2016-10-01

Peering Through the Cloud—Investigating the Perceptions and Behaviors of Cloud Storage Users

Justin Chun Wu
Brigham Young University

Peering Through the Cloud—Investigating the Perceptions and Behaviors of Cloud Storage Users

Justin Chun Wah Wu

A thesis submitted to the faculty of
Brigham Young University
in partial fulfillment of the requirements for the degree of

Master of Science

Daniel Zappala, Chair
Kent Seamons
David Wingate

Department of Computer Science

Brigham Young University

ABSTRACT

Peering Through the Cloud—Investigating the Perceptions and Behaviors of Cloud Storage Users

Justin Chun Wah Wu
Department of Computer Science, BYU
Master of Science

We present the results of a survey and interviews focused on user perceptions and behaviors with respect to cloud storage services. In particular, we study behaviors such as which services are used, what types of data are stored, and how collaboration and sharing are performed. We also investigate user attitudes toward cloud storage on topics such as payment, privacy, security, and robustness. We find that users are drawn to cloud storage because it enables robust, ubiquitous access to their files, as well as enabling sharing and collaborative efforts. However, users' preferred medium for file sharing continues to be email, due to its ubiquity and role as "lowest common denominator". Privacy and security are of great concern to users, and though users vocally describe feeling "safe" on the cloud, this is because they actively filter the content they store in cloud services. Payment is a sensitive issue, with users exhibiting a strong aversion to any form of direct payment, preferring even disliked alternative funding mechanisms such as targeted advertising. Finally, the cloud serves as an important backup location for users, although space limitations prevent them from using it as a full backup solution.

Keywords: cloud storage, personal data, privacy

ACKNOWLEDGMENTS

Thanks go to my advisor, Dr. Daniel Zappala, my committee members—Dr. Kent Seamons and Dr. David Wingate, and my labmates & friends: Mark O' Neill, Elham Vaziripour, Scott Heidbrink, and Scott Ruoti.

Table of Contents

List of Tables **vi**

1 Introduction **1**

2 Related work **5**

2.1 Users and the cloud 5

2.2 Personal information management 7

3 Methodology **9**

3.1 Survey 9

3.1.1 Survey recruitment 9

3.1.2 Survey flow 10

3.2 Interviews 12

4 Survey **15**

4.1 Demographics 15

4.2 How people store their data 15

4.3 How people share their files 18

4.4 How people feel about their data 19

4.4.1 Appeal and concerns about the cloud 19

4.4.2 Payment 20

4.4.3 Privacy and Safety 20

5 Interview results **23**

5.1 Data mobility 23

5.2 Sharing and collaboration 24

5.3 Service Providers 26

5.3.1 Google Drive 26

5.3.2 iCloud 27

5.4 Payment 29

5.5 Security and privacy 31

5.6 Data permanence and backup 35

5.7 Ideal Cloud 37

6 Discussion **39**

6.1 Reconciling conflicts 39

6.1.1 Payment vs. privacy 39

6.1.2 Convenience vs. Privacy 40

6.1.3 Cloud benefits vs. email ubiquity 41

6.1.4 Personal preference vs. collaboration demands 41

6.2 Controlling personal data 42

7 Conclusion **45**

References **47**

List of Tables

4.1 Demographics 16

4.2 Cloud survey data 17

Chapter 1

Introduction

How users interact with and manage their personal data has been a topic of great interest and has been well-studied [1, 3, 4, 12], leading to the creation of a discipline devoted to understanding user behaviors in this domain. Much of this work, however, studied user behavior within a traditional desktop environment where user data was stored and maintained on devices that they personally owned and managed. The recent ubiquity of Internet access, the advent of the cloud, and the prevalence of mobile online activity has changed this core variable. A change in such a central assumption suggests that previous findings warrant review or renewal.

Initial work indicates that cloud storage has indeed changed how people interact with their data. A field study by Odom et al. indicates that people are struggling to adjust to digital ownership [13]. Users want to move their data online, due to convenience, ease-of-sharing, and backup, but are simultaneously uneasy about the loss of control and possible loss of data that comes with online storage. This was reinforced by Kang et al., who found that users worried about their data "going everywhere" once it was put online [9]. Likewise, work by Ion et al. [6] and Ur et al. [18] indicate strong privacy concerns regarding online storage and online advertising.

Despite these concerns, cloud storage is highly popular, and many questions remain about user attitudes and behavior regarding these services. In particular, we want to know (1) how data mobility and collaboration affect use of cloud storage, (2) how people chose particular storage providers, given their various capabilities, and how they manage data

across multiple providers, (3) why people are so averse to paying for cloud storage, (4) how attitudes toward security and privacy affect use of cloud storage, and (5) how a desire for data permanence and backup affects use of cloud storage. Consequently, in this paper we study general user perceptions and behavior regarding cloud storage, using a combination of surveys and in-depth interviews covering these topics. While other work has studied some of the privacy and security issues relating to cloud storage, they have considered these issues in isolation. By studying these issues within a larger context, our work is able to understand how specific attitudes and behaviors relate to each other. For example, given that users have concerns with privacy and ownership, how do these rank against other concerns, such as payment, ease-of-use, and performance? When these concerns come into conflict, how do they reconcile these concerns?

Our work includes both a survey of 385 participants and semi-structured interviews with 18 people that cover issues from the survey in more depth. We find that:

- **The primary perceived benefits of cloud storage are ubiquitous access to personal data, collaboration, and robustness. Privacy and security ranked highest as both the most preferred feature and the largest concern.** Participants use cloud storage extensively to collaborate, and enjoy the safety that comes from reliable, robust cloud storage. Given that all services provide ubiquitous access to personal files, collaboration, in particular, is often the defining reason users choose a particular cloud storage service over another. Users may join additional services or even switch their primary service provider in order to accommodate the preferences of friends, family, or their work environment. Despite the benefits of cloud storage, users are reluctant to store data in the cloud that they consider private, and privacy concerns go well beyond financial or personally identifying information to a broad set of additional data considered sensitive.
- **Despite awareness of the ability to share via the cloud, email continues to be the preferred method of file-sharing.** Email attachments are a type of

"lowest common denominator" for users: they know their contacts all use email and all understand how to use email attachments, and thus this mechanism is often preferred to sharing via the cloud.

- **Users have a strong aversion to payment.** The vast majority of users do not pay for cloud services, and have no desire to ever do so. The easy availability of basic, free tiers of service encourages this behavior, and users go to some effort to avoid payment, including willingly fragmenting their data across multiple accounts and/or services. Affordability appeared to be a key issue in their minds, though they had a difficult time quantifying what they would consider affordable.
- **Users strongly dislike data-mining practices, but rationalize away these concerns in order to avoid payment.** Users are both aware of, and uncomfortable with, data-mining of their personal data for practices such as targeted advertising. This discomfort notwithstanding, they understand that payment must come in some form, and prefer this to direct payment for cloud storage. Some users recognize that the business interests of providers can intersect with their own, and view these activities as being done to enhance the service itself, thus enhancing their own user experience.
- **Users feel "safe" on the cloud—because they filter what data they upload to it.** While largely trusting that providers are doing their best to protect user data, concerns about government access to private data and attacks from third-parties have made users cautious. Accordingly, they actively filter the types of data they upload to the cloud.
- **The cloud is an important backup source for users.** While space limitations prevent users from relying on cloud storage services as a full backup solution, it is nevertheless a trusted backup source for them. In particular, they store data that is considered difficult or impossible to reproduce such as photos or creative content.

Based on these findings, we identify four areas where users grapple with conflicting goals—payment vs. privacy, convenience vs. privacy, cloud benefits vs. email ubiquity, and personal preference vs. collaboration demands. These conflicts are typically resolved in favor of lower cost, greater convenience, and easier collaboration. We also identify a persistent theme that users generally have a strong awareness of the benefits of ubiquitous access to personal data, yet also a strong concern over lack of control over this data. The boundaries between public and private possessions have been blurred, and users struggle to understand potential threats and the actions they can take to safeguard their personal data. Overall, there is a clear need for end-to-end encryption of cloud storage, yet no straightforward answer for overcoming the problems of usability, trust, and aversion to payment.

Chapter 2

Related work

We consider two areas of related work: research that seeks to understand user perception and behavior with respect to the cloud, and research that studied how users interact with their data prior to the advent of cloud storage services.

2.1 Users and the cloud

Research on user perceptions and behavior with respect to the cloud has focused on three areas: privacy, digital ownership, and adoption.

The most relevant work to our research are two papers that have studied user attitudes toward privacy with respect to the cloud. Ion et al. [6] conducted both interviews and a survey with participants from Switzerland and India regarding privacy of cloud storage. They found users consider the Internet to be inherently insecure and prefer local storage for sensitive data, yet paradoxically they feel safe storing data with cloud providers because they don't personally feel that they are a target. In addition, users expressed a desire to pay a small amount ($20/year) for storage services that provide strong privacy guarantees. Ur et al. [18] conducted semi-structured inteviews about online behavioral advertising, finding that people have mixed views – they simultaneously considered it useful yet had strong privacy concerns. Relatively few participants were aware that browsing history was being used to target them with advertisement. In addition, participants were confused about current methods meant to inform them of advertising practices and did not understand how to effectively control their privacy.

Odom et al. [13] executed a field study designed to analyze how people feel about digital possessions relative to physical ones. This was performed via a set of interviews with thirteen individuals who used computers daily and had an online presence. Their findings include that there were strong motivations to move data online, such as ease-of-sharing, accessibility, and backup. Participants expressed value in storing certain types of files online, e.g., photos on Facebook. This, however, was coupled with anxiety about the fact they had no "discernable control over the services" and that they might "temporarily or even permanently lose access to them." Participants described much stronger feelings of possession—and thus security and trust—with personal storage they directly owned and managed.

Additionally, work by Kang et al. [9], sought to understand the extent of users' knowledge about the Internet and how that knowledge affected their online behaviors. They found that lay people had simpler mental models of how the Internet functioned and that those with greater technical expertise perceived a greater number of threats. This discrepancy notwithstanding, they nevertheless found that behaviors did not differ substantially, finding no direct link between ones' technical background and protective actions taken online. Further emphasizing the diffusion of data online, and the potential effect this has on privacy and security, this paper's title carries a tagline of "my data just goes everywhere," a direct quote taken from one of their respondents.

Several papers have studied the adoption of cloud services. Koeler et. al [10] ran a web-based survey of 60 employees of small and medium enterprises in Singapore, asking participants about their preferences. They found that the reputation of cloud service providers as well as their support for standardized data formats were considered of greater importance than cost. Park et al. [14] focused on the extent to which users switch to the cloud, as well as major barriers and enablers in doing so. Major factors for adoption included accessibility and collaborative functionality, while major obstacles to adoption included satisfaction with current tools and the breadth of provided functionality they used in their current tools. Li et a. [11] also studied user adoption of the cloud from a multi-theoretical perspective, finding

that attitudes, ease of use, and peer pressure are leading reasons for adopting cloud services. Shin [16] created a theoretical model of cloud service adoption, which captures positive cloud features in terms of perceived usefulness and perceived ease-of-use. He found that developers should construct seamlessly connected networks, that usability facilitates other key variables (access, availability, security, and reliability), and availability is key to perceived usefulness.

2.2 Personal information management

The discipline of personal information management is concerned with the study of how people interact and manage their personal data. This includes the evaluation and design of tools for managing information, as well as qualitative research studying the behaviors involved. This is relevant to our work because it provides insight into how people interacted with their data prior to the advent of cloud storage.

An early paper by T.W. Malone [12] is perhaps the first instance of work done in this area. Hailing from an organizational management background, Malone performed a series of interviews focusing on the ways in which office workers organized their desks, seeking insight into the way in which people organize information and, accordingly, the ways in which computer storage design should reflect that mentality. He found that the purpose of personal desk organization schemes was not simply to improve the efficiency of locating sought-after information, but also to remind people of tasks they needed to fulfill. Additionally, he characterized a possible role for then-future computer systems to play: assisting the user in the cognitive load of categorizing data.

Another major work, by Barreau and Nardi [1], came over a decade later in a paper titled "Finding and reminding: file organization from the desktop." With the appearance of the computer now prevalent in the workplace, they analyzed the ways in which users both stored and located their personal data. This was accomplished in two separate studies, one which first interviewed managers on the practices within their group, and the second with the employees themselves. Their findings can be summed up in four points: 1) location-based

file access was preferable because of its reminding function, 2) users preferred simplistic over elaborate filing schemes, 3) users archived very little information (perhaps owing the limited storage capacities of the time), and 4) personal information could largely be classified as falling into one of three time-based groups: ephemeral, working, and archived.

Bergman et al. [3] studied how people locate their personal files and the effects the host operating system has on this process. They found significant process and performance differences between Windows and Mac users due to both presentation issues (Windows' default presentation scheme is not its most "optimal") as well as folder organization strategies.

Chapter 3

Methodology

We conducted a two-phase study to investigate how people use and feel about consumer cloud storage services, beginning with a survey performed via Amazon's Mechanical Turk system, followed by local, semi-structured interviews that more deeply targeted points raised by the surveys. This study was approved by the Institutional Review Board of our university.

3.1 Survey

The first half of the study consisted of a Qualtrics survey of 24 questions, including both multiple-choice and open-ended questions. These included a limited number of demographic questions, such as age and education level, while primarily focusing on cloud behaviors such as which specific services they used and what they did with those services. Data that was quantitative in nature (such as multiple-choice responses) was analyzed with built-in Qualtrics tools, while responses to open-ended questions were later coded by hand.

3.1.1 Survey recruitment

Participants for the survey were recruited using the Amazon Mechanical Turk crowdsourcing service. The recruitment message indicated that the survey was intended for users of consumer cloud services, but did not mention the details of the types of questions that would be asked. We limited the population to US residents with a 95% approval rating on Mechanical Turk. There were no restrictions on age, but due to the restriction that users have an Amazon Payments account, participants must be at least 18 years of age. Participants were paid $1.10

for taking the survey, and 385 participants took the survey over the one-week period from May 26, 2015 to June 2, 2015.

3.1.2 Survey flow

Upon accepting the Mechanical Turk task, participants were presented with an informal consent message along with a link to the survey. In order to complete the Mechanical Turk task, participants were required to enter a code that was automatically generated by Qualtrics upon completion of the survey. Once participants clicked the link to begin the survey, they were prompted to input standard demographics such as age, gender, and education level, as well as an additional self-reported metric of "technical proficiency" measured via a 5-point Likert scale.

At this point, participants were presented with the following message:

> This survey—and all following questions—have to do with how you use cloud storage services. Generally, cloud services store data for you on machines owned by the service provider, rather than storing the data on your local machine.
>
> For example, Gmail stores your email on machines operated by Google, not your machine. This provides convenience, so that you can access your email from any computer that is connected to the Internet.
>
> You can think of cloud storage services as the digital equivalent of a storage unit rental service. The physical environment—and all maintenance—is handled by the provider of the service, while you are allocated some space where you can store your data.

Participants were then provided with a list of checkbox options, of which they could make multiple selections, indicating which specific cloud storage services they use. We also included two options that are probably not typically considered "cloud storage": Facebook and Youtube. Because we are interested in how people store personal data on the Internet in

general, and not necessarily just with the cloud, we included these options along with an explicit notice to only select those options if the participant uploaded their photos (Facebook) or videos (Youtube) to these services. If participants selected multiple options, meaning that they employed multiple services in storing their data online, they were further asked what benefits they derived from using multiple services, as well as how they decided what data to store with which service. Finally, if participants selected Dropbox, they were asked whether they liked its local caching functionality.

Next, participants were given a series of ranking questions regarding broad traits of the services they used. For each of these questions, participants were presented with the following lists of options, in randomized order, and were asked to rank their preferences by clicking and dragging each item to its appropriate place. They were first asked what they liked about cloud storage, and given the options of: "I don't have to worry about hardware failures," "automated file backup," "it makes sharing easy," and "I can access my data from anywhere." Next they were asked what traits of cloud services they saw as important, between: privacy/security, reliability, accessibility[1], cost, and ease-of-use. Finally, they were asked what traits were of greatest concern to them: privacy/security, reliability, accessibility, cost, performance, and permanence.

Participants were next asked about payment for these services. They were first asked whether or not they currently pay for, or have ever previously paid for, cloud services. If they responded in the affirmative, they were asked what had made it worth the investment; otherwise, they were asked if they would ever consider paying, and if so, what changes would need to be made before they would pay for these services.

The next series of questions concerned issues of privacy and security. They were asked if there are types of data they would never feel comfortable storing in the cloud.[2] Participants

[1]Accessibility is included here because this was the text used in the survey. We understand that "accessibility" in the usability context refers to disability-centered design. However, we used "accessibility" in its broader sense of "the quality of being able to be accessed" and included a clarifying statement as part of that item, such that participants were presented with: "Accessibility—I can access my data from anywhere."

[2]In retrospect, this question was perhaps poorly asked. Our original intention was to inquire as to whether there were types of files that they stored locally, but did not feel safe moving to the cloud. The subsequent

were then asked if they had ever considered what would happen to their data if their provider went out of business, which parties they thought could view the data they stored in the cloud, whether or not they felt "safe" storing data in the cloud and what they felt "safe" meant, and finally presented with a 5-pt Likert scale ranging from "very uncomfortable" to "very comfortable" with respect to how comfortable they felt with providers using their data for things not directly related to providing the service (this included a specific mention of targeted ads as an example).

Finally, participants were asked about sharing and accessibility. More specifically, they were asked what techniques they used to share files with others, choosing from a list of options such as "email attachments" and "URL links to items stored in the cloud." They were then given a ranking question asking them to rank the relative frequency of how often they used these methods, and then asked how they determined when each method was appropriate. The final question asked participants what methods they used to access their data across different devices.

3.2 Interviews

After the data from the first half of the study was analyzed, we devised a semi-structured interview guide delving more deeply into the issues we believed to be of interest. Interview participants were first given a link to the Qualtrics survey, and then interviewed at most one day later with respect to their answers.[3] These interviews happened either in person (15 of 18), or over video chat (3 of 18), and took approximately 30-45 minutes each. Half of these interviews, those of the students, were performed in a reserved room in the university library. Of the other half, 3 were interviewed over an Internet chat service, while the remainder were interviewed in our city's public library. Each participant was compensated with a $20 Visa

interview process, however, made it clear that participants often responded with types of information that they did not have any digital copies of.

[3]Links to the survey were emailed a day in advance and some participants took the survey immediately, whereas others took it just before the interview itself.

gift card for their participation, which occurred at the beginning of each interview, along with the distribution of consent forms for taking part in the study and audio recording.

Participants were all recruited locally, and were all from Utah County, the county our university is located in. There were 18 participants interviewed over a seven-month period ranging from August 2015 to February 2016. Of these 18 participants, half were university students, and the other half were not. Student participants were recruited via a set of flyers posted around our university campus, while the non-student participants were recruited via a largely identical Craigslist ad. We further required that non-student participants be at least 30 years of age in the hopes that a non-student financial situation might lend additional perspective on the relative value of cloud storage pricing..

These interviews were semi-structured with a prepared interview guide, but discussion freely followed participants' willingness and eagerness to discuss certain topics over others. Thus interviews could at times be focused on a single topic to the exclusion of others, as in the case of two individuals who were very passionate about online privacy. The interview itself followed a review of participants' responses to the survey taken beforehand, with participants being prompted to explain themselves more fully or offer examples. However, several new items of discussion were included, such as questions about what they thought the cloud was, what they imagined an ideal cloud service to be like, as well as digital tours of participants' cloud spaces, where convenient.

As is typical for studies of qualitative data generated by this interview format, coding and analysis was performed after individual interviews had been completed, and was thus performed while this segment of the study was still ongoing. This allowed us to make determinations for the oft-used terminating condition, i.e., when participants ceased to present "new" data.

The audio for all interviews was recorded, and transcribed afterward. Coding sessions were then held by both authors going over the full transcripts. A codebook was created in accordance with the broad topics, such as "payment" or "sharing," found in the first half of

the study, the Mechanical Turk survey, with additional categories added to the codebook as they presented themselves. Each code was assigned a color or mark, and passages of the transcript were marked accordingly. After the entirety of interview transcripts were labeled in this manner, these passages were then sorted by category, printed out, and cut into individual segments such that each passage was on its own piece of paper. We then held our second set of coding sessions, again, always with both researchers present. Categories were studied one at a time, with all the passages from that category laid out on a large, empty table. We then arranged these passages into subcategories based on shared topics or sentiment. We discussed how we had created these subcategories, labeled them, taking memos on a nearby laptop throughout the process. This data is presented in the results chapter, while our final discussion on larger themes we noticed is presented in the discussion chapter.

Chapter 4

Survey

Relevant to the goals of this work, the results of the survey have been broken down into three broad categories that answer the following questions: how do people store, share, and feel about personal data in the cloud?

4.1 Demographics

Table 4.1 shows the demographic breakdown of our survey respondents. Participants were primarily between the ages of 25-34 (47.1%), while the majority of the remainder came from the surrounding two age groups of 18-24 (21.6%) and 35-45 (18.5%). Participants were primarily male, with a roughly 60:40 ratio (59.6% to 40.3%) of male to female. Most had at least some college experience: 42.5% of participants had at least some college, while 39.8% were college graduates. Relative to the larger population of Mechanical Turk [7, 8], our sample population had a higher male demographic and was more diverse in age.

4.2 How people store their data

We asked which services participants used and, if they used more than one, to identify what benefits they obtained from doing so. This data can be found in Table 4.2(a). The most commonly used services were Google Drive (66.2%), Facebook (65.7%), and DropBox (55.8%). YouTube (38.5%), iCloud (24.6%), and OneDrive (15.5%) had significantly fewer users. While these handful of services accounted for the majority of users, 7% of participants also made use of the write-in field to list alternative systems that they used. While none of these systems in

Table 4.1: Demographics

Gender	
Male	59.64%
Female	40.36%
Age	
18-24	21.61%
25-34	47.14%
35-45	18.49%
46-64	12.50%
65 and over	0.26%
Education	
High school; no diploma	1.30%
High school graduate	10.42%
College; no diploma	30.47%
Associate's/technical degree	11.98%
Bachelor's degree	39.84%
Graduate/professional degree	5.99%

particular accounted for any significant proportion, it is perhaps noteworthy that this group was larger in size than that which used OneDrive, the only one of the major service providers for which this occurred. Note that Facebook and YouTube were included because they do allow long-term file storage; participants were instructed to select these options only if they uploaded photos or videos to these sites.

Participants were allowed to make more than one selection, and the vast majority (85%) reported using more than one service to store their personal data online. Those who did so were then prompted with an open-ended question to elaborate on the benefits they derived from using multiple services as opposed to just one. Several reasons were given: to facilitate sharing and collaboration when collaborators preferred a different service be used, to organize data by keeping specific types of data in specific services, and as additional backup locations (avoiding single points of failure).

The first group is well represented by one particular response, explaining that when their personal preferences differed from others, they had to use more than one system:

Table 4.2: Cloud survey data

(a) Which cloud storage services do you use?	
Google Drive	66.23%
Facebook	65.71%
Dropbox	55.76%
Youtube	38.48%
iCloud	24.61%
OneDrive	15.45%
Flickr	11.52%
Other	7.33%
Picasaweb	3.14%
(b) How do you share files?	
Email attachment	86.83%
URL (link) to item in cloud storage	63.31%
USB flash drive	39.22%
CD/DVD	12.32%
Other	3.36%
(c) What about cloud storage appeals to you?	
Automated file backup	85.64%
Don't need to worry about hardware failures	73.11%
Makes accessing files from different devices easier	69.71%
Makes sharing files easy	54.05%
Other	2.09%
(d) What do you care about most with a storage service? (Mean rank [1-6])	
Privacy/security: only the people I allow should be able to access the files I store and share	2.64
Reliability: the service should never go down	2.92
Accessibility: I want to be able to reach my files no matter where I am or what device I am using	2.99
Cost	3.09
Ease of use	3.44
Other	5.91
(e) What are your largest concerns about using storage services? (Mean rank [1-6])	
Privacy/security: Are my files secure?	2.51
Reliability: will the service fail when I need it?	3.44
Accessibility: can I access the service from any device?	3.48
Permanence: what happens to my files if the storage company goes out of business?	3.72
Cost: is it too expensive to store the files I want stored?	3.73
Performance: is the service too slow for me to use?	4.41
(f) Do you now pay or have you ever paid for cloud storage?	
Yes	10.22%
No	89.78%
(g) Could you see yourself ever paying for cloud storage?	
Yes	24%
No	76%
(h) How comfortable are you with your provider using your data in ways that don't directly relate to the service?	
Very uncomfortable	28.69%
Uncomfortable	21.73%
Neutral	24.79%
Comfortable	18.38%
Very comfortable	6.40%

"I mainly use Google drive because it can be accessed so easily and more and more people are using it so it's becoming easier to share that way. However, some people that I share things with use other programs like Dropbox so it's easier to have both and go back and forth depending on who I'm sharing with."

The second group use multiple services the way some might use folders on a traditional file system, as described in this response,

"I like my services to be limited to specific types of files to make storage less messy. Dropbox for books and magazines; OneDrive for most MS documents."

The last group were worried about the safety of their files existing only within one cloud service:

"[I] have the files in more than one location [so] in case something happens, I don't lose everything."

4.3 How people share their files

We also investigated how participants prefer to share files over the Internet (Table 4.2(b)). Here we report both pervasiveness—how many participants reported using this method to share data—and frequency—how many participants identified a specific method as either their most frequent or second-most frequent way to share data. Surprisingly, the most preferred method by far, both by pervasiveness (86.8%) and by frequency (89%), was email. The next most preferred method was sharing a link to a file stored in the cloud (63.3% pervsaiveness, 67% frequency). Note that these links are frequently shared via email. Because the two remaining listed options—USB drives and optical media— are physically bounded, this seems to highlight the convenience of sharing data over the Internet, as opposed to in person.

Participants were also asked to explain their preferences. Those who preferred email valued it for its convenience and its ubiquity. One response effectively captured both these sentiments, saying,

"I just use an email attachment for the sake of convenience, [...] because I know that's a method that almost anyone knows how to access."

4.4 How people feel about their data

We asked participants their attitudes regarding features of the cloud, concerns they may have, payment, privacy and safety. This data is reported in sections (d) and (e) of Table 4.2.

4.4.1 Appeal and concerns about the cloud

Participants were presented with a list of possible things that appealed to them about the cloud, and were allowed to make multiple selections. This list, and participant choices, can be found in Table 4.2(c). The vast majority seemed to be drawn to cloud storage for the same reasons. Most popular was ubiquitous access to personal data (85.6%), meaning access files from different devices. Many also liked how it alleviated the concern of local hardware failures, like a failed hard disk (73.1%), as well as how the cloud facilitated sharing and collaboration (69.7%), . Additionally, about half (54.1%) valued the cloud as an extra backup source.

Participants were next asked to rank features they cared about most on a six-point scale from "most important to least." The highest average ranking was privacy/security (2.64), followed by ubiquitous file access (2.92), reliability (2.99), and cost (3.09). The last option (excluding an unused "other" option), ease-of-use, fell far behind the others (3.44).

Participants were then asked to rank their concerns about the cloud on a six point scale from "most worr[ying] to least." Privacy/security (2.51) again had the highest average ranking, and was nearly a full rank ahead of the second-highest option. The next average rankings were reliability (3.44), cost (3.48), ubiquitous access (3.72), permanence (3.73), and performance (4.41).

4.4.2 Payment

Payment is a major concern for participants. Very few participants (10.2%) had either previously paid for, or currently pay for, cloud storage (Table 4.2(f)). Moreover, most participants (76%) stated that nothing could convince them to *ever* pay for cloud storage (Table 4.2(g)). Of the roughly quarter who stated that they could envision themselves doing so, write-in responses indicated that additional storage space and security measures were the features they wanted. Generally, the perceived return on investment for current pricing schemes seemed to be an issue because "affordable" was often mentioned in the responses given here, although what exactly participants considered "affordable" was not typically specified.

This finding is intriguing as it seems to contradict a finding by Ion et. al [6], who found that users were willing to pay $20/year for a cloud storage service with increased security. We discuss this in greater detail in a later chapter.

4.4.3 Privacy and Safety

We posed a series of questions investigating how comfortable participants were with the cloud when it came to privacy and personal data. Participants were asked if there were types of data they did not feel comfortable storing in the cloud, how they felt about cloud providers using their data for reasons not directly related to provisioning the service, and whether or not they felt safe with the cloud in general (as well as what "safe" meant to them).

The majority of participants (62.8%) indicated that there are data they would not feel comfortable storing in the cloud. Elaborating further in open-ended responses on what types of data they did not want on the cloud, responses were largely unsurprising, identifying sensitive data such as financial data or personally identifying information such as social security numbers (71%). Interestingly, however, in addition to these traditional notions of "sensitive" data, roughly a quarter of these participants (27%) raised concerns that seem to be a direct consequence of the growth of social media. For example, one person specifically

stated that sensitive data included *"anything that I wouldn't want a future employer to see."* Another qualified sensitive data as *"anything illegal or questionable or that would cause problems/loss of reputation if other people saw it"*. One participant specifically mentioned pirated files that were in violation of copyright. Photo leaks seemed to be the source of at least some of these sentiments, with specific references to the iCloud celebrity hacks [2], as with one person who remarked that they would never store *"any compromising photos. Take heed, Jennifer Lawrence!"*

The majority of participants were uncomfortable with data mining that wasn't directly related to providing the service, as in the example of targeted ads (Table 4.2(h)). Presented with a 5-point Likert scale ranging from "Very uncomfortable" (1) to "Very comfortable" (5), half (50.4%) of participants identified themselves as uncomfortable (either a 1 or a 2) with this practice, including over a quarter (28.7%) who were "very" uncomfortable. It must be noted, however, that these attitudes cannot be taken without context. While participants may feel uncomfortable with these practices, it is notable that, given their strong preferences for avoiding direct payment, they are likely simultaneously active users of systems which engage in this very practice [17].

Participants also explained via open-ended responses whether or not they felt "safe" using the cloud, and what they felt constituted safety on the cloud. Most participants (78%) explicitly stated that they felt safe using the cloud, although this must be also taken with the caveat that participants described pre-filtering what types of data they allowed into their cloud storage. Those who did not feel safe attributed their feeling to issues of trust, as in this example:

> *"Absolutely not. I do not trust cloud services to protect my information either from hacking, governmental intrusion, marketing services, using it for their own benefit (selling, direct marketing, etc), and other unauthorized uses".*

Some responses were quite passionate, with one participant stating simply,

"No, I do not, as I have already explained. People would have to be stupid or insane to trust this. (or both)"

When further asked to qualify what safety on the cloud meant to them, responses largely fell into three categories: privacy, reliability, and pre-filtering. Privacy was viewed by participants as meaning that others ought not be able to view their data without their permission, while those who mentioned reliability felt safe if their data was there whenever they needed. The last group felt safe specifically because they actively filtered what data they stored in the cloud. As one participant described,

"I don't store anything particularly important on any cloud because I am concerned that such information could be stolen. For something to be safe, it would have to be impenetrable. And I don't believe any cloud service is."

Chapter 5

Interview results

We followed the survey with a series of in-depth interviews so that we could better assess sentiment toward cloud storage and how this affects behavior. In particular, we were interested in the following topics: (1) how data mobility and collaboration affect use of cloud storage, (2) why people chose particular storage providers, given their various capabilities, (3) why people are so averse to paying for cloud storage, (4) how attitudes toward security and privacy affect use of cloud storage, (5) how a desire for data permanence and backup affects use of cloud storage, and (6) what an ideal cloud storage system would do.

5.1 Data mobility

Survey participants chose data mobility as one of the primary reasons why they use cloud storage. In our interviews, this need manifested along two dimensions: the expectation that people will be "able to access [their] files whenever and wherever" (P2) and the need to access their data from across different devices and platforms. One interviewee explained that with cloud storage, "it doesn't matter whether it's iPad or Android or tablet or Windows or Mac, you can access your file anywhere" (P0).

Surprisingly, despite using the cloud for mobility, some interviewees nevertheless continued to email files to themselves as an accessibility method. One interviewee stated simply that "old habits die hard" (P0). Another explained that he preferred to email himself when it was a "one-time use," using the cloud when it was something he was "constantly updating" (P4).

5.2 Sharing and collaboration

Survey participants chose sharing and collaboration as another major reason why they use cloud storage. As one interviewee put it, *"whenever I need to collaborate with anything, it makes it really easy for me"* (P1). The fact that files could be actively shared by collaborators relieved them from having to manually synchronize efforts:

> P9: *We kind of use Drive as a go-between for working on files that everyone in the office is touching here and there.*

The use of cloud storage in the workplace was common among interviewees, even when employees are not on-site. One interviewee described using it in multiple work environments, noting,

> P14: *I have worked with several different people as a team working together on something, and so we share our work with each other. In my experience, that is much more common. This is when I was doing more graphic design stuff and web development, though. The more I get into video editing, the more I have been using cloud services to send footage to the people I have been working with as well.*

What is fascinating, however, is that despite their willingness to employ cloud storage in collaboration, email nevertheless appeared to be the preferred medium for file-sharing, corroborating our earlier survey results. This seems to be because email serves as a lowest common denominator of sorts. More specifically, while people may be uncertain of which cloud services their acquaintances are using, they know their counterpart will assuredly have email access. As one person put it,

> P1: *A lot of people I know don't have like iCloud or stuff like that. Like people in my family, to send pictures to them, I'll just send em [with email].*

Indeed, going beyond the ubiquity of email access, assumptions of familiarity with how to deal with email attachments is important. One interviewee, a vocal teacher, provided a most illustrative example:

P10: *It's easy! That's it, that's all you do! But I have a lot of older clients for voice lessons. I have a client who's 98! She wants to learn how to sing and she's 98, that's freaking amazing! But she does not know any of this, but she does know email, so for her it's easy. So for the younger ones too, I have one that's 14 and one that's 8, and the 8 year old, she can do email. She doesn't have a phone yet though. That seems to be the easiest: from 98 to 8.*

Additionally, email was often considered the most convenient. One interviewee noted that cloud sharing notified the respondent through email anyway, and *so if I can use email to start with, why not just use that?"* (P14). Another described preferring attachments because it was faster, stating that,

P15: *It's really just what is fastest. Sometimes it'll come down to thirty to sixty seconds of a difference, but I don't want to have to upload it to cloud, click share, type in their email. It's just maybe one step shorter, but I'll just click send, I start typing in their email, it auto-fills it, I click attach, it's on my desktop, send.*

The preference for email notwithstanding, links to cloud data were still frequently used. When compared against sending email attachments, two main points were identified. First, email attachments require knowing the recipients beforehand. With the cloud, on the other hand, one interviewee explained,

P2: *If it's to a whole bunch of people, or especially to people that I don't know, I'll make a public folder on Dropbox and give people the link. [...] We had a photo booth at one of our wedding receptions. [...] So I just put all the photos there into a public folder and gave it to my mom who sent it out to everyone. Because I d[id]n't know everyone that wanted to access them so that public folder was just available to whoever had the link.*

Secondly, unlike with email, which is a medium itself, cloud links can be shared over a variety of communication media. As one interviewee put it,

P8: *It's easy to just drop anywhere no matter what interface you're communicating through. It's really easy to share.*

5.3 Service Providers

Those we interviewed primarily used Google Drive and iCloud as their preferred cloud service. We explored the reasons for these choices.

5.3.1 Google Drive

In a microcosm of the more general reasons people gave for using cloud storage, interviewees who use Google Drive stated that they did so for two main reasons: collaboration via Google Docs and mobility. Google Docs, in particular, appeared to be a major draw, facilitating workplace collaboration. One interviewee described how it enabled low-latency collaboration, and thus preferring it to an alternative offering from Microsoft:

P6: *I like their online editing of Google Docs and Google Sheets, I think that's a really handy solution—the best one that I've tried so far. Microsoft has started with their Office Online—it's got a lot more tools and you can do a lot more—but the sharing and the multiple people editing is not quite as developed yet. It seems like there's a lot more latency—like my boss will type something and it will take like 5 minutes to show up whereas with Google Drive it will show up instantaneously. So even though it's more simple, it's a lot better of an interface and it has a nicer feel to it.*

P6 also noted that Google Drive worked well for sharing files between collaborators both near and far as he shared files *"between me and my boss"* and also *"with people we hire on elance"* stating that this was *"the main reason why we do it—for easily sharing files between people working on projects."*

Data mobility was another major motivator for Google Drive and cloud storage in general, particularly for students who needed access to their schoolwork both from at home

and from on campus. As one student explained, cloud storage enabled her to avoid the physical burden of bringing her laptop with her:

> P5: *I would leave my apartment—because I didn't like to take my laptop because it was too heavy—so I would always email what I was working on and so having Google Drive was very nice because I'd just sit down, and it'd pop up and it would save.*

In this manner, cloud storage replaced their previous solution to the mobility problem—email: *"Oh, you can save your files on Google Drive and then [...] you don't have to email your documents to yourself."*

Others who did not discover Google Drive through its Google Docs functionality were introduced to it by others who were already using the service, the effect of which could be quite pronounced. One interviewee mentioned how

> P4: *One of my teachers suggested to use Google Drive for an assignment to collaborate with other students. [...] I didn't even know it was a cloud, technically, at that time, but that's when I was first introduced to Google Drive and since then I've always used Google Drive.*

Indeed, one was even convinced to switch from another cloud storage provider because of how frequently another was used by those around him:

> P13: *I saw TV ads about Microsoft's cloud. Then I used Apple and learned about iCloud. [...] At my work, they started using Dropbox. [...] But my friends and family kept sending me stuff using Google Docs and I finally decided not to use Microsoft Office and just use Google stuff.*

5.3.2 iCloud

Users of iCloud readily identified two main functionalities that caused them to prefer this service: its integration into the Apple ecosystem and its use in backup of their Apple products.

Those with multiple Apple devices were pleased to find a deeper level of synchronization through integration with built-in Apple software than just general file storage, automating accessibility of data such as calendars and contacts. One interviewee explicitly attributed their preference for Apple products to this type of functionality:

> P8: *It's actively integrated. That's the kind of thing that bought me over with Apple. All the apps, the way they work between my iPhone, my iPad, and my Macbook. It's really nice.*

The other draw for Apple device users was the automated backup of their device data, and photos in particular, to iCloud:

> P8: *That's actually one thing I really like about iCloud now, if you're an iPhone user, those photos are automatically backed up.*

However, automated backup via iCloud, while a draw of the system, was also problematic for some users. One user disliked not being able to control when synchronization occurred, saying,

> P0: *That's why I stopped syncing it now. [...] I want to sync when I want to sync not when Apple thinks I want to sync.*

Illustrating how confusing this process could be for users, P8 discovered during the interview that this automated synchronization process was enabled despite their having previously disabled it.

> P8: *At one point, I had it on so that it'd always backup my photos to my computer, but that kinda got annoying because I somehow ended up with multiple copies of certain photos. Weird stuff like that. So I actually turned it off and I'm not sure right now if I have any iCloud backup from my iPhone. [checks phone] Okay, so it looks like I have it on for everything.*

5.4 Payment

In keeping with the results of our survey, while a handful of interviewees either continue to pay or have previously paid for cloud storage, for the majority were unwilling to pay. Interviewees were willing to inconvenience themselves—fragmenting their data across various services—or discomfort themselves—submitting to targeted ads, a practice they found distasteful—in order to avoid payment. Indeed, direct payment seemed a worst-case scenario, as captured by the following statements:

> P3: *I guess since they're giving the service, I'm giving them what I have and they have it. [...] They're trying to make business. It's annoying and I wish they wouldn't, but they can and they have to. [...] I really just don't want to pay. If I have to get ads, then I get ads. Whatever.*
>
> P0: *If I use their free service, you know they're gonna get something from you. [...] Eventually they don't start having ads on, they'll have to charge us money.*

Interviewees described going to some lengths to avoid payment, including fragmenting their personal data across multiple accounts or even across multiple services, even if it meant using a different service that they viewed as less-preferred, as in the following two exchanges.

> P3: *If Dropbox decided 'hey, we won't offer this unless you pay for it' I probably would switch to something different, like Google Drive.*
>
> Q: *So it being free matters a lot to you?*
>
> P3: *Yeah.*

> Q: *Would you pay for Google Drive then?*
>
> P15: *[shakes head]*
>
> Q: *Is it because the service itself isn't worth it or is it because it's too expensive?*

P15: *I'd say that the market just doesn't demand for it because there are other cloud services I can go to just as easily.*

Q: *So you can always find some free offering that will do what you need?*

P15: *I don't need it to be the best version because a copycat version of Google can do, let's say, 90%, even 80% as good a job. Maybe not as cleancut and smooth as Google can, but it gets the job done, and so why pay for a cloud?*

As with the survey, it was difficult to get interviewees to quantify what they would consider an "affordable" pricing for additional cloud storage capacity. However, it became evident that at least one factor that entered consideration was a comparison against the pricing of physical hard drives, which comparison they found unfavorable. As one interviewee explained,

P4: *It's like buying a terabyte drive. If I can go out and buy a terabyte drive and store things on there—it's portable, I can take it with me. If I can buy that for a hundred dollars, why would I buy a service that gives me a hundred gigabytes for three dollars a month. It just doesn't make sense to me to do that because I can store a hundred times as much on my terabyte [...] The three dollars a month adds up over time and you'll end up paying more than you would for the terabyte.*

Interestingly, this participant also somewhat contradictorily felt that the ad-funded model was actually conducive to the functionality of the service.

P4: *I believe that I would get a better continuous service from the one doing the data mining since that's what they're interested in, rather than the other one, because they're selling you a product to sell you a product to sell you a product. It's an added service, yes, but that's not what they're interested in so they're not going to be developing that and they won't be as fast as continuous as the other model.*

Q: *Because the other model needs to keep you.*

P4: *Yeah.*

5.5 Security and privacy

During our interviews it was clear that interviewees cared a great deal about security and privacy, often giving prolonged responses. While most interviewees generally felt safe with their use of cloud storage services, these feelings were nuanced. Firstly, privacy seemed a greater concern than security, due both to faith in cloud service providers and manual control over what files they allowed on the cloud, preferring not to upload sensitive data. While interviewees did view security as a concern, this was alleviated by the thought that their data was of no interest to a third-party attacker.

P4: *I don't have anything out there that would 'blackmail' me, I guess you could say. [...] So, I don't feel like I have anything out there that needs to be protected, but I also I do trust in that it's secure. So it's not that big of a worry for me, if that makes sense.*

P5: *I'm more worried about the documents being lost, than being hacked, or anything like that because I usually don't put secure things on there. I guess when I think 'secure' I think are they there on the Drive—on the cloud—as opposed to my identity or something. Because if somebody stole the things that are on there, I'd just be worried that they'd be gone, not that my privacy was a problem.*

P7: *If I was famous, I'd probably worry about it a lot more but I'm not famous. No one's out to get my pictures.*

Interviewees trusted providers because, having thought through providers' incentive models, and having come to the conclusion that their best interests coincided with providers' best interests, they had determined that service providers were worthy of their trust. One respondent further analogized the situation to buying a car, stating that privacy was not a

consideration for him when choosing a cloud storage provider because privacy was a universal feature across such services.

Q: *So is that because you trust Google itself? What separates them being okay from if they gave your data to someone else, then it's not okay?*

P5: *I guess I see them using that information to, not to help me, but to, improve their services rather than using it against me.*

P6: *I'm not really looking for someone that has great privacy since they all have it.*

Q: *So it's not that it doesn't matter to you but that you feel that they're already providing it.*

P6: *Right. So I go with something that has more additional features. Like on a car, wheels are important features, but that's not what I look for in a car because all cars have wheels."*

Interviewees also made it clear that trust was earned over time.

P7: *I mean, I've never had a problem in the past with it, and I don't expect to have a problem and also I trust that they won't have a business if they don't do that. I would say I have a lot of faith in the market economy that, you know, they're making a service based on consumer interest, so...*

Q: *So it's not necessarily something that they advertise that convinces you, like encryption or whatever. That's not necessarily what does it for you, it's more like experiential and an understanding of...*

P7: *Empirical evidence.*

Q: *So do you feel like Google makes an effort to protect your data?*

P10: *I do, I really do. [...]*

Q: *Is it sort of an implicit trust? Because it's Google you trust them?*

P10: *No, it's because I've been using it for ten years.*

Although most interviewees trusted service providers, others were decidedly uncomfortable trusting both providers and the Internet in general. That data on the cloud passed out of their direct control and management was worrying to them—the fact that they simply did not *know* what others did with their data concerned them. One interviewee explained her distrust of providers, stating that regardless of what they promise in their terms of service, there does not exist a watchdog to enforce this behavior: *"The real question is do we know? Are they accessing our data? No one knows"* (P0).

The notion of government access to private data prompted mixed feelings, as characterized by the following exchanges, which demonstrate diametrically opposed viewpoints.

P14: *People have a right to privacy, though I guess it depends on the type of data.*

Q: *So what kinds of data would you be okay with [the government] looking at? What kinds do you think should be off-limits?*

P14: *Well, the NSA stuff in general, I am mostly against. I can't really think of things I'd be okay with at the moment.*

Q: *So have you heard about the NSA and the government, how they were looking into these things?*

P10: *Yeah, yeah.*

Q: *So how do you feel about that? Has it changed the way you look at the cloud at all?*

P10: *No. *laughs* Not at all. To me... there's a lot of people who are severely interested in keeping government out of their stuff. I don't care. You wanna see what I got? This is what I got. What are you gonna do with it? I'm not doing illegal stuff, and so I don't care. Maybe if I did, I would, but I'm not.*

As our survey found earlier, the practice of data-mining cloud data was undesirable to most. Interviewees elaborated on these feelings at some length, finding the practice "creepy," as one put it. This was despite an understanding that algorithms, and not people, were the ones involved. One participant passionately stated that,

*I **hate** it that Google scans all my emails and I don't really mind so much getting a targeted ad, from what I wrote, I just **hate** it that they're looking at everything.*

Q: *So do you think there's a person actually reading your stuff?*

P16: *Oh, no, no. Algorithms. They just scan the file. When I say 'they', I mean their computer systems are analyzing it for big data purposes.*

Q: *So even that bothers you?*

P16: *Oh, a lot.*

One interviewee was particularly concerned about images stored in the cloud as opposed to textual data, because unlike textual mining, which they considered simple keyword matching, they felt that mining image data directly corresponded to understanding its content, which was viewed as invasive:

Google, for instance, [...] they took it to the next step. They took it from we're reading all the words you generate in any of our systems and then now they took it to pictures. Now they're like okay, how can we best advertise? They took a bunch of pictures of the same cat—not just a cat, but the same cat—they must have a pet. Let's start advertising or selling advertisements based to Petco or whatever. They've just taken it to the next level (P15).

To this end, users of iCloud noted that Apple's explicit declaration that they do not mine iCloud data as one of the reasons they preferred that cloud storage service in particular. As one interviewee put simply, *"I like that their business model isn't based on serving consumer data"* (P15).

Another interviewee, who used both iCloud and Google Drive, made a direct comparison between the two in this respect:

> "P7: *Anything personal, though, I don't like to put on [Google Drive] just because I know they do data mining. I'd like to keep private. I know that Apple's servers, they don't take money off of, like, marketing.*
>
> Q: *So the fact that Google, for example, will use your data to serve you targeted ads is something that makes you uncomfortable?*
>
> P7: *Uh-huh, yeah.*

These feelings, however, were not universally shared. Some were willing to accept their data being mined if it improved their experience in some way. Indeed, one interviewee described willingly selling browsing data through Screenwise Trends Panels, stating that *"if I'm compensated and am told about it. It's like 'alright, it's fine' "* (P6). Another explained that *"if I have to suffer through ads—it's nice to get ads that I might possibly care about"* (P9), while a third assumed their data was being used to improve the service itself, saying, *"I guess I see them using that information to, not to help me, but to, improve their services rather than using it against me"* (P5).

5.6 Data permanence and backup

Cloud storage serves as a trusted source of backup for users. Interviewees were quite explicit about the extent to which they placed their faith in the cloud, and remarks like the following were quite common:

P7: *I have maybe a false sense of security, but I do feel very secure with Google's servers and Apple's servers. [...] I probably trust the cloud more than I trust my own drive.*

P9: *It's just like in case if I ever, for any reason, need this, I know I can find it.*

For one interviewee, this trust was born of experience—they had previously experienced a hard drive failure in the workplace, but having stored their files in the cloud, they were able to recover the data they needed.

P6: *It was nice to know that those files were still accessible—most of our critical files were on the cloud so it didn't matter if we couldn't get to the drive.*

Because cost is an issue, users are reluctant to backup everything to the cloud, and so users frequently offload certain types of data to a local storage device when they no longer need it in the cloud.

P0: *Anything that I think is old should not be taking up storage space, so I'll move it to my hard drive.*

P8: *Things I might look back on every now and then but don't need access to all the time [...] are all stored on a 1 TB external hard drive. All the extra documents that I want to hold on to I'll use Dropbox mostly.*

With respect to the types of data being backed up to the cloud, users focused on data that was difficult or impossible to reproduce, like photos or creative content. Applications, on the other hand, are now seen as expendable, since they can easily be re-downloaded.

P15: *For instance, some of the things that I've written. I'll have like twenty to forty pages that took me like a year to write. Even writing an essay sucks for me, because I'll write like a five-page essay and it's just like, you know, all that creative power. If it gets deleted for whatever reason, I'll just be like, forget it,*

I'm not gonna... You can't just recreate stuff. It's just stuff that's irreplaceable basically. So writings, pictures, um... That's pretty much it.

P15: *I've lost stuff before. A lot of it's human error, but now I just backup anything that I would not wanna lose. Especially working at Apple... So many people would come to us and they'd be like 'this happened', but I'd be like, 'I'm so sorry, but that's lost.' Barring going to a data extractor, and in some cases, even then it's not recoverable. Asterisk: you'd have to go into like hugely expensive levels to extract stuff, like water damage. So yeah, losing stuff just sucks so much, you know? Because you can't like retake a picture, for example, of your grandmother in her last week. And so I want all that stuff backed up and I'm talking redundancy like it's on here, back it up to the computer, back that up to two physical hard drives, and then to two different iClouds. Like some of the stuff I have is redundant like four times over.*

P8: *I'd say the primary reason I'm not too worried about a data crash is that a lot of my data is either hosted on a cloud service or is something I can download again. [...] Like I mentioned, a lot of applications now, they're not one-time purchases. You have an account that allows you to download it. So in that sense, I'm not worried about losing any of the apps I paid for because if I get a new Mac, I can just login to the App Store and redownload whatever I've purchased.*

5.7 Ideal Cloud

At the end of each of our interviews, we gave users a prompt: if there was anything about the existing cloud that you would change or add, or if you could imagine the perfect cloud, what would it be? We were hoping that pain points in the current technology would present themselves as possible solutions. The general sentiments from participants indicate they would like a cloud that is user friendly, integrated with applications (like Google), secure, keeps data private (i.e., avoids data mining), and is free. While it is unlikely that all of these

features could be met while keeping storage free, this does offer some insight into what users value.

Despite recognizing the risks associated with data being online (snooping by providers, surveillance from the government, hacking from malicious individuals), participants were generally eager to embrace the cloud. This sentiment is best captured by one participant who explained that they wished they could access their "entire computer" over the cloud:

> P9: *I mean, even the idea of having my entire computer available - like all the programs and everything available - that would be just unbelievable. I would totally pay for that. That would be awesome, if I could access my computer anywhere, rather than just a drive account. So like how Google offers all the different features - Docs and Sheets and Draw, all the little subsets of programs that allow you to work the data that's already up there.*

Chapter 6

Discussion

In this chapter, we discuss the impact of our results with respect to several broad themes in computing.

6.1 Reconciling conflicts

By studying attitudes and behaviors in a general manner, we are able to better understand not just which concerns users had, but also how highly they ranked them relative to one another. We also identified several cases where these concerns came into direct conflict, and were able to explore some of the rationale participants use to reconcile the conflict. We now discuss four such examples: the choice between payment and privacy, the similar conflict between convenience and privacy, the decision to use email for file sharing instead of the cloud, and the importance of collaboration and sharing in determining which cloud service to use.

6.1.1 Payment vs. privacy

The conflict between payment and privacy is perhaps best observed by the difference between our finding—that users view payment so negatively that they are willing to trade some of their privacy to avoid it—and the finding by Ion et. al [6] that users were willing to pay for cloud services that provided privacy guarantees. In our survey results 76% of participants say nothing could convince them to ever pay for cloud storage, and our interview results confirm this reluctance to ever pay for the service. Furthermore, only 10% of survey respondents had

ever paid for cloud storage. Yet Ion claims 79% of users are willing to pay $20 per year for storage services with strong privacy guarantees.

We believe there are two primary reasons for this discrepancy. First, their respondents were presented with only two choices, with a bias that would lead users to choose paying for privacy. Participants were presented with a choice between two cloud storage providers. Company A "offers the service for free, but their privacy policy says that they may sell, transfer or share your personal information and documents to another company," whereas Company B charges $20 per year, but "they will not sell, transfer nor share any of your personal information to another company."

The description of Company A seems somewhat terrifying: they can sell your personal information, not just your data, and even then, they might "transfer" your documents entirely to another company, and not just gleaned metadata. In reality, however, there are not only multiple choices for cloud storage providers, but the description of Company A does not accurately reflect what current providers do with data-mining and free tiers of service. Second, the Ion work did not probe user *behaviors* in this regard. Many people will declare a choice when presented in the abstract, but make a different decision when confronted with the consequences of that choice [15]. In the Ion study, users were not *actually* making that payment. Had their participants ever actually paid for increased functionality of a service? Perhaps if more general issues of payment had been assessed as well, this would have come to light. There is, granted, the question of whether users are aware of the many choices they have, but several of our participants indicated they chose iCloud because it gave them a service that did not use online behavioral advertising or sell their data to third parties.

6.1.2 Convenience vs. Privacy

Another conflict users face is between the convenience of cloud-based sevices and privacy-invading practices from cloud providers. We confirm the findings from Ion et al. [6] that users are skeptical of the safety of putting sensitive documents online, yet do not feel they would

be a target of a hacker or of government surveillance. However, our work adds additional context to user behaviors – many practice self-filtering to avoid storing sensitive information in cloud services, and many have developed a sense of trust in major providers that they will not show their data to others. We also confirm the work of Ur et al. [18] that users find targeted advertising "creepy" and have a complex relationship with the practice. We find that users have developed multiple rationale for accepting the practice, including avoiding payment and recognizing that the business interests of large providers may intersect with their own values and preferences.

6.1.3 Cloud benefits vs. email ubiquity

Despite the importance placed upon the role of sharing and collaboration in using cloud storage—the extent to which Google Docs enable real-time collaboration was a primary draw to a number of Google Drive users, for example—we nevertheless found that email was the most frequently used mechanism for file sharing. The ubiquity of email access, the ability to interface even with those who use a different email provider, as well as its relative ease-of-use, make it preferred to sharing via the cloud.

6.1.4 Personal preference vs. collaboration demands

Users often expressed personal preferences for a particular cloud storage service, but this conflicted with the need to collaborate with coworkers or clients on projects. In practice, choices for a cloud storage provider were heavily influenced by the choices of others. Some users were introduced to the cloud via friends, family, and/or coworkers, so their choice was entirely dependent on collaboration. Others were convinced to change from their existing primary provider to another because those around them prefered a different service, as in the case of P13 who "finally decided not to use Microsoft Office and just use Google stuff" because "[his] friends and family kept sending [him] stuff using Google Docs." Likewise,

participants reported using particular cloud services because work collaborators used these systems, or because clients preferred a particular system.

6.2 Controlling personal data

Our work confirms and adds additonal context to work by Odom et al. on how people interact with digital data [13]. This is important because Odom's study was based on a field study of only thirteen participants. From our interviews, it is clear that the introduction of the cloud as well as the ubiquity of Internet access and mobile devices has changed the way people view and interact with their personal data. Users' data is now far more mobile than it was previously and is accessed and stored across a variety of devices and platforms that they may or may not own or control. New concerns of privacy and security have arisen as potential access to user data has stretched all over the globe—the threat of undesired access now stretching far beyond the limits of physical proximity to a device. Simply put, the "personal" in "personal data" has changed.

In addition to threats to "high-value" data such as financial information, personally identifying information, or other, similar types of sensitive data (such as confidential work-related files) [5], it is clear that the increased exposure that users face in the current online environment has some ill at ease even when it comes to types of data that have not traditionally been considered a concern.

For example, the thought that family pictures might be viewed by unknown individuals left P5 uncomfortable with the situation on Facebook, which they described as "out of control right now," prompting them to remove otherwise innocuous photos of their nieces and nephews from the site. The survey we conducted revealed several responses including other types of data participants considered worth protecting that similarly did not possess financial value, including two respondents who mentioned that they would not feel comfortable putting their journal or diary in the cloud, as well as a number who explicitly mentioned "sensitive" photos of themselves.

The boundaries that separate what is personal and private from what is public have been blurred by putting data online. In the physical world, users are aware of the threats posed to them and their possessions, have an understanding of the behaviors needed to protect those belongings, and have the ability to gauge the relative safety of those things. Consider, for example, the diary mentioned above by one survey respondent. A diary is traditionally considered personal and private, but is hardly worth the effort to invest any significant financial resources to protect. Its value largely lies in the fact that it is personal, and there is an implicit cultural understanding of this fact in others. As such, efforts to keep it private might be as simple as hiding it underneath one's mattress or locking it shut with a simple lock. More valuable belongings, such as expensive jewelry, might instead be kept in a security deposit box at the bank or in a safe, with an understanding of the relative amount of difficulty required to gain unsolicited access. Similarly, bars on the windows, or locks on a door allow users to—at a glance—determine whether or not a particular location is protected.

When it comes to the online world, however, participants in our study have much less of an understanding of potential threats and of what they can do to protect their data. Instead, they rely on their relative undesirability as a target (i.e., I'm no one and my data is worth nothing) as well as simply avoiding putting certain types of data online. Indeed, this self-filtering behavior is indicative of the lack of control participants feel when it comes to protecting their data: despite participants' assertions that they trust cloud storage providers and feel safe with them, they nevertheless are forced to exhibit an all-or-nothing approach to protecting their data, either uploading it to the Internet and hoping for the best or not uploading it at all, in sharp contrast with the way protecting possessions in the physical world is handled.

This effect is made more pronounced by users' aversion to direct payment and providers' subsequent turn to alternative funding mechanisms, particularly those that involve mining their personal data—an act that explicitly blurs the boundaries of what it means for data to be "personal." The majority of participants were distinctly uncomfortable with the practice,

with one participant who explicitly stating that he "hated" that Google did this (P16). When prompted, he recounted the following story, which he had found unsettling:

> *A nephew of mine moved to Norway working for General Electric in their oil and gas division. I sent him an email saying 'Hi, how are you doing'; just social stuff. I started getting ads for MRP resource planning and geology equipment for exploring oil and gas. All I did was communicate with an email address and based on that email address... They repurposed it for a bajillion ads.*

This participant well understood that there was not a specific human individual behind such actions—correctly attributing this behavior to "algorithms"—but he was nevertheless disturbed by experiences such as this, although he noted that such sentiments were at odds with his continued use of these systems. He remarked, *"given that I don't have that option, I give up the privacy and keep using it"*—a statement that seems to effectively capture the attitude of many like him.

It is this lack of control and lack of knowledge that concerns users. As P5 expressed, because they have no way of knowing exactly what providers and other potential third-parties are doing with their data, nothing would be enough to reassure them. These concerns have been further exacerbated by open knowledge of online behavioral advertising and stories in the news about government access to private data, in addition to attacks by malicious individuals. It is thus clear that more needs to be done to reassure users, both by better informing them of the threats they face online and what they can do protect themselves, and by greater transparency when it comes to how their data is being used and accessed.

Chapter 7

Conclusion

We presented the results of a two-part study on user attitudes and behaviors with respect to cloud storage services. We surveyed 385 U.S.-based users of cloud storage to identify both direct behaviors such as which specific services they used and why as well as more generic sentiments such as whether or not they felt safe using the cloud. This survey was succeeded by 18 in-depth interviews performed with a local population to further delve into these topics.

We find that users are drawn to cloud storage because it enables robust, ubiquitous access to their files, as well as enabling sharing and collaborative efforts. However, users' preferred medium for file sharing continues to be email, due to its ubiquity. Privacy and security are of great concern to users, and though users vocally describe feeling "safe" on the cloud, this is because they actively filter the content they store in cloud services. Payment is a sensitive issue, with users exhibiting a strong aversion to any form of direct payment, preferring even disliked alternative funding mechanisms such as targeted advertising. Finally, the cloud serves as an important backup location for users, although space limitations prevent them from using it as a full backup solution.

We highlighted several ways in which a more general study of the issues involving cloud storage can be beneficial, particularly with respect to conflicting concerns that users may have. We discussed several examples of this occurring, such as users' struggle to reconcile their aversion to payment with the desire for privacy, drawing particular attention to how this contrasted with a finding from an earlier work with a narrower focus.

We also elaborate on what we see as the implications of these findings: namely, that personal data is undergoing an evolution due to the introduction of the cloud and the rise of mobile devices and ubiquitous Internet access. The boundaries between what is private and what is publicly accessible have been blurred with the transition of personal data to devices and services that are no longer under a user's direct control. Aversion to direct payment means that such services are often funded by data collection practices that only further complicate this issue.

We recommend that continued effort be placed on helping users to both understand the balance between these forces and enabling them in making this decision. For example, with respect to the tension between payment and privacy, it is worth noting that a strong technical solution for privacy considerations already exists in the form of end-to-end encryption. However, this is not a straightforward fix—there are additional concerns introduced by this potential solution, namely: usability (end-to-end encryption is notoriously difficult for novice users to use), trust (knowing whether or not the software is really doing what you expect), and payment (since encryption would prevent the online behavioral advertising techniques which are currently used to fund free services).

References

[1] Barreau, D., and Nardi, B. A. Finding and reminding: file organization from the desktop. *ACM SIGCHI Bulletin 27*, 3 (1995), 39–43.

[2] BBC. Celebgate hack: Man to plead guilty to nude photos theft. `http://www.bbc.com/news/technology-35817814/`, March 2016. Online; accessed 01-May-2016.

[3] Bergman, O., Whittaker, S., Sanderson, M., Nachmias, R., and Ramamoorthy, A. How do we find personal files?: the effect of os, presentation & depth on file navigation. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM (2012), 2977–2980.

[4] Bernstein, M., Van Kleek, M., Karger, D., and Schraefel, M. Information scraps: How and why information eludes our personal information management tools. *ACM Transactions on Information Systems (TOIS) 26*, 4 (2008), 24.

[5] Bursztein, E., Benko, B., Margolis, D., Pietraszek, T., Archer, A., Aquino, A., Pitsillidis, A., and Savage, S. Handcrafted fraud and extortion: Manual account hijacking in the wild. In *Proceedings of the 2014 ACM Conference on Internet Measurement Conference*, ACM (2014), 347–358.

[6] Ion, I., Sachdeva, N., Kumaraguru, P., and Čapkun, S. Home is safer than the cloud!: privacy concerns for consumer cloud storage. In *Proceedings of the Seventh Symposium on Usable Privacy and Security*, ACM (2011), 13.

[7] Ipeirotis, P. Demographics of Mechanical Turk: Now live! (April 2015 edition). /urlhttp://www.behind-the-enemy-lines.com/2015/04/demographics-of-mechanical-turk-now.html.

[8] Ipeirotis, P. G. Demographics of Mechanical Turk. `https://archive.nyu.edu/handle/2451/29585`, 2010.

[9] Kang, R., Dabbish, L., Fruchter, N., and Kiesler, S. my data just goes everywhere: user mental models of the internet and implications for privacy and security. In *Proceedings of the Eleventh Symposium on Usable Privacy and Security* (2015), 39–52.

[10] Koehler, P., Anandasivam, A., Dan, M., and Weinhardt, C. Cloud services from a consumer perspective. In *16th Americas Conference on Information Systems, AMCIS 2010*, AIS (2010), 329.

[11] Li, Y., and Chang, K.-c. A study on user acceptance of cloud computing: A multi-theoretical perspective. *Procedia Technology 16*, 1 (2014), 85–93.

[12] Malone, T. W. How do people organize their desks?: Implications for the design of office information systems. *ACM Transactions on Information Systems (TOIS) 1*, 1 (1983), 99–112.

[13] Odom, W., Sellen, A., Harper, R., and Thereska, E. Lost in translation: understanding the possession of digital things in the cloud. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM (2012), 781–790.

[14] Park, S. C., and Ryoo, S. Y. An empirical investigation of end-users' switching toward cloud computing: a two factor theory perspective. *Computers in Human Behavior 29*, 1 (2013), 160–170.

[15] Poon, C. S., Koehler, D. J., and Buehler, R. On the psychology of self-prediction: Consideration of situational barriers to intended actions. *Judgment and Decision Making 9*, 3 (2014), 207.

[16] Shin, D. Beyond user experience of cloud service: Implication for value sensitive approach. *Telematics and Informatics 32*, 1 (2015), 33–44.

[17] Smith, C. Why facebook and google mine your data, and why there's nothing you can do to stop it. `http://bgr.com/2016/02/11/why-facebook-and-google-mine-your-data-and-why-theres-nothing-you-can-do-to-stop-it/`, February 2016. Online; accessed 01-May-2016.

[18] Ur, B., Leon, P. G., Cranor, L. F., Shay, R., and Wang, Y. Smart, useful, scary, creepy: perceptions of online behavioral advertising. In *Proceedings of the Eighth Symposium on Usable Privacy and Security*, ACM (2012), 4.

Appendix

Demographics

Age: What is your age?

- 17 and under
- 18 to 24
- 25 to 34
- 35 to 45
- 46 to 64
- 65 and over
- Prefer not to answer

Gender: Are you male or female?

- Male
- Female
- Prefer not to answer

Education: What is the highest level of schooling you have completed?

- None
- Primary/Grade school
- Some high school, no diploma
- High school graduate: diploma or equivalent, e.g. GED
- Some college, no diploma
- Associate's or technical degree
- Bachelor's degree
- Graduate/Professional degree
- Prefer not to answer

Technological familiarity: on a scale from 1-5, how would you rate your technical proficiency?

Services

This survey - and all following questions - have to do with how you use **cloud storage services**. Generally, cloud services store data for you on machines owed by the service provider, rather than storing the data on your local machine.

For example, Gmail stores your email on machines operated by Google, not your machine. This provides convenience, so that you can access your email from any computer that is connected to the Internet.

You can think of cloud storage services as the digital equivalent of a storage unit rental service. The physical environment - and all maintenance - is handled by the provider of the service, while you are allocated some space where you can store your data.

Which of the following cloud storage services do you use? (Mark all that apply.)

- [ ] iCloud
- [ ] Google Drive
- [ ] OneDrive
- [ ] Youtube (if you upload, and not just watch, video)
- [ ] Facebook (if you upload photos)
- [ ] Picasaweb
- [ ] Flickr
- [ ] Dropbox
- [ ] Other (please specify)

What is it about cloud storage services that appeals to you? (Mark all that apply.)

- [ ] I don't need to worry about hardware failures (for example, a hard drive dying)
- [ ] They make sharing files easy
- [ ] They make accessing my files from different devices easier
- [ ] Automated file backup
- [ ] Other (please specify)

If you use multiple cloud storage services, what benefits do you get from using more than one service instead of just one?

If you store files in multiple cloud services, how do you decide which files are stored in which service?

Dropbox keeps a local copy of your files on your devices in addition to a "master" copy that is kept in the cloud. This means that you have a copy of your files that can be reached without Internet access as well as a copy that can be reached by devices with an Internet connection. Is this behavior something you'd like to see in all cloud services? (Please explain why.)

- Yes
- No

What do you care about most with a storage service? Please rank how important the following attributes are to you from most important to least important. (Click the item name and drag it where you feel it should go in the list.)

- Reliability: the service should never go down
- Privacy/security: only the people I allow should be to access the files I store or share
- Ease of use
- Accessibility: I want to be able to reach my files no matter where I am or what device I might be using
- Cost
- Other (please specify)

What are your largest concerns about using storage services? Please rank how concerned you are about the following items from most worried to least worried.

(Click the item name and drag it where you feel it should go in the list.)

- Privacy/security - are my files secure?
- Accessibility - can I access the service from any device?
- Permanence - what happens to my files if the storage company goes out of business?
- Cost - is it too expensive to store the files I want stored?
- Reliability - will the service fail when I need it?
- Other (please specify)
- Performance - is the service too slow for me to use?

Do you now pay or have you ever previously paid for cloud storage? If yes, please describe what the payment was for (for example, increased storage space).

- Yes
- No

Could you see yourself ever paying for a cloud storage service? If so, what would they need to do for you to make the investment worth it to you personally?

- Yes
- No

Privacy, security, and permanence

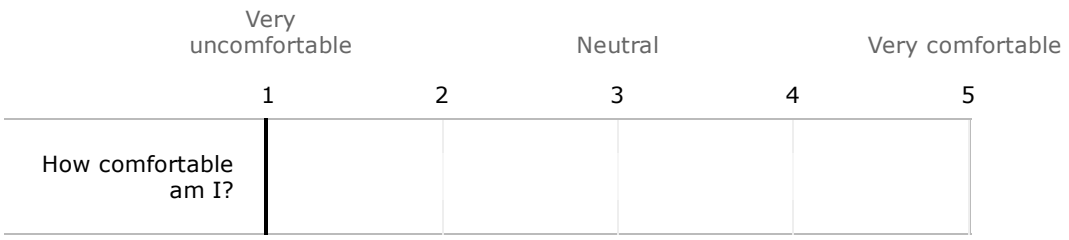
Is there data that you would *never* feel comfortable storing in the cloud? If so, what types of data?

- Yes
- No

Have you ever considered the event wherein your cloud provider goes out of business or otherwise terminates the service? Do you know what would happen to your data if this occurred?

As you understand things, which parties are able to view the data that you store in the cloud?

How comfortable are you with your cloud service provider using your data in ways that don't directly relate to the service itself (for example, to serve targeted ads)? Please rate how you feel from 1 to 5, where 1 signifies that you are very uncomfortable with such behavior and 5 signifies that you are very comfortable with such behavior.



Do you feel safe storing data in the cloud? What does "safe" mean to you in this context? Please explain.

Sharing and moving files

When you want to share files with others - friends, coworkers, etc. - how do you do so?

- [ ] Email attachment
- [ ] URL (link) to an item in cloud storage
- [ ] USB flash drive
- [ ] Burned a CD/DVD
- [ ]

Other (specify)

Which methods do you use the most? Please rank the items from most frequently used to least frequently used. (Click the item name and drag it where you think it should go in the list.)

- USB flash drive
- URL (link) to an item in cloud storage
- Email attachment
- Burned a CD/DVD
- Other (specify)

How do you decide when to use each of these methods?

(i.e. If you email a file to someone else, how did you decide that email was the best method in that case? Similarly, if you used a USB flash drive, how did you determine that that was the preferable option? etc.)

When you need to access your own files when on different devices, how do you do so? The same ways you share files with others?