Managing Visual Privacy within the Web Browser

Kirstie Hawkey and Kori M. Inkpen
Faculty of Computer Science, Dalhousie University
6050 University Avenue
Halifax, Nova Scotia, Canada B3H 1W5
{hawkey, inkpen}@cs.dal.ca

ABSTRACT
Traces of previous activity are often visible during collaboration around a personal display. A survey investigated privacy concerns associated with the traces of web browsing activity that appear in web browser convenience features. Two field studies examined patterns that occur when participants applied privacy levels to their visited pages. Results from these studies guide development of a privacy management system. This poster will present our initial design ideas to help users control the traces of their previous activity that may be visible within web browsers.

1. INTRODUCTION
Web browsers are used in a variety of contexts, including during collaboration around a personal computer. Web browsers have several convenience features (e.g. History, Auto Complete) that assist with page revisitation by storing traces of activity. However the information visible within these convenience features may be problematic in a group setting as the traces may reveal incidental information (i.e. information unrelated to the task at hand) that is inappropriate for the current viewing context (see Figure 1).

To maintain privacy in situations when web browser windows are visible to others, users must currently choose to either turn the convenience features off or periodically clear the stored information. However, those traces are often valuable for future transactions and their removal may decrease productivity. Commercial products allow users to delete web browsing traces; however the decision to erase a class of traces (e.g. History) generally erases all instances indiscriminately. Furthermore, these tools often assume that the vast majority of items are public in nature with only a small subset needing to be password protected, and that sites of both types are never viewed concurrently.

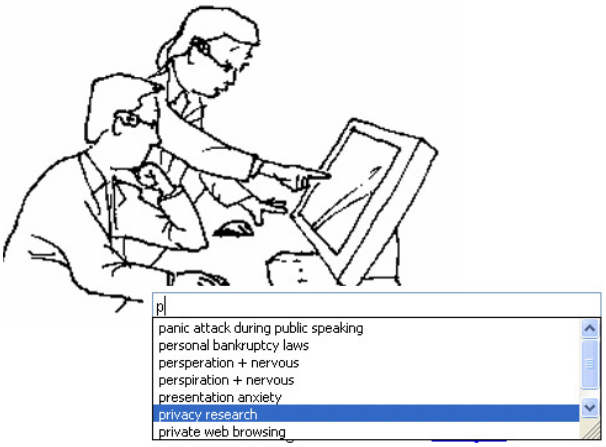

Figure 1. Incidental information privacy example. Previous search terms are revealed to a collaborator when the user begins to type "privacy research" in the form.

Little research has investigated tools for managing inter-personal privacy of information within web browsers. COLLABCLIO was developed to support automated sharing of web browsing histories [7]. It provided users with a binary classification scheme (public/private) to indicate which visited URLs should be shared. Berry et al. [2] have taken a role-based approach to enable privacy in shared views of applications such as Internet Explorer (IE) and allow protection of objects within documents. For example, in the public view of an IE window, the Auto Complete options for URLs can be masked, while the presenter retains full functionality of this feature in the private view.

Design principles are emerging for privacy management systems. Lau et al. [7] state that privacy interfaces should make it easy to create, inspect, modify, and monitor privacy policies and that the policies should be applied proactively to objects as they are encountered. De Paula et al. [3] discuss three design principles for enhancing the usability of systems with a security and privacy component: visualization mechanisms, event-based architecture, and integration of configuration and action. These principles are intended to create conditions whereby users can not only recognize issues as they arise, but also understand the issues well enough to make informed decisions and take appropriate actions.

We have used a mixed methodology approach to study the privacy of the incidental information found in web browsers. A survey examined participants' self-reported privacy concerns (see [6] for details). Two, week-long field studies examined participants' application of privacy levels (public, semi-public, private, don't save) to their actual web browsing (see [4] for details of study 1; [5] for study 2). We next discuss initial design ideas for a privacy management system based upon our research findings to date.

2. INITIAL DESIGN IDEAS
We begin with two general themes that have arisen through our research: the need for a personalized and nuanced approach. We then discuss the components of a systems approach to incidental information privacy management: classification of new traces of browsing activity, appropriate filtering of traces during collaboration, and on-going privacy maintenance.

2.1 General Themes
Privacy concerns have been known to be a highly variable and an individualized approach has been suggested [1]. Our results confirm the necessity of a *personalized approach* in order to ensure that a privacy management system in this particular domain is effective. Results from our survey revealed variability in overall privacy concerns. During both field studies, we observed variability both in terms of participants' browsing behaviours and the privacy classifications of their visited pages.

Almost all participants in the field studies utilized all privacy categories when classifying their visited web pages. This use of

all four privacy levels validates the need for a more *nuanced approach* than the Public/Private or Save/Don't Save approach currently used in web browser convenience features and privacy management tools. Users of COLLABCLIO also indicated a need for a more nuanced approach than Public/Private [7].

2.2 Systems Approach

There are three main aspects to a systems approach to privacy management: to classify web browsing traces with a specific privacy level, to then filter the information appropriately for the current viewing context, and to provide methods for users to actively maintain the system.

2.2.1 Classification of New Browsing

While a simple approach is to have users classify each trace manually, as evidenced during both our field studies, the rapid bursts of activity and the sheer magnitude of pages visited during web browsing would make this task overly burdensome. A privacy management system will likely need some type of (semi-) automated privacy classification in order to be manageable.

One approach is to *automate content categorization* so that new traces of browsing are categorized as to content and classified with a privacy level. Users would specify which privacy level to apply to each category. A comparative evaluation of participants' theoretical content categorizations and privacy levels applied to actual web browsing suggests that a personalized approach may be feasible; however, further refinement of content categorizations is needed to improve accuracy [5].

Another approach is to capitalize upon patterns inherent during web activity. For example, participants tended to partition their browsing so that private browsing is in a single window [4]. Within a window, most browsing (85% of page visits) occurs within streaks (i.e. 2+ consecutive pages) at a given privacy level and there are relatively few transitions between levels (average of 0.9 per browser window). Given these patterns, one approach may be to allow users to open *browser windows of different privacy levels*. These windows could not only filter what incidental information is displayed, but could also tag new sites visited, similar to the extensional classification described in [7].

One benefit to this approach is that users could specify at the time of initial activity which visited pages should not be saved. During our field studies, participants tended to use the "don't save" category to indicate pages that were either inconsequential or extremely private. Allowing users to stop the recording of their activity for brief periods of time will help users remove some of the most sensitive sites from their convenience features and will also reduce what data is saved. Many participants in our studies indicated a desire for a more fine-grained approach to managing which information is recorded in their convenience features.

2.2.2 Filtering Browsing during Collaboration

Whatever the classification scheme, users must be provided with mechanisms to specify the current context so that only contextually appropriate content is displayed. With *browser windows of different privacy levels*, this would be accomplished simply by opening up a window at an appropriate privacy level so that only appropriate content is display. While some users may find a simple hierarchical scheme appropriate (e.g. public, semi-public, private, don't' save); questionnaire responses during the field studies indicate that other users may require some further partitioning of their activities (e.g. work groups).

Another approach is to have users define the current viewing context. Privacy comfort levels of participants were found to be highly contextual, related to the potential viewers, the level of control, and the sensitivity of the content [6]. Furthermore these results were highly individual. Simplified configuration mechanisms may be possible for those participants not concerned along a particular dimension (e.g. level of control). An open question remains as to whether it is enough to give users pre-defined contexts to quickly toggle between or whether a more dynamic configuration of the current viewing setting is required.

2.2.3 Ongoing Privacy Maintenance

Users will require methods to check the accuracy of the classified traces of web activity and to adjust those privacy levels if necessary. Visualizations will be needed so users can easily view which traces may be revealed during browser use. It may be possible to use a content classification scheme (e.g. categories, keywords, URLs) to flag traces that may be inappropriately classified. Furthermore, many participants in all three studies indicated a desire to selectively delete traces of activity when limiting the information that might be displayed.

3. CONCLUSIONS AND FUTURE WORK

Initial patterns from both field studies look promising as mechanisms for a (semi-) automated approach to privacy management. The data is still being examined for temporal and individual patterns of privacy application that may help guide personalized solutions. Of particular interest is the identification of triggers that precipitate a switch between privacy levels (e.g. content of a page, secure pages). Once analysis is complete, a final set of guidelines will direct design of an incidental information privacy management system. The poster will feature the user interface designs under consideration to allow users to apply privacy classifications to the traces of their web activity and filter the traces appropriately during later collaboration.

4. REFERENCES

[1] Ackerman, M., Cranor, L. and Reagle, J. (1999). Privacy in E-Commerce: Examining User Scenarios and Privacy Preferences. In *Proc. of EC '99*, 1-8.

[2] Berry, L., Bartram, L. and Booth, K. S. (2005). Role-Based Policies to Control Shared Application Views. In *Proc. of UIST*, 23-32.

[3] de Paula, R., Ding, X., Dourish, P., Nies, K., Pillet, B., Redmiles, D., Ren, J., Rode, J. and Filho, R. S. (2005). Two Experiences Designing for Effective Security. In *Proc. of Symposium On Usable Privacy and Security (SOUPS)*, 25-34.

[4] Hawkey, K. and Inkpen, K. (2005). Privacy Gradients: Exploring Ways to Manage Incidental Information During Co-Located Collaboration. Ext. Abstracts CHI 2005, ACM Press: 1431-1434.

[5] Hawkey, K. and Inkpen, K. M. (2006). Examining the Content and Privacy of Web Browsing Incidental Information In *Proc. of WWW 2006*.

[6] Hawkey, K. and Inkpen, K. M. (2006). Keeping up Appearances: Understanding the Dimensions of Incidental Information Privacy. In *Proc. of CHI 2006*, 821-830.

[7] Lau, T., Etzioni, O. and Weld, D. S. (1999). Privacy Interfaces for Information Management. *Communications of the ACM* 42(10): 89-94.