Data Analytics and Artificial Intelligence
Vol: 1(3), 2021
REST Publisher
ISBN: 978-81-948459-4-2
Website: http://restpublisher.com/book-series/daai/

Unraveling Secure Storage Protocols for Public Auditability and Robust Data Integrity

Bobby. S
St. Joseph's College of Arts and Science for Women, Hosur, Tamil Nadu, India.
Corresponding Author Email: angelbobby2@gmail.com

Abstract

Cloud computing is gaining huge popularity as an IT architecture. Cloud service providers offer many services based on cloud computing. A cloud storage service is a cloud service that can provide large-scale storage capacity to solve the storage capacity shortage of local end users. However, cloud storage services may provide data security because your data is not stored on its own storage. Service orientation, loose coupling, strong fault tolerance, business model, and ease of use are key characteristics of cloud computing. Although secure cloud storage has only recently been proposed, secure network coding has been researched for more than a decade. Secure cloud storage protocol for data storage with any secure network coding protocol. The first widely accepted secure cloud storage protocol in the standard model is this one. We thus base our data on public verifiability, which includes the derivation of basic requirements and arguments about security that are only heuristically argued in random Oracle models, or arguments that are not publicly verifiable. Verify earlier studies on consistency. Lastly, we assess the protocol's performance and suggest modifications and prototypes for the future.

Keywords: Service Oriented, loose couple, Authorized Users, Storage Cost, Cloud Computing

1. Introduction

The internet has expanded recently, and one computing technology that has grown is cloud computing. It can supply resources to a user's computer or mobile device and share hardware and software resources. Because cloud computing can integrate resources, users can receive services that are more efficient. Consequently, cloud computing technology needs to meet five fundamental requirements: resource pooling, on-demand self-service, measured service, broad network access, and rapid elasticity (Mell & Grance, 2011). In order to create cloud environments and offer services to users, cloud service providers have teamed together. Three services are provided by cloud service providers: infrastructure as a service (IaaS), platform as a service (PaaS), and software as a service (SaaS). Renting cloud services is less expensive for users than building a cloud environment. The term "cloud computing" describes both the systems software in the data centers that provide those services as well as applications that are delivered as services over the internet. We refer to a cloud that is made pay-as-you-go accessible to the general public as a public cloud. Internal data centers of a company or other organization that are not accessible to the general public are referred to as private clouds. It is well known that SaaS benefits end users as well as service providers. Service providers gain from end users who can access the service at any time, from any location, share and collaborate more easily, and whose data is securely stored within the infrastructure; they also benefit from centralized control over versioning and greatly simplified software installation and maintenance. Three new features are introduced by cloud computing from a hardware stand point. • The illusion of limitless computing resources that are available instantly, so removing the need for cloud users to schedule ahead for provisioning. • The removal of an upfront commitment from cloud users, enabling businesses to start small and scale up their hardware resources only as needed. • The option to pay for the temporary use of computing resources as needed, encouraging conservation by disposing of equipment and storage when it is no longer needed. The following Figure1 shows the Users and providers of cloud computing:

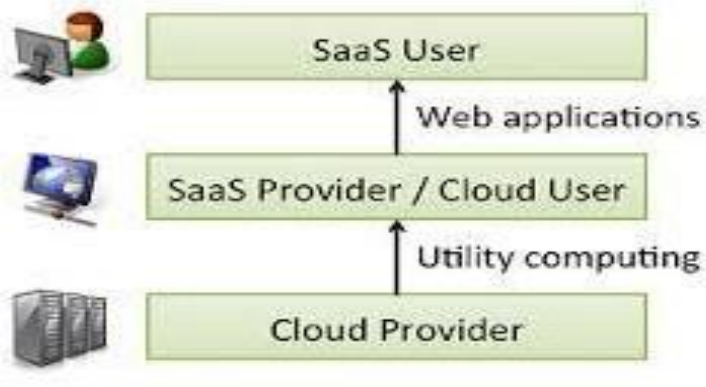

FIGURE 1. Users and Providers of Cloud Computing

Among the various cloud services (such as Google Drive, Dropbox, Amazon S3, and Microsoft OneDrive) available to the general public, cloud storage is the most widely used and well-liked. While cloud storage services offer numerous benefits, they also present several difficult problems, such as security and efficacy (Hashem, Yaqoob, & Anuar, 2015). Because users are unaware of how the cloud storage service handles their data, one of the major challenges is ensuring data integrity. Because of these advantages, the cloud service provider might choose to conceal data loss and errors in the system. Storing data on untrusted cloud storage platforms is a very serious matter for users. Numerous studies (Ateniese, Burns, & Curtmola, Provable data possession at untrusted stores, 2007; Ateniese, Pietro, & Mancini, Scalable and efficient provable data possession, 2008; Erway, K, & Tamassia, 2009; Li, Tan, & Chen, Oct. 2014) present various system and security models to address the issue of data integrity verification in cloud storage services. The verifier's role in these studies can be classified as either private or public auditability. Private auditability suggests that a productive method is for the data owner to directly verify data in the cloud storage service. Public audibility means that the owner of the data permits others to confirm that they may have numerous data files saved in cloud storage services. Sensitive data, including email addresses, medical records, and government information, could be hacked or leaked to unapproved parties. The cloud is an open platform that can be attacked by malevolent insiders as well as external parties. Typically, data security is offered by cloud service providers (CSPs) using virtualization and firewalls. However, because cloud storage servers are located remotely, these mechanisms do not shield users' privacy from the CSP itself. Encrypting data before outsourcing it to the cloud and retrieving it using keyword-based search over encrypted data is a logical way to protect sensitive data privacy. Even though encryption guards against unauthorized access, it greatly increases the computation overhead for data owners, particularly when those owners have large data files and mobile devices with limited resources. Additionally, authorized users need to interact with CSPs in order to retrieve specific files from the cloud and grant them access to encrypted data. In order to achieve efficient data retrieval, it is advisable to obtain only the most pertinent files rather than all of them. This means that files ought to be ranked, with the highest relevant files being returned to users. This is especially advantageous in the context of the "pay-as-you-use" cloud model. Thus, in a cloud environment, effective and safe mechanisms are required to safeguard sensitive data privacy. Furthermore, with cloud applications, the significance and requirement of privacy-preserving data search techniques become even more evident. For instance, big businesses using public clouds like Google or Amazon could have access to private information. Therefore, it's crucial to protect user privacy when using the cloud by concealing search terms and retrieved data. services.

Characteristics of Cloud Computing

Parallel computing includes cloud computing, data center computing, and high-performance computing, also known as supercomputing. The main focus of HPC is scientific computing, which involves a lot of computing. The most crucial factor in HPC is delay.

2.1Conceptional Characteristics – Service Oriented

Although the service-oriented concept is more useful than the grid computing concept of service-oriented architecture (SOA), Two essential elements for realizing the service-oriented concept are abstraction and accessibility. Both the threshold for application development and the requirement for cloud users to understand the specifics of cloud architecture are decreased by abstraction. By examining system parameters like processing performance and storage capacity, cloud users can effortlessly utilize the entire capacity. Cloud computing services can be broadly classified into three categories based on the type of capability they offer: software-as-a-service (SaaS), platform-as-a-service (PaaS), and infrastructure-as-a-service (IaaS) (M & M, 2012). The provision of massive computing resources, including processing, storage, and network capacity, is known as infrastructure-as-a-service. Using storage as an example, a user using cloud computing's storage service only has to pay for the portion used; he doesn't even need to purchase a disk or have any idea where the data he works with is located. Another name for IaaS is Hardware-as-a-service (HaaS). In general, platform-as-a-service supports a set of application program interfaces for cloud applications while abstracting the infrastructure. It serves as a transitional link between applications and hardware. Owing to the significance of platforms, numerous large corporations are vying for control over the cloud computing market, much like Microsoft did with personal computers. The goal of software-as-a-service is to take the place of PC applications. If you use the SaaS, you can avoid installing and running any special software on your computer. By adopting the pay-per-use model, you can lower your overall costs by not purchasing the software at a comparatively higher cost.

2.2 Technical characteristics I - loose couple

Beyond the loose coupling technique of application interaction, loose coupling is the technical foundation of cloud computing. Logically and physically, the infrastructures are distinct through virtualization or other technologies. One part's actions hardly ever impact other parts. The client-server model underpins all cloud computing, which is the most significant aspect. Clients or cloud users have loose connections to cloud providers or servers. Almost none of the users are dependent on data or controls. However, in HPC, data dependence is crucial.

2.3 Technical characteristics 2 -strong fault tolerant

In parallel computing, there are numerous fault-tolerant techniques. There are always some low-level fault correction techniques that work with particular hardware. Numerous specialized applications are researched at a high level using techniques focused on algorithms. At the middle level, one of the best strategies is checking point. The time between two failures in large-scale parallel computer systems might be less than the time it takes for an application to run. It is not necessary to maintain cloud computing systems in their entirety. Faults in cloud computing can primarily occur in four places: provider-inner, provider-across, provider-user, and user-across.

In the event of a provider malfunction, the redundancy or backup provider will take over for the compromised component. The provider-across transaction will be canceled and returned with an error hint if a fault arises among the providers. There are far too many factors that can lead to issues between a provider and a user, including traffic jams, browser crashes, timeouts on requests, busy providers, and hacker attacks.

2.4. Economic characteristics - business model

The primary feature that sets cloud computing and grid computing apart is the business model. The government and academia provide the majority of the grid computing support. Grid computing, on the other hand, is research for the advancement of information technology in the future. However, massive IT companies are the main source of support for cloud computing. In cloud computing, there are a variety of business models, particularly how-to-pay models. In many situations, pay-per-use might be the preferred option. Cloud users can be divided into two groups: end users and median users. The final user uses cloud services for personal purposes. The average user uses cloud services and provides others with expert services at a reasonable cost. Sometimes, the end user does not pay directly for cloud services. Typically, the median user pays directly for the cloud services they use. They can quickly enter the market and save money. The average user doesn't need to learn how to use tools, manage complicated hardware and software, or become proficient with cloud computing technology.

2.5 Features of the user experience: simplicity of use

An essential factor in determining the success of an application is its user experience, which falls under the category of human-computer interaction. The goal of offering cloud users a positive experience is achieved through the use of cloud services. Cloud users should have easy access to these valuable services. Achieving ease of use is at the heart of the user experience. Not only is ease of use elegant, but it is also simple. There are three reasons why cloud computing should be ease use: First, compared to other application program interfaces (API), the majority of cloud providers offer Internet-based interfaces that are easier to use. The business processing is concealed by these interfaces because they are sufficiently elegant and simple. Whether or not business processing has changed, the interfaces can remain unchanged. The user experience of web applications has been thoroughly examined. Thus, content is not a factor in the user interfaces. The entire process of developing a web application can be broken down into three stages: function design, program implementation, and user need analysis. The user experience design is the cornerstone of the entire function design in the top-down approach. Thirdly, web 2.0 facilitates more communication between web users and service providers. The original purpose of the web was to transport hypertext. The web is primarily used as a remote software interface due to the quick and rich developments of ever-more-sophisticated contents. Web 2.0 is intended to be the user experience continuum that obfuscates the distinction between software and the Internet.

2.6 Additional features

Other crucial features include virtualization, high security, and TCP/IP based. Reliable delivery and connection-oriented services are provided between distant applications via TCP/IP. A common protocol in cloud computing is TCP/IP. The majority of cloud users connect to providers via TCP/IP, even though the network protocols in the back end of the data center may be private. The features of the user experience are inspired by the HTTP protocol over TCP/IP or the Internet. Virtualized cloud resources are frequently offered over the Internet as services. Three primary methods are used to achieve high security in cloud computing. To start, a cloud computing system with loose coupling remains functional even if a portion of it is destroyed. Second, the cloud provider's abstraction, virtualization, and privation prevent the specifics of related implementations from being revealed. Third, cloud computing is protected by technology that complies with the law.

III. Design of Cloud Storage System

3.1Architecture of the System

In this work, we examine a cloud data system model that is composed of three primary components. Authorized Users, Cloud Service Provider (CSP), and Data Owner. An entity known as the "Data Owner" (DO) is one that has a lot of data that needs to be stored in the cloud. It could be a single user with mobile devices like smartphones, PDAs, TPM chips, etc. A cloud service provider, or CSP, is an organization that offers users and data owners dynamic computational resources and storage services.

3.2Authorized Users (AU): The owner of the data grants permission to the authorized users to access and share certain keying materials with them. The users with permission would obtain the encrypted data from the cloud, which they can then decrypt to obtain the original data. These three system entities typically interact with one another in the following ways: The data owner wants to outsource the set of files on the cloud server in encrypted form while still keeping the capability to search them through keyword for effective data utilization reasons. 1.When an authorized user wants to retrieve the filecollection, send a search request to the CSP. 2.Then, the CSP search the files and returns set of files and hash values files to the user. 3.Finally, the authorized user verifies the integrity anddecrypts the files and gets the corresponding plain text.

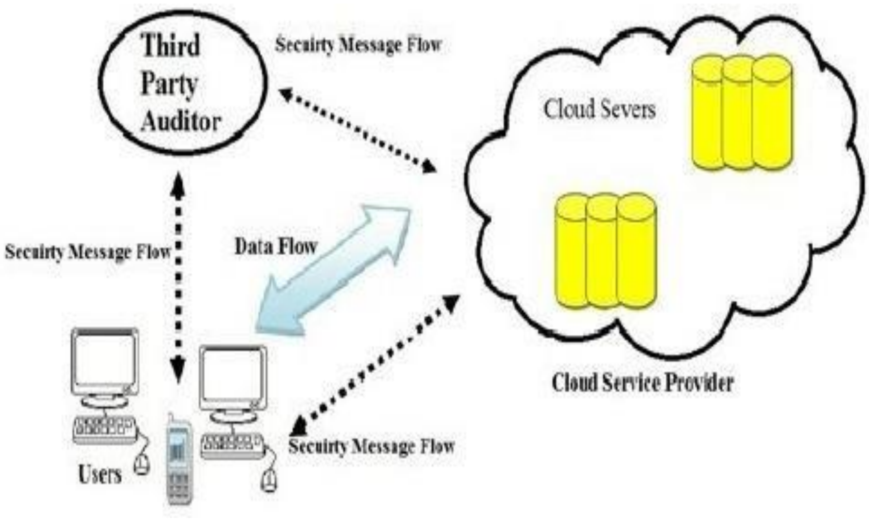

FIGURE 2. Cloud Data Storage Architecture

3.3Models of Systems

In the system model mentioned above, the data owner uses a cloud service provider (CSP) to first outsource the encrypted data files to cloud servers. He has no control over data once it is on the cloud. Even though CSP offers some common security measures to shield the data from attackers, this lack of control over data creates privacy concerns in the cloud because it is hacking (M & M, 2012). Thus, in order to safeguard the confidentiality of sensitive data that is outsourced to the cloud, we require a reliable and secure system. We take into account the following in our scheme for an effective and safe ranked keyword search over encrypted data: For users who are unfamiliar with the file collection, the search result should return the files based on a set of ranked relevance criteria to improve file retrieval accuracy. The index and data, however, should remain unknown to the cloud server since they contain important sensitive data that violates keyword privacy. The CSP only sends the top k most pertinent files to the user-inserted keywords in order to conserve bandwidth.

3.4 Model of Threat

Two primary categories of threats are taken into consideration in the threat model that disrupt cloud-based outsourced data:

3.5Internal Attacks: Users of the cloud, as well as malevolent third parties (cloud providers or client organizations), have a self-interested desire to access or reveal data kept on the cloud. They also change or adjust the information.

3.6External Attacks: These are attacks that come from uninvited sources. It is assumed that these attackers can infiltrate any storage server and thereby gain unauthorized access to the owner's data.

3.7System Objectives

We suggest an effective and safe privacy-preserving strategy with the following objectives in order to address the privacy of sensitive data kept in the cloud:

Ensuring that no malicious insiders or external parties can access the sensitive data content stored on the cloud is the first step in protecting privacy.

Index Privacy: No information about the associated keywords is disclosed by the search or query indexes.

Efficiency: Less computation and communication overhead should be required to accomplish the aforementioned objectives. Data integrity: is the ability to identify changes or deletions made to data while preserving consistency.

IV. Essential Prerequisites and Evaluative Metrics

Where they provide the basic requirements of security and performance (Ahlswede, Cai, & Li, 2000) (Ateniese, Burns, & Curtmola, Provable data possession at untrusted stores, 2007) (Ateniese, Pietro, & Mancini, Scalable and efficient provable data possession, 2008) (Charles, Jain, & Lauter, 2009) (Erway,K, & Tamassia, 2009)

4.1Security Assessment:

Blockless Verification: The auditor does not have to retrieve every audited data block from the cloud storage service in order to verify data blocks.

Stateless Verification: Since the client and cloud storage provider jointly maintain the data situation, the auditor does not need to update or maintain it.

a)Batch auditing: Since the auditor can be assigned by numerous clients, the auditor is able to confirm the information of several clients simultaneously.

b)Dynamic Data: Because the data is constantly updated, the data owner has the ability to add, edit, and remove data blocks from the cloud storage service.

c)Privacy Presenting: The cloud storage service's response does not provide the auditor with access to the assigned data.

d)Performance Assessment:

Computer Cost: We will examine the client TPA and cloud storage service costs on the computing resources to accomplish an effective public auditing.

e)Storage Cost: We will examine the client TPA and cloud storage service cost on the storage spaces since the client will upload data to the cloud storage service without making a local copy of the data files.

2. Conclusion

Users' data security is compromised because it is stored in a cloud storage service. Based on any secure network coding protocol, we have created a general construction of a secure cloud storage protocol. In order to collect files, we first created an index and saved them in the cloud. A trapdoor is made by the authorized user and sent to the host. for upcoming big data generation development. Therefore, efficiently ensuring data integrity in big data will be a significant challenge. Nonetheless, this plan needs to meet certain fundamental needs as well. Using the work that is being done and the protocols that are already in place in the secure network coding area, create a new, effective secure cloud storage protocol. Ultimately, cloud-based encrypted big data can be searched using ranked keyword search and dynamic data operation. First of all, I am glad to thank THE LORD ALMIGHTY for giving me the spirit in completing this paper. I would thankmy family for the constant support they provided throughout my preparation.

References

1. Ahlswede, R., Cai, N., & Li, S.-Y. (2000). Network information flow. IEEE Transactions on Information Theory, 46, no. 4, 1204–1216.
2. Ateniese, G., Burns, R., & Curtmola, R. (2007). Provable data possession at untrusted stores. Proceedings of the 14th ACM Conference on Computer and Communications Security, (pp. 598–609). Virginia, USA.
3. Ateniese, G., Pietro, R. D., & Mancini, L. V. (2008). Scalable and efficient provable data possession. Proceedings of the 4th International Conference on Security and Privacy in Communication Netowrks, (pp. 9:1–9:10). Istanbul, Turkey.
4. Charles, D., Jain, K., & Lauter, K. (2009). Signatures for network coding. International Journal of Information and Coding Theory, 1, no. 1, 3–14.
5. Erway, C., K, A., & Tamassia, R. (2009). Dynamic provable data possession. Proceedings of the 16th ACM Conference on Computer and Communications Security, (pp. 213–222). Illinois, USA.

6. Hashem, I. A., Yaqoob, I., & Anuar, N. B. (2015). The rise of big data on cloud computing: Review and open research issues. Information Systems, 47, no. 6, 98– 115.
7. Juels, A., & Jr, B. K. (2007). Pors: Proofs of retrievability for large files. ACM Conference on Computer and Communications Security (SP, 584– 597.
8. Juels, A., S, J. B., & Kaliski. (2007). Pors: Proofs of retrievability for large files. Proceedings of the 14th ACM Conference on Computer and Communications Security, (pp. 584–597). Virginia, USA.
9. Li, J., Tan, X., & Chen, X. (Oct. 2014). "OPoR: Enabling proof of retrievability in cloud computing with resource-constrained devices. accepted and to be publish in IEEE Transactions on Cloud Computing.
10. Li, Q., Lui, J. C., & Chiu, D.-M. (2012). On the security and efficiency of content distribution via network coding. IEEE Transactions on Dependable andSecure Computing, 9, no. 2, 211–221.
11. Li, S.-Y., Yeung, R. W., & Cai, N. (2003). Linear network coding. IEEE Transactions on Information Theory, 49, no. 2, , 371–381.
12. M, K., & M, S. I. (2012). Efficient similarity search over encrypted data. Proceedings of IEEE InternationalConference On data Engineering, (pp. 1156-67). Washington.
13. Mell, P. M., & Grance, T. (2011). The nist definition ofcloud computing," Technical Report. SP 800-145. Shacham, H., & Waters, B. (2008). Compact proofs of retrievability. Proceedings of the 14th International Conference on the Theory and Application of Cryptology andInformation Security(ASIACRYPT'08), (pp. 90–107). Melbourne,Australia.
14. Wang, C., Chow, S. S., & Wang, Q. (2013). Privacy- preserving public auditing for secure cloud storage.IEEE Transactions on Computers,, 62, no. 2, 362–375.
15. Wang, Q., Wang, C., & Ren, K. (2011). Enabling public auditability and data dynamics for storage security in cloud computing. IEEE Transactions on Parallel and Distributed Systems, 22, no. 5, , 847–859.