Incidental Information Privacy and PIM

Kirstie Hawkey and Kori M. Inkpen
Dalhousie University
6050 University Ave.
Halifax, Nova Scotia
hawkey@cs.dal.ca

ABSTRACT
Our research investigates the privacy issues regarding information visible on personal computer displays during collaboration. Personal information management systems often generate traces of activity, both when an end user explicitly saves information and through system use. These traces of activity may then be subsequently revealed as the user interacts with the system. The revealed information may not be appropriate for viewing in a collaborative setting. This paper discusses incidental information privacy and its ties with personal information management systems. A summary of our research to date is given along with a discussion of managing the visual privacy of incidental information in web browsers and other PIM systems.

Categories and Subject Descriptors
H.5.3 [**Information Interfaces and Presentation**]: Group and Organization Interfaces – *Collaborative Computing, Web-based Interaction.* H.5.4 [**Information Interfaces and Presentation**]: Hypertext/Hypermedia – *User Issues, Theory*

General Terms
Human Factors, Theory

Keywords
Privacy, incidental information, co-located collaboration, web browsers, personal information management, visual privacy

1. INTRODUCTION
Our research focus is incidental information privacy (IIP). We define incidental information to be the information visible on a display that is incidental to the task at hand (e.g., search terms revealed by Auto Complete that are unrelated to the current search, as in Figure 1). Privacy issues can arise when incidental information is visible on a display that is viewed by others, such as when colleagues gather informally around a display to collaborate on a project, when a personal computer is connected to a large screen display, or when a computer is used sequentially by those without separate accounts. This information may or may not be appropriate for the current viewing context. Privacy concerns occur when the information that is visible does not fit the persona a user is trying to maintain (e.g., evidence of visits to a bawdy humour site by an employee may be inappropriate for an employer to view on a computer used in the workplace) [6].

IIP is closely tied to personal information management (PIM) systems. Essential PIM activities include keeping information, finding and re-finding information, and maintaining and managing that information (including mappings between information and need) [12]. For incidental information found in web browsing, the focus of our research to date, a visited web page can be considered the information item. If we want to revisit that page, we have an information need. The mapping between information and need can be largely internal (e.g., our memories) and may have an external representation (e.g., Favorites, Auto Complete, History), part of which can be observed and manipulated (e.g., choice of Favorites name). Some mappings are only potential and not explicit (i.e. a search function is a potential mapping until a specific search is conducted).

Incidental information can be generated both through explicit user action (e.g., when information is saved for the purpose of re-visitation, when files are created) and by the PIM system itself (e.g., text stored for use in Auto Complete functions, accessed documents stored for use in the recent documents list). This information may later be displayed either statically by the system for the purpose of initiating user interactions (e.g., icons on the desktop, recent documents list) or dynamically in response to user interactions with applications (e.g., when entering a search term, Auto Complete shows other recently entered terms).

Many systems include advanced features to improve recognition of desired information for the end user [13]. These features can be a privacy concern as they increase the visibility of incidental information making it easier for others to see traces of previous activities with casual inspection. Examples include visualizations, such as thumbnails of web pages in history files [13], or an

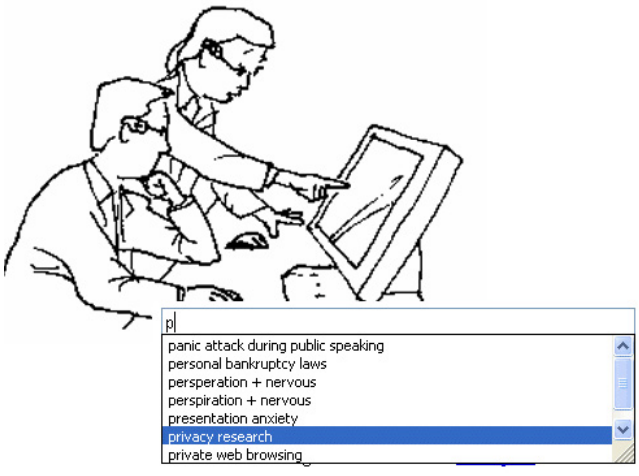

Figure 1. Incidental information privacy example. Previous search terms are revealed to a collaborator when the user begins to type "privacy research" in the form.

expanded and perhaps annotated search result (as in [3] which includes snippets of text from the retrieved information and additional annotations such as when the information was last accessed and tags applied to it).

The use of search as a method of re-finding information also introduces IIP concerns. Search often makes it easier for users to find information as there is no need to remember precisely how the information was generated or saved. However, search can make it more difficult for users to know precisely what information will appear (as opposed to when navigating through a user defined hierarchy). This problem can be exacerbated in PIM systems that incorporate results across tasks or applications. For example, if email is included in the searched documents, personal emails about difficulties working with another person on a project may be inappropriately revealed when searching for information about the project. One example is Stuff I've Seen [5] which provides a single index for all information that a person has viewed on their computer, regardless of the information type (e.g., email, URL), and then provides rich contextual cues during the search process including thumbnails, time, and author.

The intersection of privacy management and personal information management results in a challenging problem due to the complexity and volume of information [9]. Before developing privacy management solutions for incidental information, we must fully understand the dimensions of the problem. We next provide a brief overview of related work, followed by a summary of our research to date. We then discuss IIP management in web browsers and other PIM systems.

2. RELATED WORK

There has been little prior research investigating the privacy of incidental information. COLLABCLIO supported automated sharing of web browsing histories between colleagues by allowing Public/Private ratings for their visited pages [14]. Berry et al. [2] have taken a role-based approach to enable privacy in shared views of applications such as Internet Explorer (IE) and Word, allowing protection of objects within documents. An IE window can be opened with both public and private views; in the public view, features such as Auto Complete can be masked. Commercial products allow users to more easily delete traces of activity from their web browsers; although the decision to erase a class of traces (e.g., History) generally erases all instances indiscriminately. However, those traces are often valuable for future transactions and their removal may decrease productivity. Furthermore, commercial tools often assume the majority of items are public in nature with only a small subset needing to be password protected, and that both types are never viewed concurrently.

Palen and Dourish [16] describe three interrelated boundaries for privacy management: disclosure, temporal, and identity. These boundaries between what is considered public or private are continuously refined depending on the context. This model of privacy fits IIP well. Users would like to be able to control an appropriate level of content sensitivity given the context of viewing (disclosure boundary). The persistence of traces of previous activity (temporal boundary) makes it difficult for users to ensure that they are presenting themselves appropriately for their current role (identity boundary).

Although research in the domain of IIP is just beginning, previous research in other privacy domains has found that privacy concerns are highly nuanced and individual [1,15]. Lederer et al. [15] discuss how activities convey the essence of a persona. Knowledge of activities is more sensitive when identity is known as the activities can reveal hidden personae. With traces of incidental information, a person's actions in one context (e.g., personal browsing conducted at lunch) may later be viewed in another context (e.g., when collaborating with an employer). Information that is appropriate for a friend to see may not be appropriate if viewed by an acquaintance or an authority figure with whom one would prefer to present a more formal or otherwise restricted face [6].

Privacy management of incidental information can be difficult for computer users. It is not always clear exactly which traces of activities are being created and stored and which can subsequently be viewed by others during normal computer usage [17]. Nor is it clear whom all the future viewers will be and the context under which material will be viewed, particularly when devices are mobile and used in both personal and business settings [16].

PIM research also gives insight into the requirements for a privacy management system. For example, Gwizdka [7] studied email task management strategies and found that participants clustered into two groups: Cleaners and Keepers. It will be important that any privacy management system designed for incidental information be suitable not only for those willing to constantly maintain it, but also those who will be more sporadic in their efforts.

Design principles have emerged for privacy management systems. Lau et al. [14] 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. [4] 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.

3. SUMMARY OF OUR RESEARCH

Currently, users must make tradeoffs to manage the privacy of incidental information: they can either choose to work efficiently in a familiar environment, with access to their usual PIM features and screen layout, or work awkwardly in a sterile environment. Our overall goal is to provide users with tools to manage their IIP, only revealing information appropriate for the current context.

Web browsers were selected as the representative application for this research as they are often used during co-located collaboration to find information or share previously found web sites. In addition, web browsers are used for a wide variety of tasks, both personal and work related. Web browsers have many convenience features, such as History, Auto Complete, and Favorites, that assist users when navigating to previously visited pages, but also display traces of prior activity that users may prefer to remain private. The nature of these traces often leads to their unintentional viewing. For example, Auto Complete will reveal search terms previously entered; during a search for

"privacy research" a previous search for "personal bankruptcy laws" may be revealed (as in Figure 1).

We have used a mixed methodology approach in this research. An on-line survey [11] examined privacy concerns related to the incidental viewing of traces of previous web browsing activity. Topics examined include the scope of the privacy issues in this domain; how browsing behaviours affect the content that may be visible; and the role of content sensitivity, level of control, and viewer on privacy comfort levels. Two field studies [8,10] allowed us to examine how participants felt in terms of privacy about specific instances of visible content (the web pages they had visited that day) and to examine patterns in the application of privacy levels to that content. This detailed information was necessary to explore the feasibility of possible privacy management approaches.

Our research has found that the viewing of incidental information in web browsers was a concern for many participants; not only did they have regular occasions when others could view their displays, most were concerned enough to take some steps to manage the privacy of this information [11]. A *personalized approach* to IIP management is necessary as participants' self-reported privacy concerns varied [11] as did their actual browsing behaviours and privacy classifications of visited pages [8,10]. Furthermore, a more *nuanced approach* than the Public/Private or Save/Don't Save approach currently used in web browsers and privacy management tools is desired [14,8,10].

Through an examination of related work and our research results to date, several dimensions of IIP in web browsers that impact a user's comfort level have been identified [11]. Four dimensions with direct impact on the user's *privacy comfort level* in a given situation include their *inherent privacy concerns*, their *level of control* over input devices, their *relationship to the viewer* of the display, and the *sensitivity of potentially visible content*. Furthermore, visible content depends upon recent *browsing activity*, *browser settings*, and any *preventative actions* taken. Browsing activity itself may vary depending on the *location* of the activity and the *type of computer*. These dimensions are often inter-related; for example, advance knowledge of a specific viewer may trigger preventative actions to limit what is visible. As described, the dimensions are specific to traces of web browsing activity. However, while the nature of the visible content will change for incidental information generated and viewed within other PIM systems, the impact of level of control, viewer, and inherent privacy concerns will likely be consistent.

4. MANAGING IIP IN WEB BROWSERS

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.

4.1 Classification of New Browsing

While a simple approach is to have users classify each trace manually, as evidenced during 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.

An approach under evaluation is *automated content categorization* whereby 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 categorizations is needed to improve accuracy [10].

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 [8]. Within windows, most browsing (85% of page visits) occurs within streaks (i.e. 2+ consecutive pages) at a given privacy level and that 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 [14].

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 browser from recording 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 the volume of irrelevant data 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.

4.2 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 study 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 [11]. 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.

4.3 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 study participants indicated a desire to selectively delete traces of activity when managing the information that might be displayed.

5. MANAGING IIP IN PIM SYSTEMS

While our focus has been on developing a privacy management system for web browsers, lessons that we have learned may be applicable to IIP issues in other PIM systems. Rather than building privacy management systems to fix the privacy problems that arise from existing applications, it would be better to address privacy concerns during development of the applications.

Tagging is emerging as a useful method for classifying the privacy level of items in the personal information space and for filtering results appropriately. One recent paper about the PIM system Phlat [3] gives an example using a 'personal' tag to organize and filter items. However, many of the test users of Phlat did not use tags and consistent management of tags can be overly burdensome for users. The authors note that tags should be able to be applied during the workflow as information is encountered and also when decisions are being made about saving information items. Our research has identified issues with managing the privacy classification of visited web pages at the time of browsing; but, as discussed in 4.1, automated methods may assist users in their classifications. It may also be possible to automatically associate privacy tags with other tags being applied, such as tags for people, task types, and content types. For example, information tagged as being related to one person may have a different privacy association than for information associated with someone else.

6. CONCLUSION

PIM and IIP are closely tied. It is important for privacy researchers to be cognizant of PIM research and the advances in PIM systems in order to be aware of changes that will impact privacy concerns and the effectiveness of proposed privacy protection mechanisms. It is also important for PIM researchers to be aware of the various types of privacy implications for their research, be it the formal sharing of information that occurs when data is transferred or the more casual viewing of incidental information on a shared display. Our experiences researching IIP in web browsers give some guidance to appropriate mechanisms for managing IIP in other PIM systems.

7. ACKNOWLEDGMENTS

Thanks to NSERC, NECTAR, and Dalhousie University for funding support and to the members of the EDGE Lab for their feedback and assistance.

8. 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] Cutrell, E., Robbins, D. C., Dumais, S. T. and Saran, R. (2006). Fast, Flexible Filtering with Phlat - Personal Search and Organization Made Easy. In *Proc. of CHI 2005*, 261-270.

[4] 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.

[5] Dumais, S., Cutrell, E., Cadiz, J., Jancke, G., Sarin, R. and Robbins, D. (2003). Stuff I've Seen: A System for Personal Information Retrieval and Re-Use. In *Proc. of SIGIR 2003*, 72-79.

[6] Goffman, E. (1959). The Presentation of Self in Everyday Life. Garden City, New York, Doubleday Anchor Books.

[7] Gwizdka, J. (2004). Email Task Management Styles: The Cleaners and the Keepers. In *Proc. of CHI 2004*, 1235-1238.

[8] 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.

[9] Hawkey, K. and Inkpen, K. (2005). Web Browsing Today: The Impact of Changing Contexts on User Activity. Ext. Abstracts CHI 2005. Portland, Oregon, ACM Press: 1443-1446.

[10] Hawkey, K. and Inkpen, K. M. (2006). Examining the Content and Privacy of Web Browsing Incidental Information In *Proc. of WWW 2006,* 123-132.

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

[12] Jones, W. and Bruce, H. (2005). A Report on the Nsf-Sponsored Workshop on Personal Information Management, Seattle, Wa. http://pim.ischool.washington.edu/final%20PIM%20report.pdf

[13] Kaasten, S., Greenberg, S. and Edwards, C. (2002). How People Recognize Previously Seen Www Pages from Titles, Urls and Thumbnails. In *Proc. of Human Computer Interaction 2002*, 247-265.

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

[15] Lederer, S., Mankoff, J. and Dey, A. K. (2003). Towards a Deconstruction of the Privacy Space. Workshop on Ubicomp Communities: Privacy as Boundary Negotiation, UBICOMP 2003, http://guir.berkeley.edu/pubs/ubicomp2003/privacyworkshop/papers/lederer-privacyspace.pdf

[16] Palen, L. and Dourish, P. (2003). Unpacking "Privacy" For a Networked World. In *Proc. of CHI '03*, 129-136.

[17] Weisband, S. P. and Reinig, B. A. (1995). Managing User Perceptions of Email Privacy. *Communications of the ACM* 38(12): 40-47.