Assured Deletion in the Cloud

Andrea Compton[1]
12/10/14

[1] Mentor: Ming Chow, Tufts University

Abstract

As more people and companies become paperless they are relying on a new way to store information, the cloud. The cloud is a storage system that many people trust yet don't know how to control. Previously, anyone could use a hard drive and be able to at least turn it off if they felt something was being compromised. Now, the cloud can't be turned off on a whim, and many people trust their provider to keep it secure. This security is needed to protect their credentials and their data while on the network. A main issue in cloud security is deletion. When a user deletes a file how can they be sure it was actually deleted? This paper provides a background on the cloud, how files are stored and deleted, and discusses current and possible future approaches used by cloud services providers to assure customers their files were deleted adequately. In an ever-changing world, security must be kept up to date as providers attempt to satisfy customers' needs.

1 Introduction

Cloud computing originated from a need for large computational power and storage space.[1] According to the United States National Institute of Standards and Technology (NIST):

> Cloud computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction.[5]

Today, it is not only used by companies, but by individuals who are drawn to it for its benefits and ease of use. Many emerging companies such as Dropbox, Box, and SugarSync are popular options for fast and easy storage. While most cloud services offer elasticity, pay-per use options, increased storage, reduced cost, and are highly automated, security is a vital concern.[4] Many CEO's have stated that even though the cloud decreases business risk and cost savings, they have held off transferring their data because of the lack of security.[6]

2 To the Community

In this technological age we are moving away from direct storage to an easier, less painful way to ensure data integrity. Many people use cloud services to share resources and backup data because of its ease of use, without thinking about the potential security risks. Even popular mobile devices, such as Apple's iPhone, are seamlessly transitioning to the cloud. With the new update, users can place their address book, photos, and other information in Apple's iCloud, which can then be accessed from anywhere if the same

AppleID is given. This is great if you ever need to restore your phone or share pictures with friends and family; however, protecting this data is now out of your control. You have entered a trust relationship with Apple and relinquish the power to physically destroy your data. In particular, researchers have found that deleted data can be recovered from smartphones. Images, audio files, PDFs, and Word documents previously deleted from common vendors Dropbox, Box, and SugarSync, were recovered using both an HTC Android smartphone and an iPhone.[7] This paper will discuss how data is deleted in the cloud and different ways to validate its deletion.

3 Defenses

3.1 Background

When you upload your data to the cloud it is stored remotely, in an unknown server farm, alongside the data of many other customers of the same provider. Data is stored on the servers in virtual machines unique to each user and thus "cloud computing can be thought of as an evolution of outsourcing, where an organization's business processes or infrastructures are contracted out to a different provider."[6] Customers are unaware of the location of their data, as it is usually outsourced, and companies such as Google and Amazon keep their locations private.[6] In addition, cloud services providers create multiple copies of your data which are then placed on different servers, protecting against data loss yet resulting in problems assuring data deletion.

3.2 Privacy

In order to trust your cloud provider you need to know how your data is being secured. The cloud is run on a network of servers, leaving it vulnerable to network attacks including denial of service, man in the middle, network sniffing, and port scanning to name a few. This allows attackers to gain access to sensitive data if it is left unencrypted.[4] Thus trust in your provider is critical. Transferring data from virtual machines to physical machines must be secure and requires data encryption, enforcing appropriate data sharing policies and maintaining secure resource allocation and memory management algorithms.[8] If a trustworthy provider is chosen, your data may be sufficiently secure; however, your data is now completely in the hands of your cloud services provider. One option for enhanced security is to use a private cloud. While more expensive, private clouds come with many security benefits. In the Amazon Virtual Private Cloud enterprises can "connect their existing infrastructure to a set of isolated Amazon Web Services (AWS) compute resources via a virtual private network connection and extend their existing management capabilities such as security services, firewalls, and intrusion detection systems to include their AWS resources."[4] A crucial privacy issue is data deletion. When users specify they want data removed they should be assured that all copies of the data were removed completely to prevent malicious users from retrieving it in the future.

3.3 Data Deletion

The U.S. Department of Defense's definition of data deletion states two of the key concepts regarding effective deletion:

a. Clearing. Clearing is the process of eradicating the data on media before reusing the media in an environment that provides an acceptable level of protection for the data that was on the media before clearing. All internal memory, buffer, or other reusable memory shall be cleared to effectively deny access to previously stored information.

b. Sanitization. Sanitization is the process of removing the data from media before reusing the media in an environment that does not provide an acceptable level of protection for the data that was in the media before sanitizing. IS resources shall be sanitized before they are released from classified information controls or released for use at a lower classification level.[3]

Unfortunately, data is usually not deleted according to these specifications, allowing it to be accessed by archives, backups, or restorations.[11] This is exacerbated in the cloud as extra copies of data are stored, and the disk containing your data also holds data from other clients. Even if deletion is requested, it may not occur instantaneously. Many companies use garbage collection, a form of automatic memory management, which erases or overwrites the data sometimes months later. This allows users to recover deleted data for a period of time if the delete was accidental. If customers have sensitive data they want to adequately delete, providers offer encryption. Due do security concerns, providers now encrypt the data before it leaves the user's data center. The encryption keys are "maintained at the edge, on customer's premises," guaranteeing that when the key is deleted, nobody can read the data if it's adequately deleted or not.[9] This technique, crypto-shredding, provides one way to guarantee data will not be obtainable by a third party.

3.4 Assured Data Deletion

Many companies, including hospitals, banks, and law firms, must be guaranteed data was deleted adequately. Here I will present two ways to assure data deletion. The first is time-based file assured deletion where files are deleted and remain permanently inaccessible after a specified duration. To make sure all copies are unattainable after deletion, the owner first encrypts the files with a data key. A separate key manager then encrypts this data key with a control key. This control key is time-based and the key manager will remove the key at the specified time declared when the file was encrypted. Without the control key, both the file and data key remain encrypted, and are thus believed to be adequately deleted even if expired copies are not fully removed. The second approach, policy-based file assured deletion, is similar to time-based, but instead of associating files with a specified time, they are associated with a file access policy. A file access policy is "a Boolean combination of atomic policies."[10] Control keys are associated with each atomic policy and thus when policies are revoked, the key manager removes their control keys. If we assume a file is associated with one policy, then when the policy becomes revoked and no longer holds, the control key will be removed and the file will become inaccessible. For instance if a company declares a user-based policy stating that person P is an employee, all of P's files can be associated with this policy. Then if person P quits the company, the control key for this policy, and all of P's files, will be deleted. Resulting in no obtainable access to the cloud files associated with that control key. This method, known as FADE, is a secure overlay cloud storage system providing access control and assured deletion, yet has only been implemented as a prototype as performance and security tradeoffs are still being measured. Researchers have been able to show that,

while performance slows with encryption, “the performance overhead of FADE becomes less significant when the size of the actual data file content increases (e.g., on the order of megabytes or even bigger). Thus, FADE is more suitable for enterprises that need to archive large files with a substantial amount of data.”[10] Assured deletion in the cloud is not easy to obtain. Requiring encryption, and sometimes double encryption, is costly and may decrease performance, yet researchers are striving to find a solution that assures users their files are adequately deleted and convinces them to move to the cloud.

3.5 Extension: Utilizing AWS CloudTrail Logs

While assured deletion is not yet feasible, many cloud services providers offer services to better manage your files. Amazon Web Services (AWS) released CloudTrail, a web service that tracks the API calls from your AWS account and publishes the resulting log files.[2] In addition, Amazon offers a variety of CloudTrail partners, each providing analytics tools to ease the use of CloudTrail logs. To demonstrate how easily log files can be analyzed according to company needs, a java script was written to accompany this paper. Amazon provides starting scripts and ideas with supporting documentation to simplify the process. The java script looks at the JSON format provided by Amazon and takes in the user and region of each API call, which can easily be extracted by Amazon’s given scripts. A map is then created, noting where each call took place and displaying the user name, if prompted, allowing the user to visualize the data received.

4 Conclusion

Storing data in the cloud has many benefits. Used by companies and individuals, files can be accessed from anywhere and can still be retrieved upon computer failure. All of these benefits, however, are still not enough to convince everyone to transfer their data. Security is a main concern as customers relinquish complete control and must trust their providers to maintain data integrity. Adequate file deletion is critical to gaining users trust. Many companies, including AWS, offer services to ease the use of provided log files, which give users every action associated with their files. This is a great way to obtain trust from customers even though it does not assure adequate deletion. As cloud services are relatively new, researchers are still working on ways to obtain complete deletion without compromising the alluring benefits offered. Crypto-shredding, time-based file assured deletion, and policy-based file assured deletion are intriguing solutions the problem and may emerge in the near future. For now, if you want to use cloud services you should research your provider and gain sufficient trust in their system. Writing a few simple scripts and utilizing secondary services offered are great ways to keep track of your data and maintain a suitable level of accountability. While adequate data deletion may not be readily available today, if you do your research and keep track of your data's movements, the cloud is a great option that will only continue to improve.

References

[1] Alghazzawi, Daniyal M., and Syed Hamid Hasan. "Security Issues and Challenges - Cloud Computing." *International Journal of Computer Science Issues (IJCSI)* 10.5 (2013): 148-54. *ProQuest.* Web. 12 Dec. 2014.

[2] "AWS CloudTrail - Capture AWS API Activity." *AWS CloudTrail - Capture AWS API Activity*. Amazon Web Services Blog, 13 Nov. 2013. Web. 12 Dec. 2014.

[3] DoD 5220.22-M. National Industrial Security Program Operating Manual, United States Department of Defense; 2006.

[4] Jamil, Danish, and Hassan Zaki. "Cloud computing security." *International Journal of Engineering Science and Technology* 3.4 (2011): 3478-3483.

[5] Mell, Peter, and Tim Grance. "The NIST definition of cloud computing." (2011).

[6] Pearson, Siani, and George Yee. *Privacy and security for cloud computing*. Heidelberg, Germany: Springer, 2013.

[7] Samson, Ted. "Deleted Cloud Files Can Be Recovered from Smartphones, Researchers Find." *InfoWorld*. N.p., 19 Mar. 2013. Web. 12 Dec. 2014.

[8] Sathyanarayana, T. V., and L. Sheela. "Data security in cloud computing."*Green Computing, Communication and Conservation of Energy (ICGCE), 2013 International Conference on*. IEEE, 2013.

[9] Slack, Eric. "How Do You Know That “Delete” Means Delete in Cloud Storage?" *Storage Switzerland*. N.p., 16 Aug. 2011. Web. 12 Dec. 2014. <www.storage-switzerland.com%2Farticles%2Fentries%2F2011%2F8%2F16_How_do_you_know_that_Delete_means_Delete_in_Cloud_Storage.html>.

[10] Tang, Yang, et al. "Secure overlay cloud storage with access control and assured deletion." *Dependable and Secure Computing, IEEE Transactions on*9.6 (2012): 903-916.

[11] Winkler, Vic JR. *Securing the cloud: cloud computer security techniques and tactics*. Elsevier, 2011.