Ad-hoc Guesting: when exceptions *are* the rule

Brinda Dalal[1], Les Nelson[1], Diana Smetters[1], Nathaniel Good[1], Ame Elliot[2]

[1]*Palo Alto Research Center, Inc.*
3333 Coyote Hill Road
Palo Alto, CA 94304
{bdalal, smetters, nathaniel.good}@parc.com
lesnelson@acm.org

[2]*IDEO Palo Alto*
100 Forest Avenue
Palo Alto, CA 94301
elliot@ideo.com

ABSTRACT

People's work days are filled with exceptions to normal routine. These exceptions affect the security and privacy of their information access and sharing. In a recent ethnographic study of ten users in the Bay Area, we identify a number of key problems not well addressed by current data sharing technologies, and from them derive requirements for *Ad-hoc Guesting*, our term for minimal, readily available access control addressing situations not planned for in advance.

1. Introduction

Information sharing and persistent data access is increasingly critical to people's work and personal lives. Yet, corporate security policies rarely comprehend dynamic user models and people's informal and persistent practices around ad-hoc sharing. This paper reports the results of a field study focusing on people's practices around access control, security and file sharing.

Our study sought to understand three areas: under what circumstances do people or companies share or restrict access to files, what tools or behavioral norms are being used to do so, and how are people's experiences, problems and needs changing in regard to secure file sharing and access control, especially in the case of geographically dispersed colleagues, clients, friends and family members?

We identified a number of key problems users face in sharing data:

- *Sharing with myself:* users are their own most common sharing partner, effortfully moving data between their own machines, accounts and devices in order to ensure continued access.
- *Transient data:* users often need to hold data only briefly while transporting it from one place or another; and that data may linger, be lost and forgotten.
- *Transient access:* users need to access data for only short periods of time – they intend only one-time access, or to make data available in certain situations.
- *Over sharing:* users grant more access than necessary when it is difficult to limit who has access to content or how much to share with others, or when pressed for time to extract information from larger data sets
- *Ad-hoc sharing*: users often share content with groups of recipients they have not shared with before, and may not again.
- *Impedance matching:* users spend considerable time and effort tailoring content for sharing based on their understanding of recipient needs or the demands of the sharing mechanisms in use.

Based on these insights, we propose that the general nature of the problem faced by users is what we term *ad-hoc guesting:* where users need to share data securely with unplanned sets of people with whom they have not previously shared who may belong to another organization, thus cannot be "named" by traditional access control. These interactions are transitory and lightweight, often not worth the effort required to set up new sharing mechanisms or change administrative state.

2. Background and related work

Our research builds upon a growing body of literature on file-sharing and access control. Previous studies have focused on personal file sharing, specifically, in the domains of music [3,4] or photographs [1,2], or professional collaborations in corporations [6], where email is viewed as the preponderant medium for file sharing [6,7].

Ahern et al.[1] investigated sharing preferences for personal photos over a mobile phone photo sharing network, and discovered that access control mechanisms were too coarse for many users' needs. They found that end users often overloaded access control mechanisms in order to get around usability issues, such as making all their photos public to make it easier to share photos with friends and family.

Whalen et al [7] surveyed corporate users on document and access control practices, and found that end users had complex policy needs that changed over time and were inadequately addressed by current file sharing and access control mechanisms. Voida et. al [6] created a taxonomy of current file sharing technologies and their attributes. They also found that policies played an important role in users'

privacy decisions, and that current tools were inadequate in meeting people's complex requirements; a finding supported by [8]. Both of these studies considered subjects within a single organization, all of whom had access to similar, established file sharing mechanisms (*e.g.* file servers, *etc.*). They did not, however, consider the effect that these preexisting options had on the challenges users would face or the choices they would make when sharing across organizational boundaries, or operating in the absence of pre-existing shared infrastructure.

Email is routinely chosen as the preferred means of sharing files even in the presence of other alternatives [6,7]. The affordances of file-sharing through email are well-known [6,7]: for instance, you can easily share multiple files with multiple people simply by specifying recipients' addresses. People try to avoid mis-sharing by using the subject line of emails as a marker, or making explicit their requests about the content of documents or what actions they expect from recipients. However, in many instances, email servers and clients typically enforce email size limitations or limit the types of files that can be sent. Moreover, the file is only accessible to the people who received the email: if others wish to access the file, one of the recipients must find and send on the original file. Email also does not provide a means to revoke access to a file, or allow you to view how the file is being accessed. In addition, email is difficult to use for personal storage and versioning tasks. Multiple copies of similar documents can crowd an inbox, and make searching for changes between documents more difficult. Thus file sharing through email remains complex.

What this study adds to the discussion is an exploration of what properties users need from content sharing mechanisms largely in the absence of any *a priori* shared infrastructure. Based on interviews of users across various domains, we were able to explore access control and sharing issues across different types of organizations, such as those with stricter or more lax regulations and compliance policies. We examined in some depth how file-sharing and access controls were used, not used or circumvented in order to get work done. From this analysis we identified key challenges faced by those using and choosing among current file sharing technologies, including email, and derived a set of design criteria we would expect a more effective "sharing system" should meet.

3. Methods

We conducted ten 2-hour, in-depth interviews with respondents in their homes, home offices, or in cafes where people worked, in the spring of 2007. Interviews consisted of semi-structured and open-ended questions about file sharing practices and people's perceptions around access control. Participants were also asked to evaluate paper interface mock-ups relating to security and privacy (results from that research have been addressed elsewhere and will not be covered in this paper).

Participants were recruited through an online site, and pre-screened using a survey followed by a telephone call. Recruiting criteria included those who a). used laptops and desktop computers, b). used two or more mobile devices (such as a cell phone and personal digital assistant), c). worked with colleagues who were not co-located, and d). traveled frequently or had been on an overseas business trip within the last six months. We selected participants with file sharing and access control challenges, such as having to work with multiple clients from different organizations, or share data with geographically dispersed teams, or those who needed access to confidential data. Altogether, we interviewed six men and four women between 23-53 years of age, who worked in finance, health care, travel and tourism, design, engineering (civil, electrical, and software), and product management. The size of their companies ranged from 3 to 150,000 employees. Each interview was conducted by researcher pairs covering backgrounds in anthropology, design, security, and computer engineering. Respondents were asked to describe examples of their professional *and* personal practices around security, privacy and file sharing. We encouraged them to add to a preliminary list of devices, file types, content, social software applications, locations and other variables, and to select any items that triggered thoughts about their practices. Each interview was recorded and transcribed, then analyzed using a grounded theory approach [5]. Data were clustered into emergent categories and cross-cutting themes. Common themes, issues, dilemmas, and trade-offs that people made between levels of security and their ability to complete tasks were identified. Findings were discussed through three design sessions, and used to generate technology requirements and proposed solutions.

4. Summary of findings

Our study highlights distinctions between personal and professional sharing; we identified a variety of infrastructure and devices used, the types of content that people shared, the ways in which items were shared or accessed, and where people were located at the time that they shared documents with each other. 80% of respondents shared files with overseas collaborators or clients in Europe and the Asia-Pacific region, and 100% exchanged data with colleagues across the US. When working from home, consultants and employees in mid-large corporations often shared files through distributed corporate servers, and in three cases, on protected FTP sites. Predominantly, the data shared in professional settings revolved around project work: shared documents included technical specifications, meeting minutes and action items, proposals, reports and in one case, an analysis of soil samples. One of the primary affordances of using a shared server within a company was the ability to reuse documents from one project to another. Frequently, people described how they incorporated sections of an old proposal or template into a new

document, "[I] see if I can borrow text from it and pull information into what I am doing." At the same time, people found it time-consuming to browse different versions of documents to find the proposal they wished to reuse and resorted instead to telephoning or emailing their colleagues to obtain the appropriate copy. As one person explained, "Emails end up being the simplest way to do it, rather than my looking around on their server".

In contrast, people's personal file sharing practices focused on ways in which *experiences* could be shared with others, The content being shared in this case—primarily multi-media—was relational in nature, such as sharing photographs of events with family members who live overseas. We also found a surprising number of people shared the same personal account. For instance, relatives scattered across the US used a photo sharing account that had a single login and password to ensure privacy. Another set of parents set up a "family email account" and used email messages within the same account to discuss homework with their children in the evening.

All respondents used email to share files. 90% of subjects mentioned that they had multiple email accounts (largely personal accounts) and 80% said that they used personal email accounts for business.

80% of respondents, regardless of their demographics, also used a wide variety of social software, including wikis, blogs, social networking sites (including MySpace and Facebook), hosted services (such as Yahoo! Briefcase), public websites for sharing images and multi-media files (including Flickr, YouTube), and online forums and games.

4.1. Sharing with Myself

Respondents clarified two distinctions in file sharing; *sharing with self*, and *sharing with others*. File sharing with oneself serves an important function, allowing people to synchronize their activities regardless of location (work, traveling, or home), accessibility (i.e., whether people can access corporate servers while traveling), or what devices are at hand (laptop at home, USB drives or hosted services).

Sharing with oneself addresses the need to maintain persistent access, regardless of the technical or security constraints in one's environment. For example, interviewees who did not have a printer at home, often uploaded files to Yahoo! briefcase then downloaded and printed files out at their office. 80% of the respondents used USB drives (rather than laptops), to download content at client sites, especially when policies required that they contact IT administrators before accessing electronic files.

Email is a convenient and preferred mechanism for sharing files with oneself, especially for shorter term tasks. Respondents who programmed at home in the evenings described how they preferred to email snippets of code from work to their personal email accounts instead of using CVS directories on corporate servers (which involved lengthy login procedures), or when they wished to avoid having corporate IT install security policies on their personal laptops.

Most respondents had multiple email accounts (some up to 12 or 15) and used these accounts as a data management device. Different types of content were filtered into different accounts - work, friends, dating services, rental businesses, family photographs or spam. However, professional and personal accounts bled into one another, opening avenues for significant security lapses. When email or corporate servers were inaccessible, people readily sent files to consultants using their personal email accounts. While this served a short term need, people said they later ran into trouble trying to track source documents and different versions across their accounts.

4.2. Managing Transient Data

Users frequently handle what we are calling "transient data", or data useful for a single instance or for a task conducted in short order. Transient data are often placed in transient locations or on devices that serve people's short term needs– such as Yahoo! Briefcase, USB flash drives, FTP sites, or in emails to oneself or others. The "throwaway nature" of temporary storage and devices has constraints. For example, one individual remarked that she had a shoebox full of USB drives. Other respondents reported having anywhere between 2-15 active or inactive flash drives stored in their cars, briefcases, at work, or at home. "Fobiquitous tracking" is problematic for many: where does the information reside on my growing number of USB storage fobs? Is this fob the most recent one? Sensitive data might languish, unremembered, on such fobs forever.

Users dealt with such "throwaway data" at different levels of granularity up to and including entire accounts or identities. Respondents increasingly lacked the time to manage their accounts, and tended to shed rather than sort, delete or destroy private data. One individual said that she simply discarded old web accounts and opened new ones.

4.3. Transient Access

A number of individuals noted a need for transient access to data. Consultants, for example, were only supposed to have access to client data during the period of their contracts, or while working in a certain environment.

It can be difficult to go back and "fix" unwanted lingering access, as with another respondent: "But that pretty much is just a few phone calls, desperate phone calls saying 'Delete from your servers, delete from your company, make sure it's completely clean.' You're at the mercy of hoping they follow your request."

4.4. Oversharing

Oversharing occurs in situations where people share too much, or share inappropriate information with others or

themselves. For example, privacy policies are exacerbated for contractors who have limited access to corporate databases, "I have no permissions to get into anything. Other subcontractors are in the same boat, which shouldn't be a problem except that people forget it, so there's a lot of assigning out...of staff, 'Can you make sure you send her a disk?' So [the staff member sends me] files that I don't even really need...."

Time compounds the issue, and results in oversharing with oneself. A healthcare consultant noted that when she visits a client site, she lacks sufficient time to go through the client database in order to extract the data she needs, thus ends up downloading entire files (including social security numbers) onto USB drives. She remarked, "There are a lot of rules trying to get permission from state agencies [to access confidential data]. A lot of data really is protected, so a lot of times the only effective way for me to do the work really disturbs me. Like I can't get permissions, but I can dump huge amounts of data on flash drives that I can then [in theory] lose."

4.5. Ad-Hoc Sharing

Our research found an increasing trend for companies to delegate access control to other companies. Rather than set up extranet sites for consultants or provide them with logins and passwords, companies now expect their consultants to provide a secure but provisional electronic sites on which to store interim data or final reports.

4.6. Impedance Matching

People have varying degrees of perceived and actual technical skills required to use systems, and consequently, there is a disparity in the need for sharing. Often those with greater need faced the burden of the extra work to obtain or share files. As one subject reported about a newly installed web-based repository, "I think we have folks with very limited technical comfort. So for that reason I always have to upload my files [to the repo] and then email them around, so it's sort of another step rather than saving a step."

A major concern among respondents was preventing data sharing failures. The majority of our subjects spent time anticipating their own and their recipients' current state, and changed their actions according to their knowledge of or assumptions about state. People spent considerable time reformatting data for others, based on two parameters. First, they anticipated the constraints of their own or a recipient's system (such as capacity or bandwidth), and secondly, anticipated the recipient's socio-technical knowledge regarding their ability to receive data. A software engineer described the reasons why he compressed photographs for his relatives, "A lot of my relatives are not very techie, so I'll just put photographs in an email attachment. I try to compress them so they are small jpeg sizes and then all people have to do is just click [on the images]". Another respondent drew a similar distinction, "when you're trying to share with family or friends the speed of the network really decides whether or not you can share five photos or just one. If you have to upload five photos individually to send them, that's a real drag. So you need to resize all of them so they fit onto a CD. It's just a big hassle."Half of our interviewees expressed frustration in sending or receiving large files. Some specifically mentioned their personal accounts or corporate email could not handle files over 10MB. A design consultant who provides audio-visual material to his clients, was exasperated by the effort it took to reformat content for their clients, "It's absolutely absurd in this networked economy that we can't share [large] files without going into some extreme effort."

These examples exemplify how the act of file sharing induces *impedance matching*. In other words, users are forced to decide between sharing modalities based on whether the sharing mechanisms will work with a particular user (do they have X? are they on Y?) or piece of content (is the file too big?), or what sharing mechanisms work best with that user (can they be counted on to log onto a separate system?). Equally importantly, how well can you gauge the accuracy of your assumptions about another's state, (can they even receive your files)? It is clear that the onus of work currently resides on users rather than the systems they use.

5. Implications for design

Our findings lead us to identify a common class of sharing problems we term *ad-hoc* guesting. Users in our study often shared data with new and unplanned sets of people, often without assurance that they would ever share with that group of people again. We find that they preferentially and almost overwhelmingly turned to email to do so, except when their *impedance matching* processes indicate that email is unlikely to be successful. This is in contrast to the interpretations of Voida et. al. [6] who suggested that email was chosen only as a fallback alternative to other, preferred, forms of sharing. In this section, we define a set of design requirements for the problem ad-hoc guesting.

5.1. Ad hoc Guesting Design Criteria

We analyze sharing interactions in terms of two roles, initiator and responder, and two modes of sharing: sender-initiated sharing, where the user wants to provide content to someone who does not have it; and the less common receiver-initiated sharing, where a user requests content to which they currently have no access.[1]

[1] The determination of who plays the initiator role may be due to social or organizational factors; *e.g.* a consultant delivering results to a client, or as a result of impedance matching, *e.g.* the more technically sophisticated party in an exchange doing the "heavy lifting" of initiating a transfer.

1) ***No impedance matching:*** initiators should no longer need to be cognizant of the limitations of the system or of a specific recipient's system
   a) The system should work for **all types and sizes of data**, within physical limits (*e.g.* sending large files will be slower than sending small ones).
   b) Responders in particular should be required to have no more than **minimal, readily available tools** (*e.g.* email and a web browser).
2) ***Support ad-hoc sharing:*** encourage lightweight sharing interactions between arbitrary, highly dynamic groups.
   a) Use **universal identifiers**, such as email addresses; people should be able to share with anyone, inside or outside of their organization, with equal facility.
   b) **Minimize setup effort** as users will not know upfront whether they will share with a particular group or use a specific mechanism enough times to make the effort worthwhile.
   c) **Require no *a priori* preparation by responders** -- they should not be required to install software, create an account, or register a profile before someone can initiate sharing with them.
3) ***No oversharing:***
   a) **Content shared only with intended recipient**s: it is not accessible to their friends or arbitrary strangers, or to the server on which the content is stored.or its systems administrators.
   b) **Transient access management:** data can be made available for one-time or time-limited access, without requiring the user to go back and make it "unavailable" again.
4) ***Simple and self-contained:***
   a) **Interactions should be lightweight and familiar.**
   b) **One-step sharing:** additional coordination, such as follow-up emails should not be necessary; people should know that content is there *waiting for them*, or that it has been shared successfully.

6. Conclusions

In our small study, we found that exceptions to stricter security policies are increasingly becoming the rule. We have illustrated users' practices, such as mundane breaches and transient data sharing, in order to design pragmatic and lightweight alternatives. We find that people regularly bypass secure access procedures by using public web repositories, personal emails, and USB drives to transfer information (insecurely). Indeed, many situations require temporary access to data in order to complete a job or activity. Individuals spend considerable time anticipating ways in which to ensure that others can access the data they send them. Repeatedly, people are frustrated that systems and security policies prevent them from sharing large files .

We have identified a common problem and an interesting design opportunity for data sharing – that of *ad-hoc guesting*. In future work, we will explore the success of these potential designs based on the requirements proposed in this paper.

References

1. Ahern, S., Eckles, D., Good, N. S., King, S., Naaman, M., and Nair, R. 2007. Over-exposed?: privacy patterns and considerations in online and mobile photo sharing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 357-366.
2. Miller, A.D. and Edwards, W.E. 2007. Give and take: a study of consumer photo-sharing culture and practice. In *CHI '07: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 347–356.
3. Brown, B., Sellen, A. J., and Geelhoed, E. 2001. Music sharing as a computer supported collaborative application. In *Proceedings of the Seventh Conference on European Conference on Computer Supported Cooperative Work*,, 179-198.
4. Voida, A., Grinter, R. E., Ducheneaut, N., Edwards, W. K., and Newman, M. W. 2005. Listening in: practices surrounding iTunes music sharing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 191-200.
5. Glasser, B., Strauss, A., 1967. *The discovery of grounded theory: Strategies for qualitative research*. Aldine Publishing, New York.
6. Voida, S., Edwards, W., Newman, M. W., Grinter, R. E., and Ducheneaut, N. 2006. Share and share alike: exploring the user interface affordances of file sharing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 221-230.
7. Whalen, T., Smetters, D., and Churchill, E. F. 2006. User experiences with sharing and access control. In *CHI '06 Extended Abstracts on Human Factors in Computing Systems*, 1517-1522.
8. Olson, J.S., Grudin, J. and Horvitz, E.. 2005. A study of preferences for sharing and privacy. In *CHI '05 extended abstracts on Human factors in computing systems*, 1985–1988.