Improving Privacy through Exposure Awareness and Reactive Mechanisms

Apu Kapadia
School of Informatics and Computing
Indiana University
kapadia@indiana.edu

Adam J. Lee
Department of Computer Science
University of Pittsburgh
adamlee@cs.pitt.edu

Abstract

As people share increasing amounts of information with their online social networks, managing which information should be sent to whom is cumbersome. As opposed to having people prespecify access control 'policies' we advocate that system designers should consider two important aspects: 1) systems should provide 'exposure awareness' to users so that they are aware of *real* access patterns to their data, and 2) reactive mechanisms that can dynamically adjust the sharing of information based on a person's exposure. We summarize our research in the area and outline research challenges that lie ahead.

Introduction

With the simultaneous rise of smartphones, sensing devices, and social networking, people are now sharing an unprecedented amount of personal information with their social and professional contacts. Sensors within or coupled with smartphones enable various applications, such as location-based services, fitness applications, sleep monitors, and smart home thermostats, and people can share related information such as location, activity, fitness, and health information with their social networks. Moreover, such 'contextual' information is now increasingly attached to textual status updates posted to social networks; e.g., posts to social-networking services like Twitter, Facebook, and Google+ can include location

information attached by default or mood information on Facebook. Yet, people are finding it harder to control the extent to which their information is shared and often regret having shared certain information because of unanticipated events [16, 19].

Despite various privacy settings available on social-networking services, people do not adequately utilize these controls [2, 4]. Even if people try to use the available static controls (i.e., privacy policies), they are faced with two problems: 1) the many categories of experience that could conceivably be inferred by sensors make it *difficult or impossible to articulate an exhaustive list* that would instantiate an adequate privacy policy, and 2) people simply *cannot anticipate all possible access and usage scenarios* of their data beforehand with such 'set-and-forget' policies. To alleviate the first problem, Facebook [5] and Google+ [7] (through Facebook Lists [12] and Google+ Circles [1]) allow users to organize their social contacts and characterize their relationships to improve the target audience of their shared information, but these mechanisms suffer from the second problem because people cannot anticipate all target audiences beforehand [8]. Even if users take the time to construct customized lists for particular types of data, they may not have anticipated how such data will be accessed by the listed individuals — consider Alice who authorizes her boss to check her location during work hours but then later realizes her boss has been checking her location every 20 minutes. Thus, a better approach is needed that makes it tractable for people to control how, and to what extent, their information is accessed.

The above issues highlight the difficulties associated with achieving the vision of 'modular privacy' noted in a recent report from the CCC Privacy Enabling Design workshop [9]. Tailoring the sharing of potentially sensitive information to a wide audience is a difficult task in and of itself, which is compounded by factors like dynamic contexts (which can alter user *perception* of that data), dynamic data (which can be hard to protect with fixed settings), and largely opaque information-sharing networks (which obscure the true usage of data). The vast majority of data sharing controls deployed today are ill-suited to such dynamism.

Our vision: 'Reactive Mechanisms' based on 'Exposure Awareness'

Expecting people to use 'set-and-forget' type policies for a vast array of personal data, anticipating all possible access and usage scenarios, is a strategy that is unlikely to succeed for managing privacy. Instead, we advocate a **reactive** approach, where people (or automated mechanisms acting on their behalf) can make privacy decisions in response to the **actual patterns** and use of their information. This approach has two advantages: 1) people need only care about the subset of data and usage scenarios that have the potential to violate their privacy, thus reducing the amount of data to which they must regulate access; and 2) people make better decisions concerning access to their information when these decisions are made in a context where they know how their data is being accessed and used. If these mechanisms are to be successful with a large population that uses such systems, then they need to be **user-friendly** and allow people to share their information in a **pragmatic way**. In other words, we must acknowledge that expecting people to anticipate all possible uses beforehand is 'too much, too early'. At the same time, mechanisms must allow people to rein in their privacy in a practical way by controlling the spread of their information adequately and before it is 'too late'. We believe this approach will open

a door towards more sophisticated and useful techniques for managing privacy in an increasingly socially networked world in which people desire to share ever-increasing amounts of personal information.

We believe to achieve this vision, 'privacy by design' should consider the following questions in the development of technology collecting and distributing a user's information with other users of the system. For each of these, we believe the industry can already build on these suggestions, and yet much remains to be done by academic researchers on these questions.

1. *Exposure feedback.* How can people be informed about the state of their privacy in a usable and intuitive way, and how can they respond to better manage their privacy?
2. *Reactive algorithms.* How effectively can automated algorithms detect and respond to changes in the patterns of data production and access to improve the privacy of users?
3. *Interactive data management.* How effectively can the privacy needs of users be balanced with the efficiency needs of data management systems?

Exposure feedback

We first discuss our experiences with providing users with 'exposure feedback'. Systems need privacy feedback mechanisms to compute, summarize, and convey the extent to which personal data is actually being used and accessed. Our past work has addressed the lack of exposure feedback (i.e., an answer to the question "Who is accessing my data?") in location-sharing systems. For example, we proposed an intuitive mechanism for summarizing and controlling a user's exposure on smartphone-based platforms [10, 15, 18]. Our approach uses the visual metaphor of *eyes* appearing and growing in size on the home screen; the rate at which these eyes grow depends on the number of accesses granted for a user's location and the type of person (e.g., family vs. friend) making these accesses. This approach gives users an accurate and ambient sense of their exposure and helps them take actions to limit their exposure, all without explicitly identifying the social contacts making requests.

To gain a deeper understanding on how and when to provide exposure feedback to people. we conducted an experience sampling study to examine various factors contributing to when and why people disagree with their own a priori privacy settings for actual location disclosures [17]. We found that *immediate* feedback about disclosures without any ability to control the disclosures evoked feelings of oversharing. Our followup study with the same dataset [14] found that delaying non-actionable feedback almost completely eliminated feelings of oversharing. Based on these findings, we suggest making immediate feedback more *actionable* and delaying non-actionable feedback to avoid a knee-jerk reaction. We also shed light on when such situations may apply, e.g., requiring more control when sharing information with distant social contacts or when visiting atypical locations.

Reactive algorithms

Looking forward, privacy exposure information is not only useful for direct consumption by users, but also by algorithms that can act on the behalf of users. The research community needs to explore algorithms that can dynamically control the release of personal information within the social network by reacting to actual usage patterns of information. We offer two simple approaches to spur discussion:

- Based on a textual analysis of a post, automated algorithms can detect anomalous patterns of interaction by social contacts (e.g., students 'Liking' a more personal Facebook post by a faculty member) and temporarily quarantine the post until the original poster takes further action (e.g., restricting the audience of the post). This approach relies on learning but we believe is within the realm of algorithms already in use by Facebook (e.g., detecting important memories, filtering newsfeeds based on past interactions, and so on).
- Based on the frequency of access to data (e.g., Alice's location) automated algorithms could allow some degree of anonymous access, but later require identifiable feedback to Alice, and eventually no access to the data. This approach would require straightforward thresholds and could be easily deployed today.

Interactive data management

Social networking is not the only domain in which the incorrect or imprecise capture of users' privacy preferences can have serious consequences. User privacy is often seen as being at odds with the core goals (e.g., efficiency, specificity, etc.) of more traditional data management systems, as well. While investigating this space, we have found that it is often possible to support user-centric privacy preferences while maintaining reasonable performance and accuracy in distributed data management systems. We have explored this problem in the context of workplace presence systems [3], distributed relational database management systems [6], and data stream management systems [11]. In addition to myriad systems issues that arise in this space, a key challenge involves the effective capture of individuals' privacy preferences. Individuals may not fully understand *how* the data that they provide to a system is actually being used, hence may under- or over-estimate the privacy harms that come with participation. We have developed several interactive query optimization interfaces for distributed databases in an effort to inform as well as react to individuals' privacy preferences [13], but there is much work to be done in this space.

Conclusion

We believe that properly responding to individuals' privacy needs will require reactive controls that leverage individuals' mental models of privacy. Given the increasingly contextual nature of information sharing, this approach can help fill the void left by more traditional 'set-and-forget' policies. In our work, we have explored some questions regarding the timing of user feedback, the design of various exposure controls, and methods for gathering users' privacy preferences in-the-loop. However, much work remains to be done if we are to truly put individuals in control of their personal information. For instance, we have only scraped the surface of the feedback issue: how can we balance users' need for feedback with a desire to minimize the cognitive burden of *managing* their data? Further, much of our work has focused on systems under our control. How can we adapt the lessons that we have learned to more opaque systems? What would incentivize the operators of these systems to design *for* privacy? We hope the privacy community will join us in exploring reactive privacy mechanisms to ultimately impact the design of real-world systems.

Acknowledgements

This abstract summarizes the vision and work for our joint "Exposure" project, for which the authors are principal investigators. We thank the numerous students and

collaborators who have contributed to the various instantiations of this ideal. This material is based upon work supported by the National Science Foundation under Grants CNS-1016603, CNS-1252697, CNS-1017229, and CNS-1253204. 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.

References

[1] *About Circles - Google+ Help*, (accessed June 21, 2012). http://support.google.com/plus/bin/answer.py?hl=en&answer=1047805.

[2] Acquisti, A., and Gross, R. Imagined communities: Awareness, information sharing, and privacy on the Facebook. *Privacy Enhancing Technologies 4258* (2006), 36–58.

[3] Biehl, J. T., Rieffel, E. G., and Lee, A. J. When privacy and utility are in harmony: towards better design of presence technologies. *Personal and Ubiquitous Computing 17*, 3 (2013), 503–518.

[4] danah boyd, and Hargittai, E. Facebook privacy settings: Who cares? *First Monday 15*, 8 (Aug. 2010).

[5] *Facebook*, (accessed June 21, 2012). http://www.facebook.com/.

[6] Farnan, N. L., Lee, A. J., Chrysanthis, P. K., and Yu, T. PAQO: preference-aware query optimization for decentralized database systems. In *IEEE 30th International Conference on Data Engineering (ICDE)* (2014), 424–435.

[7] *Google+*, (accessed June 21, 2012). http://plus.google.com/.

[8] Grimmelmann, J. Saving Facebook. *Iowa Law Review 94* (2009), 1137–1206.

[9] Hemmings, J., Pichon, M. L., and Swire, P. Privacy by design-privacy enabling design, workshop 2 report. Tech. rep., Computing Community Consortium, 2014.

[10] Hoyle, R., Patil, S., White, D., Dawson, J., Whalen, P., and Kapadia, A. Attire: Conveying information exposure through avatar apparel (demo). In *Proceedings of The 2013 ACM Conference on Computer Supported Cooperative Work Companion (CSCW)* (Feb. 2013), 19–22.

[11] Katsipoulakis, N. R., Thoma, C., Gratta, E. A., Labrinidis, A., Lee, A. J., and Chrysanthis, P. K. Ce-storm: Confidential elastic processing of data streams. In *Proceedings of the 2015 ACM SIGMOD International Conference on Management of Data* (2015), 859–864.

[12] *Lists for Friends - Facebook Help Center*, (accessed June 21, 2012). http://www.facebook.com/help/friends/lists.

[13] Ong, N. R., Rojcewicz, S. E., Farnan, N. L., Lee, A. J., Chrysanthis, P. K., and Yu, T. Interactive preference-aware query optimization. In *31st IEEE International Conference on Data Engineering (ICDE)* (2015), 1512–1515.

[14] Patil, S., Hoyle, R., Schlegel, R., Kapadia, A., and Lee, A. J. Interrupt now or inform later?: Comparing immediate and delayed privacy feedback. In *Proceedings of The ACM SIGCHI Conference on Human Factors in Computing Systems (CHI '15)* (Apr. 2015), 1415–1418.

[15] Patil, S., and Kapadia, A. Are you exposed? conveying information exposure (extended abstract). In *Proceedings of The 2012 ACM Conference on Computer Supported Cooperative Work Companion (CSCW)* (Feb. 2012), 191–194.

[16] Patil, S., Norcie, G., Kapadia, A., and Lee, A. J. Reasons, rewards, regrets: Privacy considerations in location sharing as an interactive practice. In *Proceedings of The 2012 Symposium on Usable Privacy and Security (SOUPS)* (July 2012).

[17] Patil, S., Schlegel, R., Kapadia, A., and Lee, A. J. Reflection or action?: How feedback and control affect location sharing decisions. In *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems (CHI '14)* (April/May 2014), 101–110.

[18] Schlegel, R., Kapadia, A., and Lee, A. J. Eyeing your exposure: Quantifying and controlling information sharing for improved privacy. In *Proceedings of the 2011 Symposium on Usable Privacy and Security (SOUPS)* (July 2011).

[19] Wang, Y., Komanduri, S., Leon, P., Norcie, G., Acquisti, A., and Cranor, L. "I regretted the minute I pressed share": A qualitative study of regrets on facebook. In *Symposium on Usable Privacy and Security* (2011).