Managing the Privacy of Incidental Information During Collaboration

Kirstie Hawkey and Kori Inkpen
Faculty of Computer Science
Dalhousie University

Abstract

Privacy issues related to the viewing of incidental information can impact people's willingness to collaborate on an ad-hoc basis in a co-located setting. Management of incidental information is a complex problem due to multiple viewing contexts, individual differences, and the large volume of information involved. Solutions must balance the amount of control given to the user with the effort required to maintain the system. Our exploratory research on privacy issues for incidental information visible in web browsers has found patterns that may provide a basis for semi-automating privacy management. We discuss ongoing challenges including reducing complexity for users, enhancing visualization techniques for individual pattern analysis, and evaluating privacy management solutions.

Key words: Privacy, co-located collaboration, web browsers, field study.

1 Introduction

Colleagues often gather in an ad hoc basis around a computer to collaborate on a project. However, a great deal of incidental information about past activities on the computer is then visible with casual inspection. This information may be inappropriate for the current viewing context. The normative privacy [5] usual for personal displays does not apply during co-located collaboration; the display is an object in the collaboration.

Currently, users must make tradeoffs to manage their privacy: they can either work efficiently in a familiar environment, with access to convenience features and usual layout, or work awkwardly in a sterile environment. The growing prevalence of ad hoc co-located collaboration on mobile devices, used in a variety of contexts and settings, makes incidental viewing of information a compelling problem. The intersection of privacy management [1, 5] and personal information management [2] results in a hard problem due to the complexity and volume of information and individual differences in behaviour.

2 Research Goals

Our overall goal is to provide users with tools to manage incidental information privacy, only revealing information appropriate for the current context. We began by examining privacy issues related to the incidental information found in web browsers. Browsers have many convenience features that assist browsing but display traces of prior activity that users may prefer to remain private (e.g. AutoComplete reveals search terms and URLs). We will draw upon the knowledge gained from research in this focused area to provide guidelines for privacy management of incidental information in the general computing environment.

3 Exploratory Research

3.1 Field Study

We conducted a week-long field study [3] to explore the issue of privacy gradients and whether or not patterns of privacy exist during web browsing. We examined a 4-tier privacy gradient scheme (Fig. 1): *public, semi-public, private,* and *don't save.* Public sites are those appropriate for anyone to view (even the Queen as indicated by the crown in Fig. 1), semi-public may not be appropriate in some viewing contexts, private are only suitable for close confidants to view, and don't save includes sites that nobody, including the user, needs to see again. We recorded all web browsing conducted on the laptops of 20 participants. We developed a browser helper object to log the actual sites visited by users (URL, page title, time stamp, browser window ID). Participants used an electronic diary daily to classify the privacy level of each web page they had viewed. To maintain participant privacy, we removed the URL and page title from the diary output.

Results revealed that the privacy of incidental information during web browsing is indeed an issue:

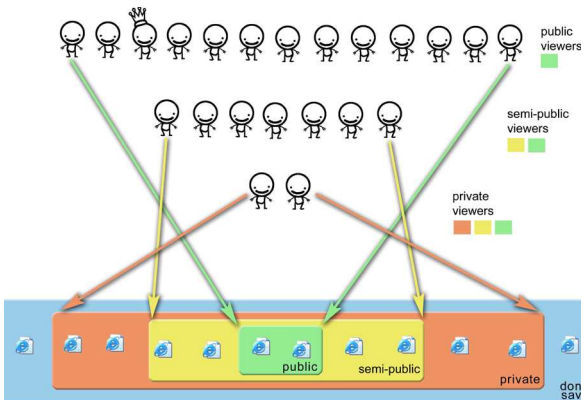

Figure 1. Privacy gradients used during field study

given advanced warning, 95% of participants indicated they would take some action to limit visibility of information. Web browsing behaviours varied considerably (e.g. privacy gradient usage and number of pages visited and windows in use); however, trends emerged in the perceived privacy of incidental information. Participants clustered into four groups depending on their overall gradient use. Patterns emerged on a per window basis with most participants having streaks of browsing at a privacy level and few transitions between levels.

A second phase of this field study has just begun with 15 additional participants: 5 non-technical laptop users, 5 non-technical desktop users, and 5 technical desktop users. The purpose of this phase is to include more non-technical users and to gain contextual information about the location of browsing and the pages viewed. We will not automatically sanitize the diary entries, but will allow blinding of sensitive entries.

3.2 Survey

An on-line survey of 155 individuals has just completed. We are exploring whether different privacy levels are required depending on the primary context of browsing: the setting of web browsing activity (home or away) and the type of computer used (laptop or single user computer). We will also examine the impact of the context of subsequent viewing and device mobility.

4 Ongoing Challenges

Privacy management of incidental information is complex due to the multiple contexts of its creation and viewing. To protect privacy within the normal computing environment, we must balance the amount of control a user has over the environment with the time and effort necessary to provide that control. Our exploratory studies are a first step in meeting this challenge as we discover patterns that may reduce the complexity and burden of managing privacy.

We have found analyses of study data difficult due to the sheer magnitude of sites visited [4] and individual differences. Numerical averages across users do not allow us to view the individual patterns at play. We will use data mining to look for more patterns, particularly those of a temporal nature. A visualization technique may also be necessary to gain a richer understanding of patterns uncovered and their applicability to individual and general solutions.

Evaluation of privacy needs and solutions must occur within a natural setting. However, unless we maintain privacy of participants' computing activities, they may not engage in the normal behaviour we seek to observe. During the first phase of our field study, we safeguarded the privacy of sites visited and only viewed data relating to gradient use on a per-window and temporal basis. We learned more general privacy perspectives through classification tasks. For the second phase, we would like to learn more about the context, but participants can safeguard the identity of particularly sensitive sites. Longitudinal evaluation will be necessary to determine if a privacy management system is effective and maintainable over time. If we occlude information to maintain privacy during evaluation, it will be difficult to validate whether the scheme is appropriate and effective. If we require participants to provide fine-grained effectiveness ratings over time, evaluation may be overly burdensome thus impacting system use.

5 Future Work

Data analyses from the field studies and the survey will soon be complete. Outcomes from the studies and the resolution of the identified challenges will guide the development and evaluation of a privacy management system for the incidental information generated during web browsing. It is clear we must utilize inherent patterns to relieve the burden of the user classifying all incidental information. Given the per window patterns of privacy streaks with minimal transitions revealed thus far, one approach may be to utilize browser windows of different privacy levels. These could not only filter which incidental information is displayed, but also classify new incidental information as it is generated.

Acknowledgements

Thanks to the members of the EDGE Lab. Funding provided by NSERC and Dalhousie University.

References

[1] M. Ackerman, L. Cranor, and J. Reagle, "Privacy in E-Commerce: Examining User Scenarios and Privacy Preferences," In ACM Conference on Electronic Commerce, pp. 1-8, 1999.

[2] R. Boardman and M. A. Sasse, ""Stuff Goes Into the Computer and Doesn't Come Out": A Cross-tool Study of Personal Information Management," In Proc. of CHI, pp. 583-590, 2004.

[3] K. Hawkey and K. Inkpen, "Privacy Gradients: Exploring ways to manage incidental information during co-located collaboration," In Extended abstracts of CHI, 2005.

[4] K. Hawkey and K. Inkpen, "Web Browsing Today: The impact of changing contexts on user activity," In Extended Abstracts of CHI, 2005.

[5] J. H. Moor, "Towards a theory of privacy in the information age," *ACM SIGCAS Computers and Society*, vol. 27, pp. 27-32, 1997.