Cloud Storage Assured Deletion: Considerations and Schemes

Richard Thames
Computer Science Department, St. Mary's University, San Antonio, Texas, USA
Email: rthames@mail.stmarytx.edu

Ayad Barsoum
Computer Science Department, St. Mary's University, San Antonio, Texas, USA
Email: abarsoum@stmarytx.edu

Abstract—**The assured deletion problem was realized with the introduction of cloud data storage. An exemplar of the broader set of cloud services, assured deletion is poorly understood by customers and complicates the work of forensic professionals. Over the last ten years, schemes that solve the assured deletion problem have been proposed. Proposed solutions have improved on each other to mitigate scaling overhead, trusted third parties, bottlenecks, single points of failure, and other inefficiencies. Cloud service providers have an opportunity to provide customers verifiable proof of deletion. In this work, we focus on the problem of how cloud data storage customers can be assured that when they attempt to delete data from the cloud, it is not retrievable.**

Index Terms— **Assured deletion; secure deletion; provable data deletion; cloud storage; cloud computing**

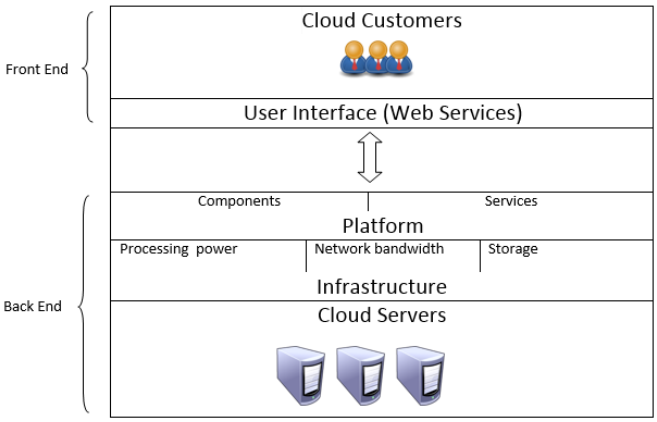

Figure 1. Conceptual framework for Cloud Computing architecture.

I. Introduction

NIST defines cloud computing as 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 [9].

Cloud computing represents the vision of providing computing services as public utilities like water and electricity. The architecture of cloud computing can be split in two: front-end and back-end. The front-end represents cloud customers, organizations, or applications, e.g. web browsers, that use the cloud services. The back-end is a huge network of data centers with many different applications, system programs, and data storage systems. It is *metaphorically* believed that, the Cloud Service Provider (CSP) has almost infinite computing power and storage capacity. A conceptual framework of cloud computing architecture is illustrated in Figure 1 with its two main parts.

Cloud services are offered by CSP and can be categorized into the service models: Infrastructure as a Service (IaaS), Platform as a Service (PaaS), and Software as a Service (SaaS). Cloud data storage services such as AWS S3, Google Cloud Storage, and Microsoft Azure Storage are examples of PaaS cloud services (Google Docs, Google Calendar, and Zoho Writer are known examples of SaaS). The NIST cloud computing definition fits cloud data storage services as they require minimal management effort, can be rapidly provisioned and released, and can be enabled ubiquitously, conveniently, and on-demand for network access.

However, while API's and management consoles abstract cloud data storage and retrieval into similar user paradigms as file systems and file managers do for data stored and retrieved on physical workstations, there are significant differences in between the two. Cloud data storage customers must consider the additional data integrity and security implications that come with storing data in the cloud. This paper focuses on the problem of how cloud data storage customers can be assured that when they attempt to delete data from the cloud, it is not retrievable. Data deletion from the cloud is a major challenge to ensure that the data has been actually deleted from the cloud servers after issuing a delete request to the CSP.

This work summarizes the academic literature on cloud storage deletion and how it (1) introduces risk that CSP customers poorly understand, (2) increases the complexity of law enforcement forensic activities, and (3) presents an

opportunity for CSP's to offer assured deletion as feature. Additionally, it follows the evolution of solutions for assured deletion of cloud-stored data that have evolved over the last ten years.

II. Cloud Storage Assured Deletion Considerations

Tanimoto et al. contended that users generally do not understand at all about how information is managed in the cloud environment [7], and a corollary to this is that they do not understand the risks that come with cloud data storage and assured deletion. Without a detailed understanding of their CSP's operation management methods, customers must trust the CSP that data is only accessible to designees of the customer, and that when data are deleted, the CSP is actually making it permanently unavailable. Tang et al. [6] indicated that keeping data permanently is undesirable, as data may be unexpectedly disclosed in the future due to malicious attacks on the cloud or careless management of cloud operators [6]. The customer must have a way of knowing if data can be made available to administrators, developers, law enforcement, or other third-parties.

Tanimoto et al. identified and enumerated risks inherent in cloud data storage—including risks related to cloud data deletion—that customers should understand and classify them in a Risk Breakdown Structure (RDB) with an appropriate mitigation. Of the twenty-three risks identified, two relate to assured deletion: (1) problem of removing data after using cloud service, and (2) problem of data deletion after cloud service use. The first risk has the Third Party Surveillance countermeasure specified as: The surveillance of data movement is requested of a third party. Cloud service provider is requested to move data [7]. The second risk has the combine with cloud service specification countermeasure specified as: Even when data cannot be deleted, the form of data is devised so that it may be uninfluential by using, for example, encryption[7].

While cloud data storage and assured deletion introduce new concerns for CSP customers, they also introduce new concerns for law enforcement and forensic experts. Reardon et al. said that data are securely deleted from a system if an adversary that is given some manner of access to the system is not able to recover the deleted data from the system [4]. They identified how unlinking at various data layers is the typical operation used when deleting data, and how this is different from secure data deletion. They surveyed the approaches that can be used to securely delete data from physical medium such that it is permanently irrecoverable. It is unlikely the secure data deletion schemes they identified are used by CSP's, but secure deletion of keys to encrypted data can be implemented such that it is functionally equivalent to secure deletion.

This technique is at the heart of many of the assured deletion schemes that will be explored shortly and is only one of the factors that increase the complexity of data acquisition and analysis for criminal investigators. Others include: decentralized data storage, data dependency chains, CSP dependence, jurisdiction and chain of custody variances, evidence segregation, and data retention. With respect to cloud data deletion, Pichan et al. indicated that the deleted data can be collected from the media using data carving methods supported by forensics tools. However, in case of cloud, the volatility and elasticity of cloud environments make it much harder to collect the deleted data [3]. Even if deleted data has been found in the cloud, attributing it to a specific user remains to be a big challenge due to the sheer volume of the data and amount of backup cloud provider would maintain [3].

III. Cloud Storage Assured Deletion Schemes

The properties of cloud services with respect to cloud data storage—minimal management effort, rapid provisioning and releasing, and ubiquitous, convenient, and on-demand network access—introduce new risks and complexities for a variety of stakeholders. Additionally, due to these properties of the cloud and how CSP's implement them on physical hardware at datacenters, assured deletion of cloud data fundamentally cannot be achieved by the same means of secure data deletion from physical devices as described by Reardon et al. New techniques to guarantee assured deletion of cloud data had to be created. The techniques surveyed in the academic literature, taken as a set over time, build on each other for the most part and represent an evolution of sorts. Each relies on data encryption and schemes to securely revoke decryption keys in place of deletion of data on physical drives. This is because the quantity and location of where data is stored for cloud implementations is unknown.

A.The Simplistic Approach

While not an actual technique in the academic literature due to high overhead, the fundamental way to achieve assured deletion of cloud data is to (1) encrypt the data such that it is computationally prohibitive to decrypt the data without the decryption key, then (2) store the encrypted data in the cloud while keeping the decryption key secure, and finally (3) securely delete the encryption key. Performing the third step in this sequence is equivalent to assured deletion of the data stored in the cloud. While this technique is straightforward, it is not practical for general use as it does not scale without significant effort and overhead. As Mo et al. noted on this scheme: If we use one key to encrypt all data, whenever we delete one data item, we have to re-encrypt all other data items with a new key because otherwise they would also become inaccessible after the old key is deleted. If we assign on key to each file, there will be numerous keys if the number of files is large. Moreover, even if we only want to delete on block in a file, we will have to retrieve the entire encrypted file from the server, decrypt it, delete that block, remove the old key, choose a new key, and re-encrypt the entire file [2].

B. Trusted Third-Party Auditors

To avoid the overhead, computation, and data transfer incurred in the previous scheme, Wang et al. proposed in

2009 the concept of a Third Party Auditor (TPA) that could reliably perform equivalent functions on behalf of the CSP customer. They defined TPA's as a party external to the CSP and CSP customer, which has expertise and capabilities that clients do not have, is trusted to access and expose risk of cloud storage services on behalf of the clients upon request [8]. Though they were primarily addressing a way to sample data upon request to guarantee it did exist intact, the scheme proposed by Wang et al. accounted for verification of dynamic data functions which included data creation, insertion, modification, and deletion. This scheme relied on the Merkle Hash Tree (MHT) to store verifiable data changes to cloud data, and a verified deletion function is provided in detail as the opposite of the insertion function.

C. Trusted Third-Party Auditors

In 2012, Tang et al. introduced a scheme for assured deletion called File Assured Deletion (FADE). This scheme did not depend on a third party for spot checks of data integrity upon request but rather created a trustworthy key management system that could serve as a trusted third party. The design intuition of FADE is to decouple the management of encrypted data and cryptographic keys, such that encrypted data remains on third-party (untrusted) cloud storage providers, while cryptographic keys are independently maintained and operated by a quorum of key managers that altogether form trustworthiness [6]. Essentially this scheme combines the simplistic approach with an access control layer that abstracts existing time-based file assured deletion techniques of decryption keys into policy-based controls for the keys. Implementing this access control layer and key provisioning, assignment, and deletion into a system of clients and key managers structured as trustworthy is how assured deletion was achieved.

D. Recursively Encrypted Red–Black Key Tree (RERK)

While FADE minimized the scalability issues of the simplistic approach and replaced the need for trusted TPA's, its system for key management and the involvement of managed keys in all data operations was architecturally a bottleneck and single point of failure. Thus in 2014, Mo et al. [2] presented a scheme which sought to address this. They explored the feasibility of permanently deleting data without involving a third party between clients and servers in a cloud system. They prevented any possibility for the cloud service providers or anyone who compromises the cloud servers to circumvent deletion or break data privacy. Their solution is based on a novel multi-layered key structure, called Recursively Encrypted Red–Black Key tree (RERK) that ensures no key material will be leaked [2]. In fact, this solution reduced assured deletion from a three-party problem to a two-party problem with properties: efficiency, integrity, correctness, and confidentiality.

E. Provable Data Overwriting

In 2016, Luo et al. [1] proposed another scheme which acknowledged the lineage of TPA's to FADE to RERK but is not a direct descendent of them. They identified that these previous models have three main deficiencies: (1) encryption occurs before data outsourcing, (2) encrypted data remains on the cloud server after the respective deletion operations occur, and (3) encryption makes cloud computation on outsourced data difficult. The solution they presented is overwriting cloud data in a predictable and provable way such that the data it replaces is functionally deleted. Their scheme disguises overwriting operations as data updating operations and then utilized the method of provable data possession (PDP) to audit the results of overwriting [1]. However, the scheme, while novel, requires conditions that cannot be guaranteed such as a CSP that only keeps the current version of the data, and consistency of all copies of data when updating.

IV. Summary and Conclusion

Assured deletion schemes have evolved over the last ten years as evidenced by the reviewed academic literature. Given the gap between simple management consoles and API's for cloud data storage and the complex hardware implementations by CSP's, the ability to perform assured deletion is likely unsettled. However, just as CSP's offer technologies that simplify storage operations, providing an assured deletion service could serve as a differentiator. This can be as a potential win-win that would benefit both CSP's and their customers. Sometimes such assurances are included in the contracts and service level agreements (SLA's) but they still require trusting the provider without any technical proof. Technical assurances—and proof deletion upon request—can give tenants confidence about how their outsourced data are handled and decommissioned. The requirements of such a service can be identified as: fine-grained, usability, cloud computation, complete deletion, timeliness, service availability, deletion of all backup copies, delete latency, error handling, and proof of deletion.

References

[1] Yuchuan Luo, Ming Xu, Shaojing Fu, and Dongsheng Wang. 2016. Enabling Assured Deletion in the Cloud Storage by Overwriting. In Proceedings of the 4th ACM International Workshop on Security in Cloud Computing (SCC '16). ACM, New York, NY, USA, 17-23.

[2] Z. Mo, Q. Xiao, Y. Zhou and S. Chen, "On Deletion of Outsourced Data in Cloud Computing," 2014 IEEE 7th International Conference on Cloud Computing, Anchorage, AK, 2014, pp. 344-351..

[3] Ameer Pichan, Mihai Lazarescu, Sie Teng Soh, Cloud forensics: Technical challenges, solutions and comparative analysis, Digital Investigation, Volume 13, 2015, pp. 38-57.

[4] J. Reardon, D. Basin and S. Capkun, "SoK: Secure Data Deletion," 2013 IEEE Symposium on Security and Privacy, Berkeley, CA, 2013, pp. 301-315.

[5] Kopo M. Ramokapane, Awais Rashid, and Jose M. Such. 2016. Assured Deletion in the Cloud: Requirements, Challenges and Future Directions. In Proceedings of the 2016 ACM on Cloud Computing Security Workshop (CCSW '16). ACM, New York, NY, USA, 97-108.

[6] Y. Tang, P. P. C. Lee, J. C. S. Lui and R. Perlman, "Secure Overlay Cloud Storage with Access Control and Assured Deletion," in IEEE Transactions on Dependable and Secure Computing, vol. 9, no. 6, pp. 903-916, Nov.-Dec. 2012..

[7] S. Tanimoto, M. Hiramoto, M. Iwashita, H. Sato and A. Kanai, "Risk Management on the Security Problem in Cloud Computing," 2011 First ACIS/JNU International Conference on Computers, Networks, Systems and Industrial Engineering, Jeju Island, 2011, pp. 147-152.

[8] Wang Q., Wang C., Li J., Ren K., Lou W. (2009) Enabling Public Verifiability and Data Dynamics for Storage Security in Cloud Computing. In: Backes M., Ning P. (eds) Computer Security – ESORICS 2009. ESORICS 2009. Lecture Notes in Computer Science, vol 5789. Springer, Berlin, Heidelberg.

[9] Peter Mell and Tim Grance. Draft NIST working definition of cloud computing. Online at http://csrc.nist.gov/groups/SNS/cloud-computing/index.html, 2009.