A Privacy Case Study: Google+ and Profiles

Shiqian Xu
Department of Computer Science
North Carolina State University
Raleigh, NC, USA
sxu11@ncsu.edu

Qian Liu
Department of Computer Science
North Carolina State University
Raleigh, NC, USA
qliu10@ncsu.edu

Feifei Wang
Department of Computer Science
North Carolina State University
Raleigh, NC, USA
fwang12@ncsu.edu

Neng Jiang
Department of Computer Science
North Carolina State University
Raleigh, NC, USA
njiang4@ncsu.edu

Jessica Staddon
Department of Computer Science
North Carolina State University
Raleigh, NC, USA
jessica.staddon@ncsu.edu

Abstract

Google+ was designed with privacy in mind in that it is built around the concept of ”circles”, a model for selective sharing. We present a preliminary case study of Google+ privacy, focusing on profile field visibility including gender, occupation and photos.

1. Introduction

Case studies of privacy products and policies are an important tool for understanding how to design for privacy. Patterns detected across case studies can identify areas for improvement in product and policy. For example, negative responses to default settings in social media [7, 8, 15] can serve to identify both ways to set defaults to better align with user preferences and design guidance for making users aware of defaults. Similarly, examples of privacy policies found to be vague or surprising in the data collection and use practices they describe [11, 2], can help policy authors produce language that is more clear and intuitive.

Google+, the social network launched by Google in 2011, is a useful product on which to base a case study because private sharing is a prominent part of its design. Google+ is built around the concept of “circles” which enable users to “share selectively” [1].

At a minimum, a privacy case study should include: (1) the design of the product or policy and any evolution over time (e.g. as tracked here in the case of Facebook settings [12]), (2) user perception and comprehension of the product/policy, and (3) user behavior related to the product/policy.

age_max	age_min
birthday	circledByCount
display_name	emails
first_name	gender
image_is_default	image_url
isPlusUser	language
last_name	occupation
org1	org2
org3	place1
place2	place3
relationship	skills
url	verified
about_me	

Table 1: Profile features available through the Google+ API

In this poster, we provide work-in-progress on the third case study component; we analyze user behavior using the Google+ API [4]. While usage of circles for sharing has been widely reported (e.g., [3]), less is known about the visibility of profiles. Based on a sample of more than $30,000$ Google+ profiles we find that, with the exception of profile photos, men tend to publicly expose more information than women, but that regardless of gender, exposure is greatest amongst those who are in a lot of circles. We also find overall exposure rates that are less than generally reported in Facebook studies, however many Facebook studies focus on exposure within a university subnetwork of Facebook, whereas we consider public exposure.

2. Google+ user profile data analysis

The data described here were gathered in December 2015, when Google+ had 418 million active users. Because the Google+ API [4] does not support random sampling we approximated a random sample by gathering the profiles of users with the top 10 most popular surnames[13, 17] in 5 countries: the United States, France, Germany, Japan and China. This resulted in $33,818$ user profiles. For each profile, we retrieved the 25 profile fields shown in Table 1, for profiles in which the fields are publicly visible (no private fields are accessible via the API). The variable names in Table 1 are generally self-explanatory with the exceptions of age_max and age_min which indicates the user age range, circledByCount which is the number of circles the user is in, org 1-3, which are the user-reported most recent places of employment or education, place 1-3 are three most recent places the user reports to have lived, and about_me is the tagline field of a profile.

USA	25132	74.3%
China	2367	7.0%
France	1276	3.8%
Germany	2511	7.4%
Japan	2532	7.5%

Table 2: Demographic Distribution

2.1 Demographics information

Out of $33,818$ records, 88.2% self-reported a gender; 58.5% as male and 29.7% female.

The most popular reported profile locations in our sample are California, New York and London (see Table 2) and the most popular occupations are software engineer, photographer and student. The top three most popular organizations are UC Berkeley, Stanford and UCLA.

2.2 Profile patterns

We define a user's profile completion percentage (PCP) as the fraction of the 25 fields that are *publicly visible* in their profile. In our sample, the median PCP is 64%. We say a profile has *high* PCP if it's completion percentage exceeds the median and low otherwise. We term a PCP, "ex-high" if it is more than 75%. In our sample, 56.5% of profiles have high PCP and 12.6% have extreme-high PCP.

We term the number of circles that contain a user as the user's *social circle size*, and we say a social circle size is *big* if it exceeds the median social circle size in our sample, 626, and *small* otherwise.

2.2.1 Gender and Social circle effects

Our gender analysis focuses on the "male" and "female" gender options (ignoring the "custom" and "decline to state" options that were little used in our sample, about 11.2% in total).

Using two-way ANOVA tests we find main effects of gender and social circle (p-value < 0.01), but no interaction effect between gender and social circle (p-value $= 0.23$). The mean profile completion is higher for users with bigger social circles ($\mu_{bigCircle}$ = 0.632, $\mu_{smallCircle}$ = 0.628)

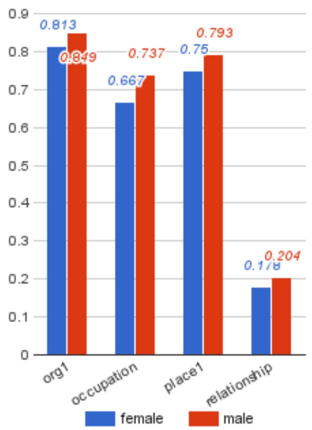

Figure 1: example completion rates by gender

	User PCP	
Feature	*High*	*ex-High*
place3	.933	.311
place2	.891	.254
org3	.877	.260
skills	.834	.325
relationship	.823	.396
birthday	.813	.418
org2	.761	.184
occupation	.733	.176
place1	.700	.162
org1	.667	.151
PCP	.565	.126

Table 3: Feature correlation.

and the mean profile completion of males is higher than for females (μ_{male} = 0.637, μ_{female} = 0.616). Figure1 shows gender impact for 4 example features: org1, occupation, place1, and relationship.

2.2.2 Relationship between PCP and Specific Features
We use the conditional probability of a having high PCP given that a particular profile field is publicly visible, to measure the relationship between features and the PCP.

Table 3 shows that the probability of having a high PCP given Place3 is publicly visible in a profile, is large ($.93$). Consequently, the single profile field, Place3, may be a good indicator of Google+ engagement. The probability of having ex-high PCP given the birthday field (month and day) is publicly visible, exceeds 40%. Indeed, users who expose both birthday and relationship status are far more likely to have a complete profile.

2.2.3 Profile Photos
In our sample, nearly 99.2% of the users changed their default profile photo to a customized one. Among the photos they uploaded, 66% of them contain a real face according to the third party API , Face++ [6]. We verified the accuracy of the Face++ API on a hand-curated sample and found a precision of $.806$ and recall of $.833$.

In contrast to the text-based profile fields, we find a slightly lower percentage of photos containing a face amongst males than females, 63.88% and 71.08%, respectively. Furthermore, we find that users with a larger social circle size are more likely to have a human face in their profile photo. Users who are circled by more than $50,000$ other users have a human face in their photo at a rate of $.745$, whereas users who are in less than 50 circles have a human face in their photos at a rate of $.62$.

	User Profile Visibility	
Profile Field	**Default Setting**	**Fraction Public**
Birthday	Your Circles	.081
Employment (Org1)	Public	.837
Gender	Public	.876
Location	Public	.779
circledByCount	Public	.874
Occupation	Public	.713
Places Lived (Place1)	Public	.779
Relationship	Extended Circles	.195
Skills	Public	.252
Tagline	Public	0

Table 4: Fields in Google+ profiles, their default visibility and the fraction of users who have public content in each field.

2.2.4 Profile Field Visibility
One indication of engagement with privacy settings is modification of defaults. Our initial analysis has not found much evidence of modification. In Table 4, the only two settings with non-public defaults (Birthday and Relationship) are also the least disclosed. The fact that many of the fields that are public by default are still public is further evidence that many users do not modify the visibility settings.

Note that in Table 4, the “Gender” field includes the fraction of users who made an entry of “male” or “female” public on their profile.

Related Work
Privacy in online social networks is a well-studied area, particularly in the context of Facebook. For example, in a seminal paper, Acquisti and Gross [5] find a high rate of personal information within a university subnetwork of Face-

book. This work continues in [16], which tracks information sharing in the CMU Facebook network over many years. While these papers generally find a remarkably high rate of personal information sharing (e.g. birthday is present in almost 90% of the profiles analyzed in [5]) they aren't directly comparable to our work which looks at profile fields that are public on the web rather than within a university network.

Both [14] and [10] analyze user Facebook data that is public on the web, but their focus is exploring how well privacy preferences match current privacy settings and they don't provide statistics on the visibility of the profile fields considered here.

Our work is similar to [18], which studies Facebook privacy settings in a population of 297 Florida college students both before and after an intervention during which students were informed about their college's social media policy. They find that "personal information pages" are publicly visible (not just within the university network) at rate of $.995$, but the authors do not describe the personal attributes they consider.

Limitations

Google+ API Limitations. Google+ API does not allow access to some profile information such as email and accurate age. These restrictions demonstrate another aspect of Google+'s privacy design, but they also limit the scope of our analysis.

Sample. Our sample, while substantial, was not selected at random and may not represent the population. In addition, we have not analyzed all of the fields available through the Google+ API.

Evolution and timing of Google+. While the privacy features of Google+ were emphasized at launch and may have attracted users, the network did not always evolve in a privacy-aware direction. In particular, shortly after launch Google+ began enforcing a "real names" policy, which resulted in many users losing access to their accounts before the policy was relaxed [9]. This policy was widely criticized on privacy grounds because it made it difficult for users to maintain a different identity in Google+ than in the physical world, particularly if that different identity did not appear to be a conventional name. Hence, while privacy was a core part of the initial Google+ design, other factors significantly influenced its evolution.

We also note that Google+ wasn't introduced until 2011, after Facebook was well-established. Given this, Google+ has likely drawn users from a different pool than Facebook and behavioral differences in Google+ may be impacted by differences in the underlying population as well as by Google+ design.

Conclusion and Future Work

We have presented an initial analysis of profile visibility in Google+ as part of a case study of a network created with privacy in mind. Our research so far has been descriptive, that is we do not have data to determine the whether profile information is withheld for privacy reasons, and if the degree of exposure meets user needs; both of which are important to assessing the success of Google+ from a privacy standpoint. In future work, we will explore user motivations will enlarge our data set to more comprehensively identify Google+ behavior patterns.

References

[1] Official Google Blog. 2011. (28 June 2011). Retrieved on January 10, 2016 from https://googleblog.blogspot.com/2011/06/introducing-google-project-real-life.html.

[2] A. Couts. 2012. Terms & Conditions: Facebook's "Data

Use Policy" explained. (2012). Retrieved on January 10, 2016 from http://www.digitaltrends.com/social-media/terms-conditions-facebooks-data-use-policy-explained/.

[3] Liz Gannes. 2011. More Than Two-Thirds of Google+ Activity Is Private. (21 July 2011). Retrieved on January 10, 2016 from http://allthingsd.com/20110721/more-than-two-thirds-of-google-activity-is-private/.

[4] Google. 2016. Google+ API. (2016). Available at: https://developers.google.com/+/web/api/rest/ [Accessed 7 Jan. 2016].

[5] R. Gross and A. Acquisti. Privacy and information revelation in online social networks *(Proceedings of themACM CCS Workshop on Privacy in the Electronic Society (WPES âĂŹ05))*.

[6] Megvii Inc. 2015. Face++ API. (2015). Available at: http://www.faceplusplus.com/ [Accessed 7 Jan. 2016].

[7] A. Khanna. 2015. Facebook's Privacy Incident Response: a study of geolocation sharing on Facebook Messenger *(Technology Science. 2015081101.)*.

[8] Micah Lee. 2012. Privacy in Ubuntu 12.10: Amazon Ads and Data Leaks. (29 October 2012). Retrieved on January 11, 2016 from https://www.eff.org/deeplinks/2012/10/privacy-ubuntu-1210-amazon-ads-and-data-leaks.

[9] R. MacKinnon and H. Lim. 2014. Google Plus Finally Gives Up on Its Ineffective, Dangerous Real-Name Policy. (17 July 2014). Retrieved January 9, 2016 from http://www.slate.com/blogs/future_tense/2014/07/17/google_plus_finally_ditches_its_ineffective_dangerous_real_name_policy.html.

[10] Michelle Madejski, Maritza Johnson, and Steven M. Bellovin. 2011. The Failure of Online Social Network Privacy Settings. (February 2011).

[11] A. Mamiit. 2015. Spotify Revamps New Privacy Policy After User Backlash. (5 September 2015).

[12] Matt Mckeon. 2016. The Evolution of Privacy on Facebook: Changes in default profile settings over time. (2016). http://mattmckeon.com/facebook-privacy/.

[13] mongabay.com. 2015. What is the most common last name in the United States? (2015). Available at: http://names.mongabay.com/data/1000.html [Accessed 10 Jan. 2016].

[14] Michael Netter, Moritz Riesner, Michael Weber, and Günther Pernul. 2013. Privacy Settings in Online Social Networks - Preferences, Perception, and Reality. In *46th Hawaii International Conference on System Sciences, HICSS 2013, Wailea, HI, USA, January 7-10, 2013*. IEEE, 3219–3228. DOI:http://dx.doi.org/10.1109/HICSS.2013.455

[15] L. Rao. 2011. Sexual Activity Tracked By Fitbit Shows Up In Google Search Results. (3 July 2011). http://techcrunch.com/2011/07/03/sexual-activity-tracked-by-fitbit-shows-up-in-google-search-results/.

[16] Fred Stutzman, Ralph Gross, and Alessandro Acquisti. (2013). Silent Listeners: The Evolution of Privacy and Disclosure on Facebook *(Journal of Privacy and Confidentiality: Vol. 4: Iss. 2, Article 2)*.

[17] Wikipedia. 2015. Common names of China,France,Germany and Japan. (2015). [China]:https://en.wikipedia.org/wiki/List_of_common_Chinese_surnames,[France]:https://en.wikipedia.org/wiki/List_of_the_most_common_surnames_in_Europe,[Germany]:https://en.wikipedia.org/wiki/List_of_the_most_common_surnames_in_Germany,[Japan]:https://en.wikipedia.org/wiki/List_of_most_common_surnames_in_Asia.

[18] J Williams, C Feild, and K. James. The Effects of a Social Media Policy on Pharmacy Students' Facebook Security Settings *(American Journal of Pharmaceutical Education. 2011;75(9):177)*.