Poking at the Cloud: Identifying Factors Behind Selective Cloud Uploading

Amanda Aizuss,
Max Chen,
Galen Harrison,
Sotiri Komissopoulos,
Blase Ur
University of Chicago
{amanda, amandaaizuss,
mchen24, harrisong,
skomissopoulos, blase}
@uchicago.edu

Abstract
Cloud storage services, such as Dropbox and Google Drive, contain many important and personal files from users, providing hackers an attractive target. Some users, however, may believe that they are protected from compromise because they do not keep anything of interest on the cloud. Not uploading sensitive information is a practice, not a single decision, requiring constant attention and negotiation. We present preliminary findings from an interview study about how users decide to upload or not upload files to the cloud.

Author Keywords
cloud storage; cloud; HCI; usability; privacy; security

CCS Concepts
•Security and privacy → **Usability in security and privacy;** *Social aspects of security and privacy;* **•Human-centered computing** → *Empirical studies in HCI;*

Introduction
Most users of cloud storage services do not put all their files on the cloud. In fact, cloud storage users derive a sense of security from this practice. If there are not sensitive files on the cloud, then those files cannot get lost or stolen. However, a user's decision to not upload a file is not just a single decision, but must be renegotiated as file context and usage change. For example, users might consider their W2 to be

private, but if their accountant asks for them to share it, they may use a cloud service anyway.

It seems very likely that the decision to upload or not upload something depends not just on the content of the file, but also on the platform being used, the expected audience for the file (which may be only the user), and the actions the uploader would like the audience to take.

Our research questions then are when do these considerations become salient and how do they interact with users' models of cloud platform security and privacy? Specifically, we would like to know:

1. How do users perceive storage providers as having agency over their files and how likely do they believe it is that a given provider will exercise that agency?
2. When users split storing files between being stored locally and the cloud, what are the major criteria by which the user determines what is kept locally?
3. When users are asked to put private data in the cloud, do they attempt to manage their privacy either via direct measures like encryption, or via indirect methods like platform or account choice?
4. In sharing files, how cognizant are users of the difference between sharing channels, and what are the major factors behind sharing channel choice?

Related Work

Recent work by Ion et al. [3] and Wu [9] has shown that most users of cloud storage services perceive it to be less safe than keeping data locally. Furthermore, these studies have found that users are unconcerned with this risk because they do not perceive the files they place on the cloud to be of value to anyone other than themselves. While both survey studies find that users report this behavior, to our knowledge no prior studies have fully unpacked users' decisions about whether to store particular files in the cloud.

Vertesi et al. took up a version of this question in interviews in which they asked subjects to create "data narratives" for how they manage their personal data [8]. Consistent with Ion et al. and Wu, they found their subjects actively engaged with the decision to share or store on the cloud. They describe users as being concerned primarily with the normative correctness of their actions. That is, users think about whether they are doing the right thing with their data. While subjects discussed the cloud, Vertesi et al. did not ask questions specifically about it, and they do not discuss the decision to store or not store files in the cloud at any length.

One question we aim to answer is to what extent users' decisions about cloud storage are security-driven. Do users make the decision to create, store, or share with security in mind? Kang et al. have found that even lay users have an awareness that many things kept on the internet may be visible to certain parties [4]. Similarly, Rao et al. found that, especially in the context of websites, the expectations of privacy by users may be mismatched with actual practice [6]. Neither Rao et al. nor Kang et al. considered cloud storage specifically within their models. The mismatch in expectations is not just a policy distinction. Clark et al. found many users were surprised by what images they had previously stored in their cloud accounts [1].

Odom et al. investigated the notion of possession as it applied to digital objects, considering in part how online storage complicates that notion [5]. In particular, they found that sharing items online frustrates a material understanding of sharing as, once shared, an item may be "possessed" by another person simultaneously and with similar agency to the original sharer. Harper et al. argue that the file abstraction continues to be an important concept for conceptualizing

user actions [2]. Together, these two studies suggest the decision of how and what to share may be significant in organizing particular sorts of actions.

Sleeper et al. consider how users may select sharing channels to achieve particular aims [7]. While their study includes channels like social networking sites, they do look at Dropbox and Google Drive as sharing channels. In particular, they find that sharing behavior to a large degree depends upon users' initial expectations of the actions they need to take and the interactions they expect to have with their audience.

Methodology

As a pilot, we conducted six semi-structured interviews with users of cloud file storage services to gauge their cloud storage habits and perceptions of privacy in the cloud. The interviews consisted of: (1) general questions about usage and perceptions of cloud services; (2) accessing a cloud service and answering questions about files stored with that service; and (3) a series of hypothetical file-sharing/storing scenarios. We recruited six undergraduate students who reported using at least one cloud storage service.

As a pilot study, our work is limited by the representativeness of the subjects, who were predominantly drawn from a college population and may have different privacy concerns compared to other segments of the population. Due to the preliminary nature of this study, as well as the small sample size, we did not attempt to find statistically significant differences. Instead we identified a set of themes and developed hypotheses for a future study.

Interview Protocol

For each subject, we started out by asking about their use of cloud storage services. We asked them specifically about whether they used Google Drive, Dropbox, or iCloud, we also asked if they used a cloud storage service other than those three. For each provider, we asked about the frequency and nature of their use.

We then asked subjects to access each service they used and to examine the files they saw stored there. We asked the subjects to access the services using their most common method for doing so. Subjects were then asked questions about the contents of their account. We asked subjects about the types of files they saw, how many were shared with other people, and whether they saw anything there which they would be concerned about losing or having unauthorized people access.

Once the subjects had completed these tasks, we asked them about how they would complete a few hypothetical potentially privacy-sensitive tasks. These tasks were the creation and maintenance of a budget spreadsheet, a private diary, travel information including a passport scan, and then storing passwords. For each task, we asked a follow-up question about how the subject would complete some task which involved sharing or multi-device access. We did this because we wanted to understand how the subjects would choose to manage privacy with respect to the cloud.

Limitations

The hypothetical tasks necessarily contains a confounding factor. Asking the subject about how they would share a file may implicitly alter how the subject considers the file's privacy. For example, when we ask about a budget spreadsheet, someone might consider it to be highly private. However, when asked how they would share it, subjects might either change their assessment of their budget's privacy (due to the fact that it is something they want to share) or they might simply believe that the cloud has sufficient privacy guarantees. More elaborate role play could be useful in understanding this particular boundary, as it would permit the researchers to engineer a more specific situation.

Figure 1: Demographic information of the interview subjects

Subject	Gender	Technical background
P1	Male	Yes
P2	Female	No
P3	Female	No
P4	Male	Yes
P5	Female	No
P6	Female	Yes

Another issue with the current protocol is that it asks the subject to look at files through a perspective they are used to seeing ("the most common method for accessing") as well as asking the subjects to self-report. Despite the fact that the subjects are looking at their own files, there is still the possibility that the impressions they report will be based more on their (mistaken) mental image.

Results and Hypotheses

The full table of demographic information can be seen in Figure 1. We considered subjects to have a technical background if they had either taken computer science coursework or reported working with computers.

Users believe that multiple parties can access their accounts, but that those entities have no interest in doing so
Our subjects also reported that they were not concerned about the privacy of their files because they did not believe they had anything of interest in their cloud storage accounts. Some interesting qualifications to this are that P2 and P6 were concerned that someone might look their schoolwork, and P3 was worried about interview notes for a student organization they were a part of being seen. This is consistent with Ion et. al's findings [3].

Dividing files between locally stored files and the cloud is driven by more than just security considerations
While most users discussed keeping certain items locally, the reasons for doing so varied. While P1 chose not to place password and bank account information on Dropbox due to security concerns, P5 only kept things locally if she did not need to share them, and P4 kept app-specific files locally. It is worth noting that in P1's case the privacy concern is confounded by the absence of a need to share.

Privacy management when sharing takes place via channel choice (specifically email)
P2, P3, P4, and P6 all stated that they would use email to share sensitive data. While the aforementioned issue with the role-playing questions make it hard to pull apart whether this is due to a downgrading in privacy concern, this is a compelling hypothesis to test against in further experiments.

Users choose sharing channels on the basis of social or organizational pressures
All subjects reported that they had more than one cloud storage service. In almost all cases, the subjects noted that they used one service frequently, while the others were used only for specific tasks. P2 and P3 reported using a university-hosted file sharing service only for the purposes of their jobs. P3 stated that she preferred Google Drive because "almost everyone has a Gmail or Gmail-compatible account".

Future work should in particular investigate the dynamics of platform choice as it relates to privacy management. Understanding the relationship between organizational, interpersonal, and privacy considerations when sharing or storing files will help us to better design methods for managing privacy when sharing. In particular, we think that by expanding our line of questioning about hypothetical sharing scenarios, or possibly even observing sharing behavior as it happens, we expect to be able to better address this question.

REFERENCES

1. Jason W Clark, Peter Snyder, Damon McCoy, and Chris Kanich. 2015. "I saw images I didn't even know I had": Understanding user perceptions of cloud storage privacy. In *Proc. CHI*.
2. Richard Harper, Eno Thereska, Siân Lindley, Richard Banks, Phil Gosset, William Odom, Gavin Smyth, and Eryn Whitworth. 2013. What is a file?. In *Proc. CSCW*.
3. Iulia Ion, Niharika Sachdeva, Ponnurangam Kumaraguru, and Srdjan Čapkun. 2011. Home is safer than the cloud! Privacy concerns for consumer cloud storage. In *Proc. SOUPS*.
4. Ruogu Kang, Laura Dabbish, Nathaniel Fruchter, and Sara Kiesler. 2015. "My data just goes everywhere": User mental models of the internet and implications for privacy and security. In *Proc. SOUPS*.
5. William Odom, Abigail Sellen, Richard Harper, and Eno Thereska. 2012. Lost in translation: Understanding the possession of digital things in the cloud. In *Proc. CHI*.
6. Ashwini Rao, Florian Schaub, Norman Sadeh, Alessandro Acquisti, and Ruogu Kang. 2016. Expecting the unexpected: Understanding mismatched privacy expectations online. In *Proc. SOUPS*.
7. Manya Sleeper, William Melicher, Hana Habib, Lujo Bauer, Lorrie Faith Cranor, and Michelle L. Mazurek. 2016. Sharing personal content online: Exploring channel choice and multi-channel behaviors. In *Proc. CHI*.
8. Janet Vertesi, Jofish Kaye, Samantha N. Jarosewski, Vera D. Khovanskaya, and Jenna Song. 2016. Data narratives: Uncovering tensions in personal data management. In *Proc. CSCW*.
9. Justin Chun Wu. 2016. *Peering Through the Cloud: Investigating the Perceptions and Behaviors of Cloud Storage Users*. Ph.D. Dissertation. Brigham Young University.