Users' Lives a Little Easier

1.3 Agenda: Claims to Support in this Thesis

Perception of Password Strength Over the past decade, users received many hints and advice to construct strong passwords. Their understanding of a secure password has changed and is sometimes wrong. We show that this is the case in section @TODO REF SEC

Password Composition Policies As many web sites require or allow some kind of registration, their operators implement different password composition policies. We show that the criteria are manifold and largely inconsistent. Consequently, users approach enrollment with their preferred password, and are forced to apply heuristics to modify the password, depending on the policy in use.

Password Value We present a framework to assess the value a user associates with a specific password. The users might not realize that they re-use the password for accounts with different values. Knowledge about a password's value is important to design persuasive strategies to protect it, e.g. by discouraging its usage on low value accounts. (See PSST).

1.4 Main Contributions

1.4.1 Insights into the Psychology of Passwords

1.4.2 Designing Around Password Reuse

1.4.3 When and Why to Apply Nudges

1.4.4 Holistic Password Support

1.5 Thesis Overview

Chapter 1:
Chapter 2:
Chapter 3:
Chapter 4:
Chapter 5:
Chapter 6:
Chapter 7:

Part I

AUTHENTICATION ON THE WEB

Chapter 2

Foundations of Studying Passwords

2.1 A Brief History of Passwords

when was this first used why?

benefits, obvious drawbacks, attack scenarios online offline

2.2 Metrics and Statistics

The initial approach towards estimating the strength of passwords was to look at their *Shannon entropy*(@@Quelle).

Lately the community reached widespread consensus that the realistic strength of password can be defined as *the number of attempts that an attacker would need in order to guess it* [**?**]

2.2.1 Entropy vs. Guesswork

@@TODO tell the story of all the attacks – Melicher, Johnson, Ur etc. Password Guessability Service etc.

2.2.2 The zxcvbn Approach

Daniel Wheeler presented an approach towards password strength estimation by looking at a conservative expected guess attempt number [**?**]. The idea is to utilize pattern matching against dictionaries and leaked password corpora and then calculate the minimum rank over a series of frequency ranked lists. In other words, the approach is heuristic instead of probabilistic, because the ranking is based on searching through the patterns and ranking them not only based on their likelihoods, but other factors like keyboard sequences. The implementation of the algorithm is called zxcvbn[1]. Wheeler showed that in an online attack scenario [4] the algorithm estimates the number of guesses accurately within an order of magnitude of 2 – consistently better than NIST guidelines to date and KeePass strength estimators. That is, utilizing 100,000 tokens stored within a 1.5 Megabyte file zxcvbn conservatively estimates the number of guesses required to crack a password. Beyond the online-attack threshold, the results are mixed, but we can observe that adding more tokens to the dictionary improves accuracy even more. The great benefit of using this method is its speed and size that make it a lightweight tool that is prepared for widespread adoption, to relieve users from "LUDS" policies

[1] The name zxcvbn originates from the bottom row on a QWERTY keyboard. Many users mistakenly consider this approach secure because the resulting password looks fairly random.

(lowercase, uppercase, digits, symbols). We can conclude that zxcvbn is a reasonable tool when we collect meta statistics about passwords, e.g. in studies where it is ethically questionable to collect plain text passwords [**?**]. Wheeler also points out that at this point there is no study comparing the effects the different estimators have on user behavior.

2.3 Authentication without Passwords

Bonneau et al. argue [1] that passwords are an imperfect technology that are difficult to replace. One, industry has found ways to work around the many drawbacks and can compensate breaches to a large part. Second, alternatives to passwords are often privacy invasive. For example, identity providers like Google or Facebook often collect large amounts of personal data on the users. Third, Bonneau et al. point out that empirical evidence from practice often contradicts results produced by academic research. They advocate that researchers rethink their model of users, who often behave too predictably and whose behaviors one should not try to change. Instead, academia could tackle new approaches that would be dangerous to the success of businesses and thus are seldom tried out.

The authors point out how user models assumed by researchers often do not apply in reality, for instance, their behavior is anything but random: Users pick from a limited set of passwords that is far smaller than random passwords. Just as [5], the paper strongly discourages focusing on offline attack scenarios and suggests that users should at most try and protect themselves against online attack scenarios. Consequently, the lessons from the past about attempts to improve password strength through changing user behavior are questionable in the authors' eyes. Even more so, because other attacks (phishing, malware, eavesdropping, stealing from servers or identity tokens) are not fended off at all by stronger passwords. Online attacks can drastically be mitigate by rate-limiting and contextual information (e.g. geolocation). Yet, the authors see that not all sites employ these methods, "probably to avoid denial of service". Users also often receive too much advice from security experts that is often contradictory and in extreme cases boiled down to "Pick something you cannot remember and do not write it down".

Bonneau et al. answer the question whether we still need passwords with a differentiated "yes": *Passwords appear to be a Pareto equillibrium*[2]. Also the learning curves for password authentication at a new service is virtually non-existent. However, as many researchers in the community, the authors see the feasibility of multi-factor authentication in progressive or continual ways as the most promising future. The challenges for privacy and usability we find in these approaches are mostly ill-defined or not validated. The user experience of such systems may improve while the drawbacks come at a high price that the users may not understand at all.

2.3.1 Biometrics

2.3.2 Multimodal Authentication

One Time Passwords

OTP also have the possibility to replace a password that you have to memorize, but there are disadvantages as well.

2.3.3 Shared Authentication and Hardware Tokens

Single Sign-On (SSO). Identity Providers, Reference Models:

[2] TODO add definition of this here. It's about game theory.

- **OAuth** Twitter, Google
- **OpenID** Google, PayPal
- **Facebook Connect**

Problems: Successful identity providers such as Facebook take a central role as the “sole identity provider, which does little for privacy” [1].

Privacy The biggest players for shared authentication are Facebook, Google, and Twitter. However, it is exactly these three companies for which users raise the most privacy concerns. On the one hand, users trust the companies because they know how much data they store and manage, and only few breaches are known. On the other hand, users are doubtful that their credentials will be in safe hands and not shared with third parties (users lack the understanding of how shared authentication works internally and cannot separate privacy and security).

Single Point of Failure Embracing the opportunities of shared authentication, users integrate it into their password management / coping strategies **??**.

2.4 Passwords on Mobile Devices

touch some small aspects, especially Melicher’s Paper

set the stage for the emoji passwords later.

2.5 Passwords are Here to Stay

Chapter 3

Decision Making in Usable Security

Related Work Overview *

- Forget Papers: [7] [8] [6] [?]
- Chiasson: [2]

3.1 Cognitive Illusion and Biases

3.1.1 Definition

3.1.2 Scenarios

3.2 Persuasion and Nudging

3.2.1 Persuasive Authentication Framework

3.2.2 Privacy Nudges

3.3 Personality Traits

3.3.1 Inventories

3.3.2 Personality Heuristics

3.3.3 Privacy Decisions

3.4 The Persuasive Authentication Framework

3.5 Nudging

3.5.1 Personalization and Individualization

Already the PAF introduced the personality dimension for effective persuasion. Recently, Egelman and Peer advocated to use psychometric cues to contextualize privacy or security messaging [?]. In two studies they looked at the predicting power of different psychometric scales, among which we find the five-factor model ("Big Five"), the General Decision Making Style (GDMS) or the Domain-Specific Risk-Taking scale (DoSpeRT) @@TODO dreimal Quelle. They conducted two online experiments using Mechanical Turk. In the first round they used the Ten-Item-Personality Index and correlated the scores with several privacy metrics (e.g. the Privacy Concerns Scale or the Internet Users Information Privacy Concerns scale). They realized that the Big Five model was a rather weak predictor for the different scales, which lead them up to their second experiment, where they focused on decision-making metrics. Here, they observed stronger correlations, between e.g. the rational trait of the GDMS and both PCS and IUIPC. Following these observations, Egelman and Peer interpret them as evidence that privacy attitudes originate from rational decision-making, and also from people's intuitions, which sounds paradoxical at first, but it were different respondents who produced this result. They conclude that the decision-making metrics were about three times as powerful regarding their predictive power than the big five model regarding privacy attitudes and behavior. Finally, they propose the Security Behavior Intentions Scale (SeBIS). This scale intends to isolate confounding factors when correlating personalty traits with security behavior. Also, the authors offer a number of hypotheses and unanswered research questions that highlight how little exploiting personality traits in security nudges has been investigated.

Chapter 4

The Human Side: Selection and Coping

Florêncio and Herley probably conducted the largest study to date on password habits. Their intention was to find out among other things A) how often people type passwords, B) how many sites share a password C) how many distinct passwords a user has, and D) how the strong the passwords are. They utilized the Windows Live Toolbar for Internet Explorer to collect in-the-wild data from up to 500000 users during three months of running the collection

[3]. The established protected password lists (PPL) to avoid intruding into people's privacy. They found that users had about 7 distinct passwords in 2007, and that passwords are re-used at about 6 sites in average. Interestingly, they found that stronger passwords are not re-used as often as weak passwords (only around 4 sites). It was not possible to trace the incoming data back to a specific user, which might have resulted in over counting of entries. Also, it was not measured how long the actual password entries takes. If users only used regular dictionary words without any modification, the key logging module of the toolbar would record a password reuse event (PRE) every time the user entered the word – also in regular text searches, for example. Another limitation could be that they used entropy as a proxy for password strength. However, as discussed in the previous chapter, we have seen that this metric is more robust for system-generated passwords and that strength estimation has evolved over the past ten years.

@@TODO cite Wash paper @SOUPS 2016.

4.1 Methodology: Running Password Studies

4.1.1 Ecological Validity

4.1.2 Mechanical Turk Studies

4.2 How Users Select Passwords

4.2.1 Problem: Weak Passwords

4.3 How Users Traditionally Cope with Passwords

4.3.1 Tools

Problem: Accessibility

Word documents post-its (use a screenshot of french newspaper that was featured on tv and you could see one of their passwords in the back. spouses can access them .

4.3.2 Problem: Password Re-Use

Policy Fulfillment

Re-Use Approaches

4.4 Helping Users

4.4.1 Advice and Guidelines

4.4.2 Memorization Techniques

4.4.3 Real-Time Feedback

4.4.4 Password Managers

Part II

EVALUATING THE PROBLEM

Chapter 5

Mental Models of Password Strength

5.1 Background and Context

5.1.1 Research Questions

5.2 Related Work

5.3 Approach: PASDJO - The Password Game

5.4 Log Analysis

5.4.1 Sample

5.4.2 Results

5.5 Discussion

5.6 Limitations

Chapter 6

Field Study: Authentication Time

6.1 Background and Context

6.2 Results

TODO

6.3 Password Habits

A lot of Herley and Florencios work focused on quantifying the “password problem” as an everyday task (@REF SEC RW). However, the studies were conducted almost a decade ago. @TODO elaborate on argument: More internet services, rejection of SSO etc.

We have therefore reason to believe that the amount of accounts has increased compared to the 2000s. As such, we hypothesized that users face a higher number of authentication tasks, in particular with passwords on the web, each day.

6.4 Password Managers

6.4.1 Criticsm

Leakage LastPass breach in 2015 [1]

[1] https://blog.lastpass.com/2015/06/lastpass-security-notice.html/, last access March 30 2016

Chapter 7

Chapter 8

Mental Models of Password Strength

8.1 Background and Context

8.1.1 Research Questions

8.2 Related Work

8.3 Approach: PASDJO - The Password Game

8.4 Log Analysis

8.4.1 Sample

8.4.2 Results

8.5 Discussion

8.6 Limitations

Chapter 9

Password Authentication in the Years 2016 and 2017

Chapter 10

Survey: Policies against Reuse

10.1 Background and Context

10.2 Method

10.3 Results

Policies are Mostly homogenous, with slight differences

Policies do not prevent password-reuse

10.4 Limitations

Chapter 11

Field Study: Password Selection on Real-World Accounts

11.1 Background and Context

Roskilde sample data.

11.2 Privacy-respectful Password Logs

Show how my anonymous zxcvbn contributed to the anonymous collection of real-world data.

11.3 Results

Chapter 12

3 Studies on Detecting Password Reuse in Real Time

12.1 Background and Context

12.1.1 Research Questions

12.2 Related Work

12.3 Approaches

12.3.1 Heuristics

12.3.2 Probabilistic Methods

12.3.3 Machine Learning

12.4 Feasibility Studies

12.4.1 Samples

12.4.2 Results

12.5 Data Consolidation and Proof-of-Concepts

12.5.1 Samples

12.5.2 Results

12.6 Discussion

12.7 Limitations

Part III

INFLUENCE STRATEGIES

Chapter 13

Redesigning Verbal and Non-Verbal Password Feedback

13.1 Background and Context

13.1.1 Research Questions

13.2 Related Work

13.3 Case Study: Work Passwords

13.3.1 Samples

13.3.2 Results

Feedback needs to be Personalized

Visual feedback first, verbal feedback next

Verbal feedback requires more verbal feedback

Nudging by Examples most promising

13.3.3 Limitations

13.4 Case Study: Participatory Design of a Password Meter

13.5 Case Study: Jumping on the Bandwagon

Mention that LastPass already has something like this now. – find out when they introduced it (write them an email)

13.6 Discussion

13.7 Take-Aways

Chapter 14

Extending the Password Space with Emoji

14.1 Background and Context

14.1.1 Research Questions

14.2 Related Work

14.3 Case Study: Work Passwords

14.3.1 Samples

14.3.2 Results

Feedback needs to be Personalized

Visual feedback first, verbal feedback next

Verbal feedback requires more verbal feedback

Nudging by Examples most promising

14.3.3 Limitations

14.4 Case Study: Participatory Design of a Password Meter

14.5 Case Study: Jumping on the Bandwagon

Mention that LastPass already has something like this now. – find out when they introduced it (write them an email)

14.6 Discussion

14.7 Take-Aways

Chapter 15

Understanding Password Selection Through the Lens of Personality Traits

15.1 Background and Context

15.1.1 Research Questions

15.2 Related Work

15.3 Case Study: Work Passwords

15.3.1 Samples

15.3.2 Results

Feedback needs to be Personalized

Visual feedback first, verbal feedback next

Verbal feedback requires more verbal feedback

Nudging by Examples most promising

15.3.3 Limitations

15.4 Case Study: Participatory Design of a Password Meter

15.5 Case Study: Jumping on the Bandwagon

Mention that LastPass already has something like this now. – find out when they introduced it (write them an email)

15.6 Discussion

15.7 Take-Aways

Chapter 16

Password Selection and the Decoy Effect

16.1 Background and Context

16.1.1 Research Questions

16.2 Related Work

16.3 Designing Decoy Options

Talk about how we did the iterations.

16.4 Evaluating the Design

16.4.1 Methods

16.4.2 Survey Results

Mostly focus on the contents of the paper here

16.4.3 Live-Deployment Data

Talk about a couple of findings of the Roskilde dataset.

16.4.4 Limitations

16.5 Discussion

16.6 Take-Aways

Chapter 17

Side-Effects and Corroborative Results

17.1 What users do to create strong passwords

- Hier alles zusammenfassen was in den Studien herausgefunden wurde, was Leute tun, um starke Passwörter zu generieren. - In alle BAs/ MAs reinschauen weil da jeder etwas dazu geschrieben hat.

Part IV

THE PASSWORD SUPPORT TOOLKIT

Chapter 18

Overview

The Password Selection Support Toolkit (PSST) is a holistic approach to help users of any experience level manage their password portfolio of any size. Ideally, continuous usage of the PSST enables and encourages users to

a) improve their strategies

b) strengthen the passwords

c) boost memorability of passwords

All this leads to an improvement of user experience as both pragmatic and hedonic qualities are maximized.

18.1 Privacy Concerns

To make sure users can fully benefit from the power of the PSST, it is required that no information about the user's behavior travels through the web, i.e. to remote servers. Such knowledge would deeply compromise protection against foreign access. Attackers would intensify attacks on the stored behavioral data that would allow them to reverse engineer credentials. This has to be avoided at all costs. Consequently, the calculations need to take place on the client-side. With increasing computational power even for consumer electronic devices this does not seem to be a tough restriction.

18.2 Adapting to Existing Strategies

As discussed in Chapter **??**, each individual user handles password selection and memorization differently, even though patterns occur.

The most prevalent factors for password selection are:

- Account Value

- Memorability

- **Perceived Strength**. As we have seen, people's perception of strength has changed, so as to speak "they were enlightened". Everyday practice has taught users that difficult passwords are good, but everybody thinks that they seem hard to do. The PSST allows to shift this perception.

The PSST lets the user know, that their strategy is legitimate. It acknowledges them and does not want to change them. Even password re-use can be beneficial.

The PSST addresses password strategies, by learning them through both active involvement and observation.

18.2.1 Learning Account Values

The PSST adjusts nudges and persuasion to the users' perceived value of a password.

Measuring Password Value

Users often fail to assess the value of passwords correctly as we have seen in (@TODO REF SEC).

18.3 Composition Policy Fulfillment

18.4 Supporting Memorization

After the users picked their password, be it completely independently or with aid by the PSST, they are now facing the critical stage of memorizing it. In case they simply re-used or modified an old password, this phase is very short. They might write down which of their ready-made credentials they utilized. This is necessary and helpful if the account value is not clear or might change in the future.

However, in case the PSST manage to persuade a user towards are slightly different option than they would have chosen without it, memorization is extremely important. As such, the PSST provides means to support this process in @TODO ways:

- training.
- **Storing**. The password is stored and can be retrieved from a local database.

18.5 Hedonic Qualities

Hedonic UX qualities are often overlooked. The design and incorporates them from the very beginning.

Chapter 19

Proof-of-Concept: Personalized Password Manager

19.1 Background and Context

19.2 Design Iterations and Evaluations

19.2.1 Persuasion in Password Managers

Talk about ATH Results here

19.3 Field Study

19.3.1 Sample

19.3.2 Results

19.4 Discussion

Part V

CONCLUSION

Chapter 20

Summary

20.1 Classification and Dimensions

20.2 Lessons Learned

What would I do differently the next time I take on a PhD project?

Chapter 21

Crowd Sourced Experiments

21.1 mTurk Studies

21.1.1 Privacy

- Egelman: [?]

Bibliography

[1] Joseph Bonneau, Cormac Herley, Paul C. Van Oorschot, and Frank Stajano. 2015. Passwords and the Evolution of Imperfect Authentication. *Commun. ACM* 58, 7 (2015), 78–87. DOI: http://dx.doi.org/10.1145/2699390

[2] Sonia Chiasson, Alain Forget, Robert Biddle, and P C van Oorschot. 2008. 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*. 121–130. http://dl.acm.org/citation.cfm?id=1531514.1531531

[3] Dinei Florêncio and Cormac Herley. 2007. A Large-Scale Study of Web Password Habits. In *Proceedings of the 16th international conference on World Wide Web (WWW '07)*. ACM, 657–665. DOI:http://dx.doi.org/10.1145/1242572.1242661

[4] Dinei Florêncio, Cormac Herley, and Paul C Van Oorschot. 2014a. An Administrator's Guide to Internet Password Research. In *Proceedings of the 28th Large Installation System Administration Conference (LISA14)*. USENIX Association, Seattle, WA, 35–52.

[5] Dinei Florêncio, Cormac Herley, and Paul C. Van Oorschot. 2014b. Password Portfolios and the Finite-Effort User: Sustainably Managing Large Numbers of Accounts. In *Proceedings of USENIX Security Symposium*. USENIX Association, San Diego, CA, USA, 575–590. https://www.usenix.org/system/files/conference/usenixsecurity14/sec14-paper-florencio.pdf

[6] Alain Forget and Robert Biddle. 2008. Memorability of Persuasive Passwords. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '08)*. 3759. DOI: http://dx.doi.org/10.1145/1358628.1358926

[7] Alain Forget, Sonia Chiasson, and Robert Biddle. 2007. Persuasion as Education for Computer Security. In *Proceedings of E-Learn: World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education*. Association for the Advancement of Computing in Education (AACE), Chesapeake, VA, 822–829.

[8] Alain Forget, Sonia Chiasson, P C Van Oorschot, and Robert Biddle. 2008. Improving Text Passwords Through Persuasion. In *Proceedings of the 4th Symposium on Usable Privacy and Security (SOUPS '08)*. ACM, New York, NY, USA, 1–12. DOI:http://dx.doi.org/10.1145/1408664.1408666