Contextual Identity: A Framework for Usable Privacy

No Author Given

No Institute Given

1 Introduction

I am large, I contain multitudes. — Walt Whitman, *Song of Myself*

People reveal different aspects of themselves depending on context. At any given time, they may be acting as a friend, relative, spouse, co-worker, acquaintance, or stranger. This observation is not new; Erving Goffman called this impression management, and Carl Jung called it persona theory [10,15]. In this short paper, we call it contextual identity.

Helen Nissenbaum has long argued that privacy violations come not from the simple sharing of personal information, but rather from sharing that information in a way that violates social norms, or in the wrong context [29]. The desire for spontaneous, positive human interaction often necessitates sharing personal information. Take, for example, PatientsLikeMe [14]. This site functions as a support group and also a place for people to share health information, such as drug regimens, treatment history, side-effects, and more. Though this information is traditionally considered extremely sensitive, the users of this site clearly consider the benefits of sharing (longitudinal graphs of their own health history, building community with people who share their disease) to outweigh the privacy risks. So long as the contextual integrity of the shared data is kept intact, users will continue to trust and use the site. Selling user data intended for one purpose for another (e.g., to insurance companies to discover uninsurable patients) would be a clear violation of that expectation.

In offline environments, managing contextual identities is more intuitive than in online environments. There are cues to help you determine where you are, who your intended audience is, how many people will overhear you, and how likely your information is to be re-broadcast in a different context. It is also true that in offline environments, humans don't have perfect memories and will eventually what forget. However, with the advent of increasingly vast searching, indexing, and archiving cabilities, one cannot rely on forgetting in online enviroments [21]. Thus, managing contextual identities in online environments has become increasingly difficult and fraught with mistakes.

One possible solution to this problem is never to post personal information online, or at least to consider the entire world might be the audience. However, this solution defies human nature. Given that there are 1 billion active Facebook users and that 30-40% of spoken communication is devoted to informing others about ourselves, humans are bound to fall short of these guidelines [33,38]. The security and privacy community should recognize this and develop techniques to meet users' needs rather than expecting human nature to change.

We attempt to formalize the notion of contextual identity, use it as a framework to understand existing privacy violations, and propose mitigations to those violations.

2 Definitions

Let

- R be a set of actors and their relationships to the user
- T be a set of tasks and their side-effects when performed in service of R
- P an impression that actors in R have of the user when performing tasks in T

Let a **contextual identity** I be a loose coupling of (R, T, P). A user may have multiple of these. These identities change over time, as users form and dissolve relationships, and focus on different sets of tasks. The desirable outcome is that the user can accomplish tasks in T while being able to predict R and P.

This definition focusses on relationships because most privacy concerns stem from causing problems with people the users knows [36]. This definition focusses on tasks because users are task-oriented, and it is easy to map user actions (page navigation, sending email, posting, filling forms) to tasks. (include?) The definitions of R and P are deliberately underspecified, as this is work in progress.

3 Previous work

Many authors have discovered attacks to link different contextual identities to the same person. Narayanan and Shmatikov present a re-identification technique to merge anonymized social graphs from different networks and prove that Netflix and IMDB and Flickr and Twitter can be linked [27,28]. Lindamood et al. and Mislove et al. show how to infer previously undisclosed information from released social networking data [18,24].

Many authors have noted how re-broadcasting information information out of context and making information discoverable can lead to user distress (even if the information was previously public, but not easily discoverable) [7,5,29].

Barth et al. provide a formal model for contextual integrity of user data [3]. In contrast to this work, the contextual integrity model focusses on data as principal rather than users.

Although network attacks are very powerful, at the current time they require a level of expertise and resources not typically held by those causing the most worry to users. A brief survey of social networking advice reveals that the major causes of concern come not from a dedicated graph expert, but people that the user knows: friends, family, and co-workers [9,32]. Surveys conducted by Wang et al. show that typical regrets from posting on Facebook stem almost exclusively from fear of negative interactions with people that the users knows [36]. The consequences of these negative interactions can lead to loss of employment or breaking personal relationships. By contrast, much privacy research has focussed on network privacy, where the bad actor is a behavioral tracking service, or a state agent.

4 Contextual identity violations

4.1 Redistributing information out of context

Re-distributing information out of its original context often leads to embarrassment [29]. In November 2007, Facebook Beacon allowed third-party sites to publish purchases,

travel bookings, movie rentals to the user's activity stream. Because of poor opt-out and lack of visibility into what was being published, user outcry was immediate [22,26]. In December 2007, Google Reader exposed RSS feeds of user-marked news stories to the user's Google Talk contacts[1]. Although this feed was always public, prior to this launch it was not discoverable, leading to a bad experience for many users [12]. In September 2012, Facebook imported old wall posts from 2008 into the new Timeline interface. Although wall posts were always visible from profile pages, the new Timeline interface brought old wall posts (which users used to treat as private messages, before the advent of "Like" and comment buttons) to the attention of an audience that the user never intended [19].

In all these examples, the audience R of the identity expanded without user intervention.

4.2 Unsatisfiable policies

Sometimes service providers have policies which preclude isolating multiple identites. For example, Facebook and Google have a "Real Names" policy, which requires users to register for accounts with their legal name [4,13]. These policies presume the users can be one person to everyone. These policies ignore that community-building happens in many different contexts, that users have many legitimate reasons for presenting different identities in different contexts, and that users don't necessarily want those identities to be linked. For example, disallowing avatar handles as a primary identifier makes building a gaming community difficult at best, and linking a gaming identity to a professional identity has already caused problems for one exposed has already caused problems for Maine Senate candidate and World of Warcraft gamer Colleen Lachowicz in 2012 [11].

It is impossible for users who want to isolate multiple contextual identities to participate in these networks without violating the terms of service.

4.3 Federated login

Facebook Connect is a login platform that allows third-party websites to authenticate users using their Facebook identity [25]. Because it is against the terms of service to have multiple Facebook accounts, using Facebook Connect has the unwanted side-effect of linking multiple contextual identities.

Other login platforms and protocols, including OAuth and BrowserID do not suffer from this policy or design error [1,6].

4.4 User interface errors

Even in the absence of changes made by service providers to re-contextualize personal information, it is all too easy for users to broadcast information to an unintended audience. This phenomenon is so common on Twitter that it has its own name, "DM fail",

[1] Google Talk contacts are everyone with whom a user has chatted, which might include coworkers, supervisors, and friends

or Direct Message fail: when the user posts a public message instead of a private, direct message. Representative Anthony Weiner was a victim of this mistake when he inadvertently sent compromising pictures of himself publicly [30]. Considering that this mistake requires mistyping a single character (`@` instead of `d`), it's no surprise that DM failures are exceedingly common.

Similar to DM failures, posting to the wrong account is also a common mistake. Because many jobs require posting on social networks on behalf of the company, many people now have multiple accounts for personal and business use. KitchenAid, Chrysler, and Google are just three examples whose employees have made this mistake in the past year [23,31,37].

4.5 Third party tracking

Even if a user does not participate in social networks, they are still subject to tracking by virtue of browsing the web. Third-party cookies, flash cookies, web bugs, device fingerprinting, geolocation, history sniffing and more are techniques that a dedicated third party could use to collect browsing history and link contextual identities.

5 Possible mitigations

5.1 Inferring contextual identities

Accurately detecting the user's current contextual identity, whether heuristically or through asking the user, is paramount to helping the user manage it.

Although a person's contextual identities may be subconsciously clear, expecting users to enumerate their identities and manually curate them is probably too burdensome. Firefox has a Profile Manager, but not many people know about it or how to use it, to the extent that it will be removed[2].

Tang et al. show that most people do not take the time to create labels and maintain friend lists [34]. However, the same authors show how to use heterogenous networks to infer relationship types (e.g., manager-subordinate) through the graph itself is unlabeled.

It may also be possible to infer tasks from the user's activity. For example, the `about:profile` Firefox extension uses the Open Directory Project and Alexa to categorize URLs that the user has visited [17].

5.2 Automatic mitigations for information leakage

Given the set of relationships R and tasks T, we can categorize a user's contextual identities and prevent information leakage between them. For example, preventing third-party cookies from being read and set across task boundaries means that targeted ads for dating would not follow the user when visiting a technology news site. Knowing the

[2] See `https://bugzilla.mozilla.org/show_bug.cgi?id=214675#c53`

user's current contextual identity also allows the browser to suppress URL bar suggestions that aren't relevant to that identity, protecting the user from both shoulder-surfing and distraction.

Privacy modes in Firefox, Chrome, Safari, and Opera are a good first step, but are no means a complete solution [2]. There is no standard behavior for private browsing mode, and thus interaction with extensions, treatment of cookies, history, and bookmarks upon entering and exiting this mode are different and serve different use cases. Depending on user desire, R when using private browsing mode may be other household members, or it may be a service to which the user wants to limit cookie leakage. Another deficit is that private browsing mode does not adequately serve long-term tasks (such as curating a set of adult entertainment sites) in any major browser.

Another good partial solution is to prompt the user when they're about to link multiple contextual identities through a federated login service. For example, Persona is the suite of tools that implements the BrowserID protocol, and it supports multiple email identifiers [1]. In this case, when the user decides to register for a service using Persona, the task T they are performing is authentication, the relationships R are ones they will engage in when using the service, and P is tied to the email identifier they use to register for the service. Because we can categorize the type of service by URL as `aboutprofile` does, Persona could prompt the user when they're about to link the same email identifier to radically different contextual identities, such as ones for dating and professional use.

5.3 Recovering from mistakes

Auditing A major cause of user embarrassment is exposing information to the wrong audience, as illustrated in section 4. A browser is in the perfect position to intermediate social network posts, aggregate them locally, and present them to the user when they desire to audit their digital footprint. For example, a user could see all of the comments and posts they made in the last week and redact content that is sensitive or too negative, topics that are frequently regretted [36]. Although this approach does not help when the service provider decides to change the audience on the user's behalf, auditing gives the user a good opportunity to review posts that might cause regrets. In the case of high-volume users, such a tool could even incorporate sentiment analysis to aid the user in finding particularly problematic content.

Profiles can be generated by user-supplied application-level data, but also implicitly through browsing via techniques in 4.5. Collusion is a tool for visualizing cookies [35], in particular the third-party cookies set by tracking sites which are typically ad networks. Once the user knows which sites are tracking them, they can choose to block cookies from those sites.

Remembering to forget Humans don't have perfect memories, and neither should social networks [21]. Many users already engage in manual auditing and deletion of old posts [9]. A better solution would be to build tools that let the user manage this more easily. Both Twitter and Facebook provide APIs for deletion[3]. One way to reduce the

[3] At the time of this writing, Google Plus offers a read-only API

likelihood of linked contextual identities would be to provide applications that expire posts after some time.

5.4 Preventing mistakes

Slowing information flow Sometimes a user might want to prevent mistakes from being made in the first place. If a user is concerned about a particular service provider, they can install an extension like Disconnect to disallow cookies from that provider unless they explicitly want to interact with that site [16].

Protocols like Do Not Track are also a good first step to solving this problem [20]. Although Do Not Track is unenforceable from the client side, it provides a clear signal to service providers and ad networks that the user does not want to be tracked. As adoption increases, ignoring this signal will lead to loss of reputation for service providers that explicitly act against their users' wishes.

Visual reminders of context To mitigate privacy violations caused by poor user interfaces, the onus must be on software developers to provide sufficient visual cues of context. For example, many but not all Google products support multiple accounts. However, the visual cue to indicate which account is active is a small, easy to ignore text name at the top right of the page.

Much work has been done for visual verification of domain authenticity in the context of phishing, but not much has been done for visual verification of which user credentials are active [8].

6 Open problems

The mitigations proposed in this work become more complex in the face of shared devices and overlapping identities. There are good reasons to want to intermix work and personal identities (increased exposure, serendipitous connections) and so many people use the same account for both personal and work reasons.

Mobile growth is driving most internet growth, and much of that growth comes developing countries. The literature on how device sharing works in those markets is completely lacking. Managing contextual identities on shared devices promises to be a difficult problem.

7 Discussion

People have multiple contextual identities. Sharing personal information in those contexts is a great way to foster postive human interaction. We hope that contextual identity will serve as a framework for both developers to understand their users' privacy needs, and to develop tools that allow users to act in their contextual identities in a spontaneous, relaxed, and interesting way.

References

1. Ben Adida. Deploying BrowserID at Mozilla. http://identity.mozilla.com/post/12950196039/deploying-browserid-at-mozilla, November 17 2011.
2. Gaurav Aggrawal, Elie Bursztein, Collin Jackson, and Dan Boneh. An analysis of private browsing modes in modern browsers. In *Proc. of 19th Usenix Security Symposium*, 2010.
3. Adam Barth, Anupam Datta, John C. Mitchell, and Helen Nissenbaum. Privacy and Contextual Integrity: Framework and Applications. In *IEEE Symposium on Security and Privacy*, 2006.
4. Facebook Help Center. Facebook's Name Policy. http://www.facebook.com/help/?page=258984010787183, October 10 2012.
5. Monica Chew, Dirk Balfanz, and Ben Laurie. (Under)mining Privacy in Social Networks. In *Web 2.0 Security and Privacy*, 2008.
6. Ed. D. Hardt. The OAuth 2.0 Authorization Framework. http://tools.ietf.org/html/draft-ietf-oauth-v2-31, July 31 2012.
7. danah boyd. Privacy and publicity in the context of big data. http://www.danah.org/papers/talks/2010/WWW2010.html, 2010. Raleigh, NC.
8. Rachna Dhamija and J. D. Tygar. The battle against phishing: Dynamic Security Skins. In *Symposium on Usable Privacy and Security*, 2005.
9. Alicia Eler. Top 5 Facebook Privacy Tips. http://www.readwriteweb.com/archives/top_5_facebook_privacy_tips.php, April 2012.
10. Erving Goffman. *The Presentation of Self in Everyday Life*. Anchor, 1959.
11. Elizabeth Hartfield. Maine Candidate's World of Warcraft Persona Stirs Debate. http://abcnews.go.com/blogs/politics/2012/10/maine-candidates-world-of-warcraft-persona-stirs-debate/, October 5 2012.
12. Miguel Helft. Google Thinks It Knows Your Friends. http://bits.blogs.nytimes.com/2007/12/26/google-thinks-it-knows-your-friends/, December 26 2007.
13. Google Help. Google+ Page and Profile Names. http://support.google.com/plus/bin/answer.py?hl=en&answer=1228271, October 10 2012.
14. Benjamin Heywood, James Heywood, and Jeff Cole. PatientsLikeMe. http://www.patientslikeme.com/, October 10 2012.
15. C. G. Jung. *Two Essays on Analytical Psychology*. London, 1953.
16. Brian Kennish. Meet Disconnect. http://byoogle.blogspot.com/2010/12/meet-disconnect.html, December 2010.
17. Edward Lee. about:profile – Analyzing Data in Firefox. https://blog.mozilla.org/labs/2012/10/about-profile-analyzing-data-in-firefox, October 2012.
18. Jack Lindamood, Raymond Heatherly, Murat Kantarcioglu, and Bhavani Thuraisingham. Inferring private information using social network data. In *WWW*, 2009.
19. Matthew Lynley. Facebook on Privacy Scare: Nothing to See Here. http://blogs.wsj.com/digits/2012/09/24/facebook-on-privacy-scare-nothing-to-see-here/, September 24 2012.
20. J. Mayer, A. Narayanan, and S. Stamm. Do Not Track: A Universal Third-Party Web Tracking Opt Out. http://tools.ietf.org/html/draft-mayer-do-not-track-00, March 7 2011.
21. Viktor Mayer-Schoenberger. *Delete: The Virtue of Forgetting in the Digital Age*. Princeton University Press, July 2011.

22. Caroline McCarthy. MoveOn.org takes on Facebook's 'Beacon' ads. http://news.cnet.com/8301-13577_3-9821170-36.html, November 20 2007.
23. Mark Memmott. KitchenAid apologizes for 'offensive tweet' about Obama's grandmother. http://www.npr.org/blogs/thetwo-way/2012/10/04/162293140/kitchenaid-apologizes-for-offensive-tweet-about-obamas-grandmother, October 4 2012.
24. Alan Mislove, Bimal Viswanath, Krishna P. Gummadi, and Peter Druschel. You Are Who You Know: Inferring User Profiles in Online Social Networks. In *WDSM*, 2010.
25. Dave Morin. Announcing Facebook Connect. http://developers.facebook.com/blog/post/2008/05/09/announcing-facebook-connect/, May 9 2008.
26. Ellen Nakashima. Feeling Betrayed, Facebook Users Force Site to Honor Their Privacy, November 30 2007.
27. Arvind Narayanan and Vitaly Shmatikov. Robust De-anonymization of Large Sparse Datasets. In *IEEE Symposium on Security and Privacy*, 2008.
28. Arvind Narayanan and Vitaly Shmatikov. De-anonymizing Social Networks. In *IEEE Symposium on Security and Privacy*, 2009.
29. Helen Nissenbaum. *Privacy in Context: Technology, Policy, and the Integrity of Social Life*. Stanford University Press, 2010.
30. Adam Ostrow. Rep. Weiner meant to send lewd photo as direct message. http://mashable.com/2011/06/06/anthony-weiner-direct-message/, June 6 2011.
31. Erik Qualman. Chrysler Fires Agency over F*** Tweet. http://www.socialnomics.net/2011/03/14/chrysler-fires-agency-over-f-tweet/, March 14 2011.
32. Adam Rosenberg. 5 Essential Facebook Privacy Tips. http://mashable.com/2010/05/18/facebook-privacy-tips/, May 2010.
33. Diana I. Tamir and Jason P. Mitchell. Disclosing information about the self is intrinsically rewarding. In *Proceedings of the National Academy of Sciences of the United States of America*, May 2012.
34. Jie Tang, Tiancheng Lou, and Jon Kleinberg. Inferring Social Ties across Heterogenous Networks. In *5th ACM Symposium on Web Search and Data Mining*, 2012.
35. Atul Varma. Collusion. http://www.toolness.com/wp/2011/07/collusion/, 2011.
36. Yang Wang, Gregory Norcie, Saranga Komanduri, Pedro Giovanni Leon, Lorrie Faith Cranor, and Alessandro Acquisti. "I regretted the minute I pressed share": A Qualitative Study of Regrets on Facebook. In *Symposium on Usable Privacy and Security*, 2011.
37. Todd Wasserman. Google Engineer Accidentally Posts Rant about Google+. http://mashable.com/2011/10/12/google-engineer-rant-google-plus/, October 12 2011.
38. Mark Zuckerberg. One Billion People on Facebook. http://newsroom.fb.com/News/One-Billion-People-on-Facebook-1c9.aspx, October 2012.