Contextual Identity: Freedom to be All Your Selves

Monica Chew, Sid Stamm
Mozilla
{mmc, sid}@mozilla.com

***Abstract*—We examine *contextual identity*, a notion that individuals reveal different aspects of themselves depending on context. At any given time, an individual may act as a friend, relative, spouse, co-worker, acquaintance, or stranger. We analyze contextual identity from the perspective of user choice and control, survey contextual identity violations, and propose new research directions to enable users to have better control over their contextual identities.**

I. Introduction

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

The desire for privacy balances the desire for spontaneous, positive human interaction and sharing personal information. The act of sharing alone does not negate expectation of privacy [8]. Helen Nissenbaum has long argued that privacy violations come not from the simple sharing of personal information itself, but rather from sharing that information in a way that violates social norms [33].

For example, lesbian, gay, bisexual, and transgendered youth are particularly at risk for contextual identity violations. Access to support is crucial for LGTB-identified youth, who are at higher risk for bullying and suicide [4], and much of that support is only accessible through the Internet in socially conservative areas. In 2011 Bobbi Duncan and Taylor McCormick, two students at the University of Texas at Austin, were inadvertently outed when their choir director added them to a Facebook group for Queer Chorus [13]. Even though both students had blocked their parents from seeing any personal posts revealing their sexuality, Queer Chorus is a public group and so Facebook published notifications to Duncan's and McCormick's friends, including their parents. Duncan later attempted suicide [11].

Managing contextual identities is more intuitive in offline environments than in online environments. Offline, 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. In offline environments, humans don't have perfect memories and will typically forget. However, with the advent of increasingly vast searching, indexing, and archiving capabilities, one cannot rely on forgetting in online environments [25]. Managing contextual identities online has become increasingly difficult and fraught with mistakes; we must provide users with better tools to control their contextual identities.

II. Related Work

The notion that individuals inhabit multiple personas has existed in multiple fields for many years; Erving Goffman called this impression management, and Carl Jung called it persona theory [14], [18].

Barth et al. provide a formal model for contextual integrity of user data [3]. Their work models the flow of personal information in terms of knowledge states held by agents acting in given roles, and works well for expressing privacy legislation such as HIPPA and COPPA. Two limitations of this work are that it does not model information where the subject involves more than one person, and it only takes the type of information into account rather than the content or tone of the information. For example, "Bob told Alice that Charlie cooked a delicious dinner" is out of scope of the model. In the context of social networks, information is rarely about a single person, and the tone of the message is as important to its effect on a user as the contents.

Many authors have discovered methods to link different contextual identities to an individual. Narayanan and Shmatikov present a re-identification technique to merge anonymized social graphs from different networks and prove that user identifiers from Netflix and IMDB and Flickr and Twitter can be linked [31], [32]. Lindamood et al. and Mislove et al. show how to infer previously undisclosed information from public social networking data [21], [28].

Many authors have noted how re-broadcasting information information out of context and making information discoverable can lead to distress, even if the information was previously public, but not easily discoverable [9], [5], [33].

Evidence suggests that a person's major privacy concerns come from others he or she knows: friends, family, and co-workers [12], [37]. Surveys conducted by Wang et al. suggest that typical regrets from posting on Facebook stem almost *exclusively* from fear of negative interactions with people the individual knows [40]. The consequences of these negative interactions can lead to loss of employment or breaking personal relationships. Most previous privacy research has focused on distant adversaries, where the bad actor is a behavioral tracking service, state agent or unknown eavesdropper; in contrast, we assert that people are concerned about their interactions with others close to them.

III. Contextual identity violations

In this section we discuss three types of *contextual identity violations*. We consider a contextual identity violation to be when multiple identities are linked without owner intent, or when someone cannot choose which identity to assert in a given context. In these examples the individual may not be conscious of what a contextual identity is: nevertheless each example illustrates a loss of control over what aspects of self to reveal in a given context.

Although additional access control and improved usability can improve contextual identity management, they are not a panacea: many of the below examples fall out of scope of traditional access control models.

A. Redistributing information in a different context

Redistributing information out of its original context often leads to embarrassment [33]. In November 2007, Facebook Beacon allowed third-party sites to publish purchases, travel bookings, movie rentals and more to the user's activity stream. Difficulty of opt-out and lack of visibility into what was being published led to user surprise and dismay [26], [30]. In December 2007, Google Reader exposed RSS feeds of user-marked news stories to the user's Google Talk contacts, which includes everyone with whom a user has chatted whether they be co-workers, supervisors, or friends. Although this feed was always public, prior to this launch it was not discoverable, leading to a bad experience for many users [15]. In February 2010, Google Buzz launched, a product that exposed the user's most frequent Gmail and Google Talk contacts and their publicly available (through previously less discoverable) news and photos. The combination of exposing contacts (which could include abusive ex-husbands, co-workers, and friends), as well as aggressively linking photos and news items, resulted in a huge backlash [16], [17]. 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 people used to treat as private messages before the advent of "Like" and comment buttons, to the attention of an audience that the posting party never anticipated [22].

B. Unaccommodating policies

Some service providers have policies which preclude isolating multiple contextual identities. For example, Facebook and Google have a "Real Names" policy, which requires people to register for accounts with their legal name.[1] These policies ignore that community-building happens in many different contexts, that individuals have legitimate reasons for presenting different identities in different contexts, and they don't necessarily want those identities to be linked. For example, disallowing avatar handles as a primary identifier makes building a gaming community difficult. It is impossible to isolate multiple contextual identities in these networks without violating the terms of service.

Even worse, many of these providers now serve as login platforms for external sites. For example, Facebook Connect allows third-party websites to authenticate individuals using their Facebook identity [29]. Because it is against the terms of service to have multiple Facebook accounts, using Facebook Connect may have the unwanted side-effect of linking multiple contextual identities across multiple sites.

Other federated login systems have support for multiple identifiers, but this feature can be poorly implemented or difficult to discover, leading to low use. Some login platforms and protocols, such as OAuth and BrowserID, do not suffer from this policy or design error, giving a user better control over which contextual identity to assert [1], [6].

C. Confusing user interfaces

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", 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 published compromising pictures of himself [34]. Considering that this mistake requires mistyping a single character (`@` instead of `d`), it's no surprise that DM failures are so 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, it is common to have multiple accounts for personal and business use. KitchenAid, Chrysler, and Google are three companies whose employees have made this mistake in the past year [27], [35], [41].

IV. Research directions

We propose new research directions based on the following questions:

- How do users think about identity?
- How do users manage identity?
- How can we improve tools for managing identity?

A. How do users think about identity?

We hypothesize many people are not consciously aware of having multiple contextual identities. In order to develop helpful tools, we need to understand people's mental models of identity and how information is shared on the Internet. For computer scientists, authentication is intrinsically linked to identity, and even in the absence of authentication, using the same device over time implicitly creates an identity through tracking techniques and local information. However, these points are far from obvious to a typical web user, especially

[1] Facebook's policy: http://www.facebook.com/help/?page=258984010787183, Google's policy: http://support.google.com/plus/bin/answer.py?hl=en&answer=1228271

for those who don't distinguish the web application from the browser, from the operating system, or from the device.[2]

The computer science community must free itself from software-specific notions of identity if we are to help users to whom the very concept of identity is an enigma. To understand people's mental models, we might ask:

- How is identity represented online?
- What do you need to represent your identity?
- Is your identity tied to your device?
- If you check your mail, read news, log on to a social network at a computer at the public library, does that affect your identity?
- What does it mean to share devices? If you lend your tablet to your sister, is she representing herself, or you?
- If you register for an account on a service, do you expect that to have an influence on your online identity? What if you never use that account again?
- If you visit a website, do you expect that to have an influence on your browsing experience or online identity? If so, for how long?
- Which parts of your online identity do you expect to be visible to your housemates, friends, relatives, or employers?
- Which parts of your online identity do you expect to be visible to websites you use? What about websites you don't use?
- Which parts of your identity do you want to share or keep secret, and from whom?
- How does your identity change over time?

Pew Internet is a good source for phone survey data on privacy and social media, and danah boyd's work on youth and social media is excellent, but there is otherwise a scarcity of research in these areas [7], [10], [23], [24].

B. How do people manage identity?

Those who are aware of having multiple identities engage in the following techniques to separate, link, and curate identities. These techniques are by no means complete or optimal: people may have latent needs that are not met by current tools. Studying usage and common problems of these techniques is a worthy goal.

1) Separating identities: The following examples illustrate how individuals attempt to preserve boundaries between identities. None of these techniques are fail-safe, but they serve as useful reminders that perfect solutions may not be necessary or even desirable, especially in the case where information leaks by casual or accidental inspection.

Multiple accounts: The long-time existence of support for multiple accounts in email clients suggests that many people have multiple email identifiers, which could be considered as a proxy for identity if tied to service accounts. Similarly, data suggests that a large minority of Twitter users have multiple Twitter accounts [20].

Multiple browsers: Using multiple browsers is a useful technique for managing multiple accounts. Some services support multiple accounts but not multiple login (e.g., Twitter). For services that do, the user interface may be so confusing that it leads to errors, so the best recommendation may be to use multiple browsers [41].[3] For those who want to separate work and personal browsing, using multiple browsers is the easiest solution.

Multiple devices: Many people have multiple devices and many use them for different purposes. Some of this difference is due to the nature of the device (e.g., GPS and mapping software are probably more often used in mobile devices) and some may be due to policy (e.g., limiting work activities to a corporate laptop). Using multiple devices implicitly creates multiple browser states and thus multiple identities. However, these identities may still be linked through authentication data or other techniques.

Multiple profiles: Several browsers support separation of profile data, including cookies, history, passwords, and other local storage. Firefox and Chrome support multiple profiles, though neither of these implementations is discoverable or easy to use.[4] The threat model does not include users with malicious intent in the same household.

Private browsing mode: Private browsing mode exists in all major browsers, but has no standard behavior [2]. Interaction with extensions, treatment of cookies, history, and bookmarks upon entering and exiting are different across browsers. Similar to multiple profiles, private browsing mode is not secure against all local attacks.

Cookie blockers: Disconnect and ShareMeNot are browser extensions to disallow interacting with service providers like Google and Facebook unless users explicitly choose to interact with that site [19], [36]. Collusion is a tool for visualizing and blocking cookies [39], in particular third-party cookies set by tracking sites which are typically ad networks.

2) Linking identities: For many users, the natural separation that occurs when creating multiple accounts, using multiple browsers and devices is a drawback, not a benefit. These users want fewer contextual identities and use the following techniques to merge them.

Building social graphs: Users can explicitly create links between contextual identities, e.g. linking to their blog from their Flickr profile, or resharing a blog post via Twitter. It is also easy to implicitly create links between contextual identities, sometimes accidentally: generating two isomorphic graphs at different services is often enough to re-identify an individual [31]. The plethora of cross-posting software suggests that users often want to link identities.

[2] What is a browser? https://www.youtube.com/watch?v=o4MwTvtyrUQ

[3] Using multiple browsers as an alternative to multiple login: https://support.google.com/accounts/bin/answer.py?hl=en&answer=179235

[4] Chrome: https://support.google.com/chrome/bin/answer.py?hl=en&answer=2364824, Firefox: https://bugzilla.mozilla.org/show_bug.cgi?id=214675#c53

Building reputation online is a long and arduous process; a user who has built up high credibility in one social network may want to transfer that credibility when using a new service by using the same identifier or somehow proving in the new service that they control other credible identities. The advent of verified account mechanisms in Twitter, Google Plus and Facebook suggests this problem is on the rise.

Synchronizing browser data: All five major browsers (Firefox, Chrome, Safari, Internet Explorer, and Opera) offer synchronizing a subset of browser data across devices. Multiple cross-browser applications also exist to synchronize bookmarks and passwords.[5]

3) Curating identities: Given the large number of social network users and that 30-40% of spoken communication is devoted to informing others about ourselves, many users are bound to share information they later regret [38].

Service settings: Features such as Facebook friends, Google circles and privacy settings presumably allow users to manage their privacy and identity. However, shifting implementations and complex, interacting features make these configurations unpredictable for many users. According to Pew Internet, 71% of users change privacy-related settings; yet the large number of users who experience regret on social networks indicates that these settings are not working as intended [24].

Auditing: Wang et al. suggest that common techniques for handling regret online include manual deletion of regrettable posts, self-censoring or delaying posts that they predict might bring regret [40]. Pew Internet reports that 57% of Internet users periodically search for themselves to manage reputation, with young adults being the most active [24].

C. How can we improve identity management tools?

The multiplicity of social media means that in order to be competitive, social networks encourage cross-linking, resharing, and resyndication — all of which promote linking identities. Linking identities is already easy; users need more tools to help separate and curate identities. Without knowing more about how people think about identity, we concentrate solely on automating manual processes that users already perform.

Mitigating accidental linkage: Federated login services can prevent linking contextual identities to the same user identifier. For example, a federated login service that supported multiple identifiers could prevent the user from associating the same identifier to radically different contextual identities, such as ones for dating and professional use.

Auditing: A browser is in an ideal position to intermediate social network posts, aggregate them locally, and present them to its user when she wants to audit her digital footprint. For example, people could inspect all of the comments and posts they made in the last week and redact content that is sensitive or too negative, a frequent cause of regret [40]. Such a tool could incorporate sentiment analysis to find particularly problematic content.

Expiring posts: Humans don't have perfect memories, and we question why social networks do [25]. Many users already engage in manual auditing and deletion of old posts [12]. A better solution would be to build tools that let the user manage this more easily through an API: both Twitter and Facebook provide APIs for post deletion.

Expiring contextual identities: Engaging in a long-term task such as house-buying often requires the construction of a contextual identity. The tasks associated with this identity have an externally imposed end date and may have long-lived side-effects, such as long-term cookies that reveal sensitive information (e.g., the neighborhood of the house purchase, income information derived from purchases, other demographic data). For contextual identities that have outlived their usefulness, people would benefit from destruction of side-effects such as cookie data and service accounts.

V. Summary

People have multiple contextual identities for many different reasons. Sharing personal information in appropriate contexts fosters positive human interaction and doesn't negate the need or desire for privacy. We hope that contextual identity serves as a useful notion for developers to understand their users' privacy needs and seek to understand identity use better so we can develop tools for effective identity management.

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] Human Rights Campaign. Growing up LGBT in America. http://www.hrc.org/files/assets/resources/Growing-Up-LGBT-in-America_Report.pdf, June 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.

[5] An example of bookmark sync is xmarks.com, one cloud-based password manager is lastpass.com.

[7] danah boyd. Why Youth (Heart) Social Network Sites: The Role of Networked Publics in Teenage Social Life. In David Buckingham, editor, *MacArthur Foundation Series on Digital Learning — Youth, Identity, and Digital Media Volume*. MIT Press, 2007.

[8] danah boyd. Making Sense of Privacy and Publicity. http://www.danah.org/papers/talks/2010/SXSW2010.html, March 2010. SXSW, Austin, Texas.

[9] danah boyd. Privacy and publicity in the context of big data. http://www.danah.org/papers/talks/2010/WWW2010.html, 2010. WWW, Raleigh, NC.

[10] Maeve Duggan and Joanna Brenner. The Demographics of Social Media Users — 2012. http://pewinternet.org/Reports/2013/Social-media-users.aspx, February 14 2013.

[11] Bobbi Duncan. Second Chance. http://ridingincarswithducks.com/2012/01/second-chance/, January 9 2012.

[12] Alicia Eler. Top 5 Facebook Privacy Tips. http://www.readwriteweb.com/archives/top_5_facebook_privacy_tips.php, April 2012.

[13] Geoffrey A. Fowler. When the Most Personal Secrets Get Outed on Facebook. http://online.wsj.com/article/SB10000872396390444165804578008740578200224.html, October 13 2012.

[14] Erving Goffman. *The Presentation of Self in Everyday Life*. Anchor, 1959.

[15] 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.

[16] Harriet J. Fuck you, Google. http://www.fugitivus.net/2010/02/11/fuck-you-google/, February 11 2010.

[17] Todd Jackson. Millions of buzz users, and improvements based on your feedback. http://gmailblog.blogspot.com/2010/02/millions-of-buzz-users-and-improvements.html, February 2010.

[18] C. G. Jung. *Two Essays on Analytical Psychology*. London, 1953.

[19] Brian Kennish. Meet Disconnect. http://byoogle.blogspot.com/2010/12/meet-disconnect.html, December 2010.

[20] Ilya Kochanov. How Many Twitter Accounts Do You Have? http://techcrunch.com/2008/01/09/how-many-twitter-accounts-do-you-have/, January 9 2008.

[21] Jack Lindamood, Raymond Heatherly, Murat Kantarcioglu, and Bhavani Thuraisingham. Inferring private information using social network data. In *WWW*, 2009.

[22] 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.

[23] Mary Madden. Privacy management on social media sites. http://pewinternet.org/Reports/2012/Privacy-management-on-social-media.aspx, February 24 2012.

[24] Mary Madden and Aaron Smith. Reputation Management and Social Media. http://www.pewinternet.org/Reports/2010/Reputation-Management.aspx, May 26 2010.

[25] Viktor Mayer-Schoenberger. *Delete: The Virtue of Forgetting in the Digital Age*. Princeton University Press, July 2011.

[26] Caroline McCarthy. MoveOn.org takes on Facebook's 'Beacon' ads. http://news.cnet.com/8301-13577_3-9821170-36.html, November 20 2007.

[27] 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.

[28] 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.

[29] Dave Morin. Announcing Facebook Connect. http://developers.facebook.com/blog/post/2008/05/09/announcing-facebook-connect/, May 9 2008.

[30] Ellen Nakashima. Feeling Betrayed, Facebook Users Force Site to Honor Their Privacy. http://www.washingtonpost.com/wp-dyn/content/article/2007/11/29/AR2007112902503.html, November 30 2007.

[31] Arvind Narayanan and Vitaly Shmatikov. Robust De-anonymization of Large Sparse Datasets. In *IEEE Symposium on Security and Privacy*, 2008.

[32] Arvind Narayanan and Vitaly Shmatikov. De-anonymizing Social Networks. In *IEEE Symposium on Security and Privacy*, 2009.

[33] Helen Nissenbaum. *Privacy in Context: Technology, Policy, and the Integrity of Social Life*. Stanford University Press, 2010.

[34] 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.

[35] Erik Qualman. Chrysler Fires Agency over F*** Tweet. http://www.socialnomics.net/2011/03/14/chrysler-fires-agency-over-f-tweet/, March 14 2011.

[36] Franziska Roesner, Chris Rovillos, Tadayoshi Kohno, and David Wetherall. ShareMeNot. http://sharemenot.cs.washington.edu/, July 2011.

[37] Adam Rosenberg. 5 Essential Facebook Privacy Tips. http://mashable.com/2010/05/18/facebook-privacy-tips/, May 2010.

[38] 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.

[39] Atul Varma. Collusion. http://www.toolness.com/wp/2011/07/collusion/, 2011.

[40] 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.

[41] Todd Wasserman. Google Engineer Accidentally Posts Rant about Google+. http://mashable.com/2011/10/12/google-engineer-rant-google-plus/, October 12 2011.