"I Saw Images I Didn't Even Know I Had"

Understanding User Perceptions of Cloud Storage Privacy

Jason W. Clark[†] jclarks@masonlive.edu
Peter Snyder[*] psnyde2@uic.edu
Damon McCoy[†] mccoy@cs.gmu.edu
Chris Kanich[*] ckanich@uic.edu

[†]Department of Computer Science
George Mason University

[*]Department of Computer Science
University of Illinois at Chicago

Abstract

More than a billion people use cloud-based storage for personal files. While many are likely aware of the extent to which they store information in the cloud, it is unclear whether users are fully aware of what they are storing online. We recruited 30 research subjects from Craigslist to investigate how users interact with and understand the privacy issues of cloud storage. We studied this phenomenon through surveys, an interview, and custom software which lets users see and delete their photos stored in the cloud. We found that a majority of users stored private photos in the cloud that they did not intend to upload, and a large portion also chose to permanently delete some of the offending images. We believe our study highlights a mismatch between user expectation and reality. As cloud storage is plentiful and ubiquitous, effective tools for enabling risk self-assessment are necessary to protect users' privacy.

Author Keywords

Cloud; Privacy; Security; Threat Modeling

ACM Classification Keywords

H.5.m. Information Interfaces and Presentation (e.g. HCI): Miscellaneous

INTRODUCTION

Cloud storage is immensely popular. The ubiquity of Internet connectivity, the low price of high density storage, and the efficiency of modern datacenters enables companies like Google, Apple, and Dropbox to provide multiple gigabytes of resilient, high speed, globally accessible storage without charging the user for the privilege.

The cloud has revolutionized services like email, web hosting, and file backup. While copious free cloud storage has no doubt greatly improved overall user experience, the risks of having multiple copies of files stored under different user agreements with different security safeguards introduces the possibility of greatly increased risks for the user.

Automatic storage greatly compounds the risk of exposure - a compromise at any one storage provider, or on the device itself, can leak private photos to unauthorized users. It is also possible that users are not even aware that their images are being saved to cloud services. The composition of multiple apps and services makes the situation even more insidious, as it creates a mismatch between a user's intent and the result. For instance, if a user takes a compromising photo, emails it to a friend, and then later deletes it from his camera, it is possible that even if the recipient deletes the email and the sender deletes the image, the image could still persist in automatic cloud-based photo backup services (e.g. Dropbox or iCloud), the "Sent Messages" folder of the sender, and even the local device memory of the recipient (and thus, transitively, the recipient's own cloud-based photo backup services).

Under normal circumstances, these solutions add a layer of convenience and resiliency for the user. However, when stored in the cloud, private files are often visible to anyone with the proper credentials or the ability to reset the account's password. The Apple iCloud breach of August 2014 which exposed several celebrities' private images is one high profile example of this risk. While these victims were high profile celebrities, phenomena like revenge porn and sextortion bring this risk to many other individuals. Being able to control the existence and proliferation of private images is an absolute necessity in our hyper-connected world.

As handheld, Internet-connected cameras with automatic cloud backup are ubiquitous, it is imperative that users understand how to fully control all copies of the images that they create, share, and save. The first step to empowering users is to understand their mental models and expectations related to cloud storage.

This paper reports the results of an exploratory study that evaluates participants' security postures toward and mental models of cloud based storage. Through the use of custom built software, we are also able to report on participants' expectations related to specific photographs stored within their cloud storage accounts. We find that 16 out of 30 participants' were unaware of private images that were stored in their cloud storage based webmail accounts. The results show that their understanding of threats of how images might become public are realistic. However, they largely perceive the consequences of their private images becoming public to be embarrassment and are less aware and concerned of extortion. Once they used our automated image audit software on their webmail accounts, 11 out of 30 participants chose to delete stored private images to protect themselves from these images becoming public.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
CHI 2015, April 18–23, 2015, Seoul, Republic of Korea.
Copyright © 2015 ACM 978-1-4503-3145-6/15/04 ...$15.00.
http://dx.doi.org/10.1145/2702123.2702535

METHODOLOGY

Between the two surveys, participants used our 'Sensitive Image Audit' tool (or SIA tool), which we designed and implemented, to visualize and explore the images stored in their e-mail account.

SIA tool

The SIA tool works by searching users' gmail account for all images attached to messages. It then saves those images to the local disk. The user is then prompted to delete any images they wish to remove from their local disk. The SIA tool then finds which images have been deleted from the disk, and modifies the related messages on the gmail server to remove those attachments. This leaves the original version of the message in the users Gmail account with the attachment gone.

Participants

We recruited 30 participants from Craigslist to get a more diverse sample compared to posting fliers around campus. Each participant was offered $60.00 USD in compensation for their time and effort. The participants needed to be at least 18 years old and regular Google Mail users. The subjects' average age was 28, 50% were African American, 20% of Asian descent, 20% Caucasian, 7% Hispanic, and 3% other. 90% had at least some college and 50% had completed at least a 4 year degree.

Participants were heavy Internet users spending an average of 7.6 hours online. All of the participants accessed their email from home and most from work. Our pool of participants were also younger and more educated than the overall population demographics of the United States.

Surveys

All of the surveys were given through SurveyMonkey. We developed two surveys that were presented to the participants at different stages during the experiment. The initial survey (S1) consisted of 55 questions related to demographics, Internet use, and privacy strategies. The second survey (S2) asked 23 questions after the participant's use of the SIA tool as well as specific image related questions. While questions were asked about a broad range of subjects, questions relevant to this paper include: S2-2: Are there images saved in your email that you did not realize was there and that you would not want made public? S2-10: Are there images you deleted? S2-17: Why did you want to delete the image? What threat did it pose?.

Approach

The study was approved by our Institutional Review Board. Each of the 30 participants were provided with a private and secured office room. They were given access to a laptop running a virtual machine with the aforementioned software installed and the links to the surveys pre-loaded. After each participant completed their tasks we reset the virtual machine to ensure that no information from the previous participant was left on the laptop.

The participants' first task was to complete the preliminary demographics based survey. Upon completing this survey, the participants ran the SIA tool. At this point, the participants were given the option to deleted images if they so desired. Next, the participants completed a second survey associated with how they protect and share their images. Lastly, the participants completed a brief exit interview where we discussed the results of the SIA tool and asked more open ended questions that were not found on any of the surveys. These interviews were recorded with the consent of the participants and transcribed.

Bias and Limitations

One limitation is that our study only considered participants with Google Mail accounts, which might represent a biased group of users. Another potential source of bias are priming effects due to the order in which our survey questions were asked. We attempted to limit such bias in our survey design, but found afterward that some issues were missed. For example, question S2-8: What threat (source) did you see, that caused you to not to want the image made public? and S2-9: What consequences do you see if your image was made public? These questions were asked after priming questions such as S1-49 through S1-53, which all ask for a self evaluation of the user's concern with respect to several different Internet threats. For this reason, we focus on factual responses as well as relative popularity of answers where the bias is likely constant across all possible answers.

The full list of questions asked during this study can be found in the online appendix in the ACM digital library.

RESULTS

In this section, we describe three main themes regarding our research subjects' use of cloud storage: subjects' understanding and concerns regarding potential security breaches, subjects' understanding and concerns regarding actions that can increase risk of adverse consequences, and subjects' behavior and willingness to take action to mitigate or defend against these threats. We combine results from the software based security audits, individual subjects' quotes, and aggregate response metrics to support describe each theme.

For brevity's sake, we make the distinction between *threats*, which we define as the perpetrators and the potential harmful outcomes like stolen data or financial fraud, and *risks*, which we define as user behaviors which make harmful outcomes more likely or more damaging.

Understanding Threats

Security breaches are potential threats to users' well being. Here we describe both users conceptualization of who or what these threats are with respect to cloud storage, and what aspects of these threats are most important or concerning to the users.

Possible adversaries

While priming effects invalidate these questions' utility for judging how concerned users are about adversaries, the relative popularity of different adversaries is interesting: unintentional sharing by oneself was the most common threat (19/30) compared to cybercriminals (18/30) or acquaintances (15/30),

perhaps indicating an apprehension regarding their own understanding of technological functionalities like automatic cloud uploads.

Furthermore, when asked what their primary threat of account break-in was during the exit interview, 23 respondents mentioned strangers (e.g. cyber hackers,"people out of foreign countries," "person who just wants to use my credit card"), while only three mentioned acquaintances in some capacity. This result indicates the subjects' mental models are focused mainly on global rather than local threats like inadvertent sharing by oneself or break-ins by acquaintances.

Possible Ramifications of a breach

The main ramification people are concerned with regarding stolen photos is embarrassment - 25/30 respondents mentioned embarrassment as a consequence, while only 8/30 mentioned being concerned regarding extortion and a different (but not distinct) 8/30 were concerned with stolen photos incriminating them.

Understanding Risks

Task-oriented users rarely stop to think about the security implications of their actions. However, characterizing how users conceptualize the risks involved with how they use cloud storage is an important aspect of understanding how they interact with these services.

Private Content

The main focus of our study is private photos - 16/30 respondents said that they have images saved in their email that they did not realize were there and that they would not want made public. To gauge how well users understand the risk of online account compromise, our survey asked subjects about their password hygiene habits. 14/30 respondents admitted to sharing their passwords with other people, and 9/30 admitted to saving passwords in their email. This latter action can lead to privilege escalation for attackers who gain access to an email account and then can gain access to other more sensitive accounts like brokerage or banking accounts.

Passive Storage

Several actions taken by software on behalf of users are automatic, passive, and/or opt-out. For instance, applications like Dropbox, Google Drive, and iCloud all encourage users to back up all photos taken on a mobile device automatically. Another example is the "Sent Mail" folder in a gmail account: even though some users' mental models understand email as a communication medium rather than a storage medium [6], a copy of every mail ever sent is stored in the web-based cloud storage of several different kinds of email accounts. This is often not immediately apparent to the end user. When this feature is combined with the use of mobile phone cameras to take private photos, a very dangerous interaction happens: users do not realize they are saving these private photos to the cloud, and even if they specifically intend to delete a photo, it might still exist on their device or in the cloud. This sentiment was expressed best by participant 5: "I saw images I didn't even know I had."

Understanding Countermeasures

Users exhibit a wide range of behaviors intended to limit the possibility of harmful outcomes. Here we characterize both the ways that users mitigate the potential for harm currently, as well as with the assistance of our privacy audit tools.

Current Strategies for Privacy Protection

The easiest and most effective defense against becoming a victim of photo theft is simply to not take or save private photos. We asked two questions of users regarding their photo protection strategies: one in the initial interview regarding how they share photos online, and one in reference to individual photos after the users had used our redaction tool.

With respect to possessing and sharing photos, several users exhibit thoughtful strategies for minimizing their risk of unwanted exposure. For photos that they are not willing to upload to the cloud, most subjects store the photos locally, on a hard drive or removable media (18/30), and several (11/30) choose not to store the photos at all and delete them instead. For photos that they do wish to share with others, subjects largely rely on service-provided privacy settings (22/30) and password protection (17/30) to limit photos' audiences.

Tool-assisted Strategies

The primary component of our user study was to instruct research subjects on the use of SIA tool and record their discoveries and reactions. After using the tool, 16/30 subjects reported finding pictures that they both did not realize were there and would not want to be made public. While we do not claim to have a representative sample, this initial finding suggests that users' expectations regarding cloud storage of personal photos, perhaps the most intimate medium available, are strongly at odds with reality either due to passive storage, information overload, or other factors. Participant 5 summed up the sentiment that removing these images has become increasingly difficult: "Now you've got to look at probably 15 different places if you want to get rid of that photo."

Not only were several subjects surprised to find unwanted photos stored in their accounts; 11/30 subjects chose to permanently delete those photos during their use of our tool. Most users were concerned both regarding pictures of one's self (9/11) and of friends (7/11). This finding echoes that of Klasnja et al.'s study of security concerns when on Wi-Fi, where subjects showed a strong concern for the security and privacy of their communication partners, sometimes even more than themselves [5].

Subjects indicated that they rely heavily on social bonds for defense; 22 of the 30 subjects said they ask others not to share private photos as a means of defending their privacy. When asked whether they would use this tool again if it were available, 23/30 answered in the affirmative. While the proximity to the revelation of private images in the account and the overall goal of the study likely biases this survey question, 4/30 subjects independently inquired during the exit interview about the availability of this software for further use.

RELATED WORK

While our study focuses on the specific domain of image security, several other research efforts have explored users' understanding of cloud computing based security and privacy.

Wang et al. [8] explored private content exposure concerns for users and their Facebook posts (including images) which they made but then regretted. Garg et al. [2] explore the reason for privacy failures and discover that privacy behaviors can be explained by risk perception, control usability, or privacy preferences, sometimes even by the same actor.

Ion et al. [3] investigate privacy concerns for users in the space of consumer cloud storage, and find that users prefer local storage for sensitive documents, which was not the case in our results where a large percentage of users had unintentionally stored sensitive images online. Odom et al. [6] studied how young people value and form attachments to virtual possessions with the goal of comparing them to their physical counterparts. They found that an increasing portion of users' possessions are digital, and that many used email as a way of moving "digital assets" between accounts.

The work of Klasnja et al. [5] presented an exploratory study of how users understand Wi-Fi and the associated privacy risks. They show that while users are aware of "expert hackers," they are less aware of the immediate risk of eavesdropping on unencrypted communications. The authors posit that end-user awareness tools and infrastructural improvements are necessary to address privacy and security problems with Wi-Fi use.

Multiple tools exist to improve users' cloud security. Egelman [1] performed a laboratory experiment to study the privacy tradeoff offered by Facebook Connect: they observed that most users understood the privacy convenience tradeoff and thus consciously chose to forfeit privacy for convenience. Kelley et al. [4] used the concept of a nutrition label as their inspiration for implementing a similar label for privacy. Snyder and Kanich [7], designed a system called "Cloudsweeper" which gives users the opportunity to encrypt or redact sensitive, unexpected, and rarely used information to mitigate the risks of cloud storage accounts without sacrificing the benefits of clouds storage or computation.

FUTURE WORK & CONCLUSIONS

Our initial prototype SIA tool was well received and many of the participants expressed interest in using this tool again. This suggests that users want to know when they have unintended images stored and that providing them with tools could be an effective strategy for enabling them to remove images that were unintentionally stored. To address this gap, we are working on a powerful and flexible image auditing tool that can interoperate with most cloud storage providers, including backup, online social networking and webmail. The goal of this tool is to mitigate the problem of unintended image storage by allowing users to audit their images and make changes to a single cloud storage account or global delete and update images across all accounts in one place.

We have explored how our participants perceive the existence of sensitive images in their cloud email accounts and the privacy threats associated with storing these images in cloud storage. By allowing users to audit their stored images, we were able to discover that there is both a lack of awareness of this storage and a lack of desire to maintain copies of these images in this way. Furthermore, several participants deleted these images as soon as they were revealed by our tool. Clearly, more is being done on users' behalf than they necessarily desire. Cloud based storage providers must find a way to balance automatic features with the possibility that those features are potentially harmful to their users.

ACKNOWLEDGMENTS

We thank our study participants and the anonymous reviewers for their insightful comments. This work was made possible by National Science Foundation grant CNS 1351058 and a gift from Google.

REFERENCES

1. Egelman, S. My profile is my password, verify me!: the privacy/convenience tradeoff of facebook connect. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM (2013), 2369–2378.
2. Garg, V., Benton, K., and Camp, L. J. The privacy paradox: A facebook case study (2014).
3. 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.
4. Kelley, P. G., Bresee, J., Cranor, L. F., and Reeder, R. W. A nutrition label for privacy. In *Proceedings of the 5th Symposium on Usable Privacy and Security*, ACM (2009), 4.
5. Klasnja, P., Consolvo, S., Jung, J., Greenstein, B. M., LeGrand, L., Powledge, P., and Wetherall, D. When i am on wi-fi, i am fearless: privacy concerns & practices in eeryday wi-fi use. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM (2009), 1993–2002.
6. Odom, W., Zimmerman, J., and Forlizzi, J. Teenagers and their virtual possessions: design opportunities and issues. In *Proceedings of the SIGCHI conference on Human Factors in computing systems*, ACM (2011), 1491–1500.
7. Snyder, P., and Kanich, C. Cloudsweeper: enabling data-centric document management for secure cloud archives. In *Proceedings of the 2013 ACM workshop on Cloud computing security workshop*, ACM (2013), 47–54.
8. Wang, Y., Norcie, G., Komanduri, S., Acquisti, A., Leon, P. G., and Cranor, L. F. I regretted the minute i pressed share: A qualitative study of regrets on facebook. In *Proceedings of the Seventh Symposium on Usable Privacy and Security*, ACM (2011), 10.