Respecting User Privacy in Cross-System Personalization

Yang Wang

Cylab Usable Privacy and Security (CUPS) Lab
Carnegie Mellon University
wang@cs.cmu.edu

Abstract

This paper discusses one area of our research interests that fit with the scope of the workshop. Specifically, our domain of interest discussed here is cross-system personalization (CSP), an innovative technology that enables consistent personalized user experience across different applications, platforms and even devices. Despite the potential benefits to both service providers and end users, CSP raises thorny privacy issues. This paper discusses some of these potential privacy issues in CSP, argues for the importance of user control in the data management, and suggests directions for future research.

Introduction

Cross-system personalization (CSP) refers to "personalization that shares information across different systems in a user-centric way" [1]. In a converged service environment, CSP enables services or applications that adapt to each user based on the user's service consumption data from multiple service domains (e.g., music and news) and multiple service platforms (e.g., IPTV and mobile phone) [2]. Imagine the personalized radio (e.g., Pandora) on your smart phone playing music that is (partially) based on what news and shows you watched on your IPTV, and/or the Youtube videos you saw on your laptop. CSP has the potential to strengthen the benefits of personalization: further engage and retain end users, help select targeted ads, etc. However, since CSP usually relies on collecting, merging and mining user data gleaned from multiple applications/platforms, it is subject to legal privacy requirements and evokes privacy concerns in end users.

Legal Requirements

Privacy laws and regulations usually lay out both organizational and technical requirements for information systems that store and/or process personal data, in order to ensure the protection of these data. Those requirements prescribe, e.g., proper data acquisition, retention, transfer, and processing [3]. Our earlier work involved a general analysis of impacts of various European Union directives[1] and privacy laws on personalization [4]. Here we discuss several aspects of legal privacy requirements that are particularly relevant to CSP.

Purpose-Specific Data Collection and Usage.

The Czech Republic Privacy Act [5] mandates that:

> *Personal data that were obtained for different purposes may not be grouped.*

The German Telemedia Act [6] requires that:

> *Personal profiles retrievable under pseudonyms shall not be combined with data relating to the bearer of the pseudonym.*

These legal requirements reflect a fundamental privacy principle that underlines many privacy laws, namely, purpose-specific data collection and usage. This principle conflicts with the practice of merging data across multiple sources (and presumably collected under different purposes). Without users' consent (opt-in), one may question the legality of CSP driven by merging and sharing user data across applications.

Parsimonious Data Retention and Processing

Another related privacy principle has to do with data parsimony – only collect and use data to the extent that it is needed. For instance, the German Telemedia Law [6] also requires that:

> *Usage data must be erased immediately after each session except for very limited purposes*[2].

This specification could affect CSP systems that utilize a user's usage data across sessions on the same or on different systems over an extended period of time. This data parsimony imperative may again jeopardize CSP systems that rely on tracking users across sessions and applications.

CSP Deployed across Different Jurisdictions

Many service providers that espouse the idea of CSP operate internationally (e.g., Alcaltel-Lucent) . That is to say their CSP systems are likely to be deployed to different countries and thus need to observe the laws of different

[1] EU member states need to implement the requirements from these EU directives in their national privacy laws.

[2] Examples include fighting fraud and bill tracking.

jurisdictions. A CSP system that operates lawfully in one country may violate the privacy laws of another country. CSP designers need to take this into consideration. Our previous work proposed a software architecture that mitigates this problem in web personalization by individually catering the processing of personal data at a website to the privacy requirements of every single user [3]. We plan to investigate the applicability of this approach in the context of CSP.

End User Privacy Concerns

Teltzrow and Kobsa [7] present a meta-analysis of various studies of Internet users' privacy concerns and their impacts on personalized systems. They conclude that web users are not only quite concerned about being tracked online but also counteract, e.g., by providing false information to websites. This dramatically affects CSP because such systems need to track a user across multiple applications. Unfortunately there is currently little academic knowledge/research of end user's privacy concerns about being tracked across systems. A better understanding of users' privacy concerns in the context of CSP is needed to search for usable solutions.

From our ongoing study, our participants almost universally express distaste of the idea of being tracked across different sites, esp. when it is used for marketing or advertising purposes. They were somewhat less averse to content recommendation based on cross-site tracking.

Future Research Directions

Privacy is not a new research topic for personalization. There are a substantial amount of prior knowledge and many existing techniques that we can build upon. In the area of usable privacy and security [8], researchers have been studying people's privacy concerns and practices in various contexts (e.g., [9]), and developing usable end-user privacy management tools (e.g., [10]). However, to what extent these privacy concerns and tools apply in the context of CSP is still an open question.

In the area of privacy-enhanced personalization [11], most solutions follow either an architectural approach that the personalization system architecture respects certain privacy constraints (e.g., [3]) or an algorithmic approach in which the personalization algorithms manifest some privacy-preserving characteristics (e.g., [12]). There is virtually no work on empowering end users to manage their privacy in personalization. One exception is scrutable personalization [13] in which tools are provided to enable end users to scrutinize the underlying user model and adaptation process, primarily in educational settings.

In the following, we outline a number of promising research directions that we believe may be particularly fruitful for the future. Nearly all of them involve some form of user empowerment.

Privacy Dashboard

One of the hurdles that prevent users from making rational privacy decisions is information asymmetry – users usually do not have sufficient contextual information to make informed decisions. Inspired by the idea of Google Dashboard, we are building a prototype of privacy dashboard for social network services (SNSs). This application functions as a SNS aggregator. Once a user logs into her various SNSs, the application displays all the content and the privacy settings associated with each piece of content on these SNSs. We plan to conduct a user study to test its effect on users' privacy decision making.

Sticky Policy

We also plan to explore technical and usable mechanisms for supporting sticky policy. In the domain of SNSs, we are interested in mechanisms that protect each piece of content based on its sticky privacy policy. We are currently considering emerging techniques and standards such as content-centric networking (CCN) [14] and User managed access (UMA) [15].

Privacy Nudges

Humans exhibit various systematic (thus predictably) cognitive and behavioral biases in our privacy decision making (e.g., the dichotomy between stated privacy preferences and actual privacy decision/behavior) [16]. One exciting area of privacy management is to nudge users' privacy decision making behavior towards their stated preferences or increasing their welfare or decisions they will not later regret [17].

Privacy Sampling

Because of the potentially large number of applications, we do not want to overwhelm our users by asking them for their preferences every time they encounter a new application. One simplification is to "sample privacy" – each user is only asked to provide a small set of privacy decisions initially. The CSP system will (incrementally) build a privacy model for each user that can predict his/her unspecified privacy decisions. Users can of course choose to override these predicted privacy settings as they wish. One case in which this strategy has been applied is an application for sharing location information between friends that yields fairly high (about 90%) prediction accuracy [18].

Visualization of Privacy Settings and Support for Social Navigation

We can create intuitive visualizations of individual users' privacy settings (e.g., [19][20] for privacy policies). We can also explore the idea of social navigation [21] in this context – providing visualizations of other people's (friends and families) or group's privacy settings, and share them with one another [22]. For example, knowing aggregated statistics, such as the percentage of users who chose to disclose a particular piece of service consumption data, may help users make their own decisions [23].

Conclusion

Cross-system personalization has a huge potential of transforming user experience and boosting business, but considerable privacy issues remain to be resolved. In this paper, we highlight some potential privacy issues in CSP, advocate more privacy research in this emerging area, and suggest future directions that can potentially empower end users to better make informed privacy decisions.

References

[1] B. Mehta, C. Niederee, A. Stewart, M. Degemmis, P. Lops, and G. Semeraro, "Ontologically-Enriched Unified User Modeling for Cross-System Personalization," in *User Modeling 2005*, 2005, pp. 119-123.

[2] A. Aghasaryan, S. Betgé-Brezetz, C. Senot, and Y. Toms, "A profiling engine for converged service delivery platforms," *Bell Lab. Tech. J.*, vol. 13, no. 2, pp. 93-103, 2008.

[3] Y. Wang and A. Kobsa, "Respecting Users' Individual Privacy Constraints in Web Personalization," in *UM07, 11th International Conference on User Modeling*, pp. 157–166, 2007.

[4] Y. Wang and A. Kobsa, "Impacts of Privacy Laws and Regulations on Personalized Systems," in *PEP06, CHI06 Workshop on Privacy-Enhanced Personalization*, pp. 44-46, 2006.

[5] CZ, "Act 101 of April 4, 2000 on the Protection of Personal Data and on Amendment to Some Acts," 2000. [Online]. Available: file://Lit1/Czech-PrivacyLaw-2000.pdf.

[6] DE-TS, "German Teleservices Data Protection Act, as amended on 14 Dec. 2001," 1997. [Online]. Available: file://Lit1/germany-iukdg-eng.htm.

[7] M. Teltzrow and A. Kobsa, "Impacts of User Privacy Preferences on Personalized Systems - a Comparative Study," pp. 315--332, 2003.

[8] L. F. Cranor, "Towards usable Web privacy and security," in *Proceedings of the 14th international conference on World Wide Web*, pp. 352-352, 2005.

[9] N. Sadeh et al., "Understanding and capturing people's privacy policies in a mobile social networking application," *Personal Ubiquitous Comput.*, vol. 13, no. 6, pp. 401-412, 2009.

[10] C. Brodie, C. Karat, J. Karat, and J. Feng, "Usable security and privacy: a case study of developing privacy management tools," in *Proceedings of the 2005 symposium on Usable privacy and security*, pp. 35-43, 2005.

[11] A. Kobsa, "Privacy-enhanced personalization," *Commun. ACM*, vol. 50, no. 8, pp. 24-33, 2007.

[12] B. Mehta, "Learning from What Others Know: Privacy Preserving Cross System Personalization," in *Proceedings of the 11th international conference on User Modeling*, pp. 57-66, 2007.

[13] M. Czarkowski and J. Kay, "A Scrutable Adaptive Hypertext," in *Proceedings of the Second International Conference on Adaptive Hypermedia and Adaptive Web-Based Systems*, pp. 384-387, 2002.

[14] V. Jacobson, D. K. Smetters, J. D. Thornton, M. F. Plass, N. H. Briggs, and R. L. Braynard, "Networking named content," in *Proceedings of the 5th international conference on Emerging networking experiments and technologies*, pp. 1-12, 2009.

[15] Kantara Initiative, "User managed access (UMA)." [Online]. Available: http://kantarainitiative.org/confluence/display/uma/Home.

[16] R. Acquisti and J. Grossklags, "What Can Behavioral Economics Teach Us About Privacy?."

[17] A. Acquisti, "Nudging Privacy: The Behavioral Economics of Personal Information," *IEEE Security and Privacy*, vol. 7, no. 6, pp. 82-85, 2009.

[18] P. G. Kelley, P. H. Drielsma, N. Sadeh, and L. F. Cranor, "User-controllable learning of security and privacy policies," in *Proceedings of the 1st ACM workshop on Workshop on AISec*, pp. 11-18, 2008.

[19] R. W. Reeder, P. G. Kelley, A. M. McDonald, and L. F. Cranor, "A user study of the expandable grid applied to P3P privacy policy visualization," in *Proceedings of the 5th Symposium on Usable Privacy and Security*, pp. 1-1, 2009.

[20] J. Kolter and G. Pernul, "Generating User-Understandable Privacy Preferences," in *Availability, Reliability and Security, International Conference on*, pp. 299-306, 2009.

[21] A. Dieberger, P. Dourish, K. Höök, P. Resnick, and A. Wexelblat, "Social navigation: techniques for building more usable systems," *interactions*, vol. 7, no. 6, pp. 36-45, 2000.

[22] J. Kolter, T. Kernchen, and G. Pernul, "Collaborative Privacy – A Community-Based Privacy Infrastructure," in *Emerging Challenges for Security, Privacy and Trust*, 2009, pp. 226-236.

[23] Sameer Patil and Alfred Kobsa, "Enhancing Privacy Management Support Instant Messaging," *Interacting with Computers*, Forthcoming. .