A Survey of Cloud Computing Secure Data Storage and Network

[1]Bobby. S, [2]Swathi.M
[1]Assistant Professor, [2]Student,
[1,2]Department of Computer Science, St. Joseph's College of Arts and Science for Women, Hosur, Tamilnadu, India

Abstract - Cloud computing has been popular as the IT architecture. Cloud service providers offer many services based on cloud computing. Cloud storage service is the cloud services which can provide a huge storage space to solve the bottleneck of the storage space of local end users. However, cloud storage service may have data security because the user's data is not stored in their own storage. The service oriented, loose coupling, strong fault tolerant, business model and ease use are main characteristics of cloud computing. Secure cloud storage proposed only recently while secure network coding as been studied for more than ten years. A secure data storage cloud storage protocol given any secure network coding protocol. The protocol is the first publicly verifiable secure cloud storage protocol in the standard model. Therefore, we survey the previous researches of data integrity based on public auditability which includes collecting basic requirements and either not publicly verifiable or security argument is only argued heuristically in the random oracle model. Finally, we propose future development and prototype the proposed protocol and evaluate its performance.

Keywords: *data security, secure coding, prototype, protocol*

I. INTRODUCTION

Cloud computing is a computing technology, and the internet has grown in recent years. It can share the software and hardware resources, and provide resources to a user's computer or mobile device. The user can obtain a more efficient service because cloud computing can integrate resources. Therefore, in order to achieve cloud computing technology, it must satisfy five basic features: On-demand self-service, Broad network access, Resource pooling, Rapid elasticity and Measured service (Mell & Grance, 2011). Thus, cloud service providers have joined to build cloud environments and provide services to the user. Cloud service providers offer three services including Software as a Service (SaaS), Platform as a service (PaaS) and Infrastructure as a Service (IaaS). The cost for users to rent cloud service is cheaper than the cost for users to build cloud environment.

Cloud Computing refers to both the application delivered as services over the internet and systems software in the datacenters that provide those services. When a cloud is made available in a pay-as-you go-manner to the public we call it a public Cloud. We use the term Private Cloud to refer to internal datacenters of a business or other organization that are not made available to the public.

The advantages of SaaS to both end users and service providers are well understood. Service providers enjoy greatly simplified software installation and maintenance and centralized control over versioning end users can access the service anytime anywhere share data and collaborate more easily and keep their data stored safely in the infrastructure. From the hard ware point of view, three aspects are new in cloud computing.

- The illusion of infinite computing resources available on demand, thereby eliminating the need for cloud for cloud computing users to plan for ahead for provisioning;
- The elimination of an up-front commitment by cloud users, thereby allowing companies to start small and increase hardware resources only when there is an increase in their needs.
- The ability to pay for use of computing resources on a short term basis as needed, thereby rewarding conservation by letting machines and storage go when they are no longer useful.

The following Figure1 shows the Users and providers of cloud computing:

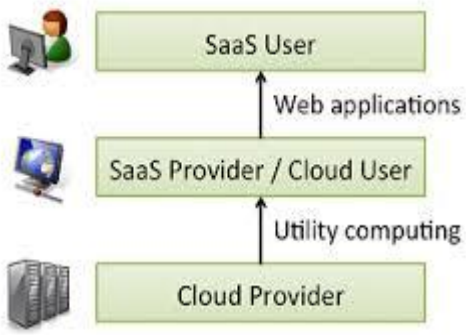

Figure 1: Users and Providers of Cloud Computing

Cloud storage service is the most common and popular service among many cloud services(e.g. Google Drive, Dropbox, Amazon S3 and Microsoft OneDrive) for general users.

Although cloud storage service has many advantages, it brings a lot of challenging issues which include efficacy and security (Hashem, Yaqoob, & Anuar, 2015) One of the big challenges is verifying the integrity of the data because users cannot know how the cloud storage service handles their data. Therefore, the cloud service provider may hide data loss and data errors in the service because their benefits. It is very serious when a user stores data in untrusted cloud storage. In order to solve the problem of data integrity verification in the cloud storage service, many studies present different system and security models (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). In these studies, the role of the verifier can fall into two categories: Private auditability and Public auditability. Private auditability implies the data owner directly verifying data in the cloud storage service is an efficient way. Public auditability implies the data owner allowing other to verify the data owner's may have a lot of data files which are stored in cloud storage service.

The sensitive information such as e-mail, health records, and government data may leak to unauthorized users or even be hacked. Since ,the cloud is an open platform; it can be subjected to attacks from both malicious insiders and outsiders. The Cloud service providers (CSPs) usually provide data security through mechanisms like firewalls and virtualization. However, these mechanisms do not protect users' privacy from the CSP itself due to remote cloud storage servers

A natural approach to preserve the privacy of sensitive data is to encrypt data before outsourcing it into the cloud and retrieves the data back through keyword based search over encrypted data. Although encryption provides protection from illegal accesses, it significantly increases the computation overhead on the data owners especially when they having resource-constrained mobile devices and large size of data files.

Further, the authorized users want to retrieve the certain files from cloud, need to communicate with the CSPs and allow him to operate over the encrypted data. To meet effective data retrieval, it is preferred to get the most relevant files instead of getting all files i.e., The files should be ranked and only highest relevant files are send back to the users, which is highly desirable in the "pay-as-you-use" cloud model.

Therefore, the efficient and secure mechanisms are needed to protect the privacy of sensitive data in a cloud environment. Moreover, the importance and necessity of privacy preserving of data search techniques are even more pronouncing in the cloud applications. For example, large companies that are operating on the public clouds like Google or Amazon may access the sensitive data, search patterns, hiding the query and retrieved data has great importance in ensuring the privacy of that using cloud services.

II. CHARACTERISTICS OF CLOUD COMPUTING

The cloud computing, High performance computing (HPC) or supercomputing and data center computing all belong to parallel computing. HPC focuses on scientific computing which is computing intensive and delay are the most important criteria in HPC.

A. Conceptional Characteristics – Service Oriented

The service oriented concept is similar to but more practical than the concept of SOA in grid computing, Abstraction and accessibility are two keys to achieve the service oriented conception. Abstraction reduces both the need for cloud user to learn the detail of cloud architecture and the threshold of application development. Cloud user can consume all the capacity easily by exploring system parameters such as processing performance and storage capacity. In general, according to the type of provided capability the services of cloud computing are broadly divided into three categories: Infrastructure-as-a service (IaaS), platform-as-a-service (PaaS), and software-as-a-service (SaaS (M & M, 2012). Infrastructure-as-a-service is the delivery of huge computing resources such as the capacity of processing, storage and network. Taking storage as an example, when a user uses the storage service of cloud computing, he just pays the consuming part without buying any disk or even knowing nothing about the location of the data he deals with. Sometimes the IaaS is also called Hardware-as-a-service (HaaS).

Platform-as-a-service generally abstracts the infrastructure and supports a set of application program interface to cloud applications. It is the middle bridge between hardware and application. Because of the importance of platform, many big companies want to grasp the chance of predominating the platform of cloud computing as microsoft does in personal computer time. software-as-a-service aims at replacing the applications running on PC. There is no need to install and run the special software on your computer if you use the SaaS. Instead of buying the software at a relative higher price, you just follow the pay-per-use pattern which can reduce you total cost.

B. Technical characteristics I - loose couple

The loose coupling is the technical fundament of cloud computing and goes beyond the loose coupling method of application interaction. Though virtualization or other technologies, the infrastructures are separate in logic or physic. The behavior of one part hardly affects other parts. Most important of all, whole cloud computing runs in a client-server model. The clients or cloud users connect loosely with servers or cloud providers. All users have almost no data or control dependence. But the data dependence plays a key role in HPC.

C. Technical characteristics 2 -strong fault tolerant

There are many faults tolerant methods in parallel computing. At low-level, there always exist some fault correction mechanisms with specific hardware. At high-level, many specific applications are studied with methods aiming at algorithms. Checking point is one of the most effective methods at middle-level. In large scale parallel computer systems, the interval of two failures may be shorter than application execution time. It is unnecessary to keep the whole states of cloud computing systems. There are mainly four places where faults maybe occur in cloud computing: provider-inner, provider-across, provider-user and user- across.

If a fault occurs in provider, the backup or redundancy of provider will substitute for the failed part. If a fault occurs among providers, the provider-across transaction will be canceled and return with an error hint. There are too many reasons such as congestion, browser collapse, request time out, provider busy and hacker attack can cause faults between provider and user.

D. Economic characteristics - business model

The business model is the key characteristics to distinguish grid computing and cloud computing. The grid computing is mainly supported by government and academe. On the other hand, the grid computing is a research for future development of information technology. But the cloud computing is mainly supported by gigantic IT companies. There are many business models especially how-to-pay models in cloud computing. Pay-per-use may be the favorite one in many cases.

There are two categories of cloud users: End user and median user. End user consumes cloud services for self use. Median user consumes cloud services and cost efficiently supplies professional services to others. End user sometimes doesn't pay for cloud services directly. Median user usually pays for consumed cloud services directly. They save money on jumping to market quickly. For median user, it is no need to manage complex hardware and software, learn how to use tools and gain experience with cloud computing technology.

E. User experience characteristics - ease use

User experience which belongs to the subject of human computer interaction is an important criterion when evaluating whether an application is successful or not. The cloud service is a means toward the end of providing a good experience for cloud user. The valuable services should be easily accessed by cloud user. The core of user experience is achieving ease use. Ease use is not only simple but also elegant.

There are three reasons why cloud computing should be ease use:

First, most cloud providers offer Internet-based interfaces which are simpler than other application program interfaces (API). These interfaces are simple and elegant enough to hide the business processing behind. The interfaces can stay the same ignoring whether the business processing has changed or not.

Second, user experience of web applications is full studied. So the user interfaces are indepent of content. The development of web application has a full suite of flow which can be divided into three stages including user need analysis, function design and program implementation. In top-down method, the user experience design is the fundamental of whole function design.

Third, the web 2.0 increases the interactions between web users and providers. The web was originally designed to transport hypertext. As the rapid and rich developments of increasingly sophisticated contents are appearing, web is usually used as a remote software interface. The web 2.0 is supposed to be the continuum of user experience and blurs the line between software and the Internet.

F. Other characteristics

There are other important characteristics such as TCP/IP based, virtualization and high security. TCP/IP gives reliable delivery, a connection-oriented service between remote applications. TCP/IP is widely used in cloud computing. Although the network protocols may be private in the back end of data center, most cloud user connect to providers through TCP/IP. The HTTP protocol over TCP/IP or Internet inspires the user experience characteristics. Cloud resources are often virtualized as a service over the Internet. High security of cloud computing is achieved mainly through three ways. First, the loose coupling makes cloud computing system run well when part of it is destroyed. Second, the abstraction, virtualization and privation of cloud provider avoid exposing the details of corresponding implementations. Third, technology cooperating with law is the guard of cloud computing.

III. ARCHITECTURE OF CLOUD DATA STORAGE

A. System Architecture

In our work, we consider a model of cloud data system, which consisting of three main entities. Data Owner, cloud service provider (CSP) and Authorized Users.

Data Owner (DO): Is an entity that has large amount of data to be stored in the cloud, can be individual user having mobile constrained devices such as smart phones, PDA, TPM chip, etc...

Cloud Service Provider (CSP): Is an entity, provides data storage services and computational resources dynamically to the data owner and users.

Authorized Users (AU): The data owner allows the authorized users to use their files and share some keying material with the data owner. The authorized users would retrieve the data from the cloud in an encrypted form and by decrypting it they get the original data.

The typical interactions between these three entities of the system are as follows:

1.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.

2. When an authorized user wants to retrieve the file collection, send a search request to the CSP.

3.Then, the CSP search the files and returns set of files and hash values files to the user.

4.Finally, the authorized user verifies the integrity and decrypts the files and gets the corresponding plain text.

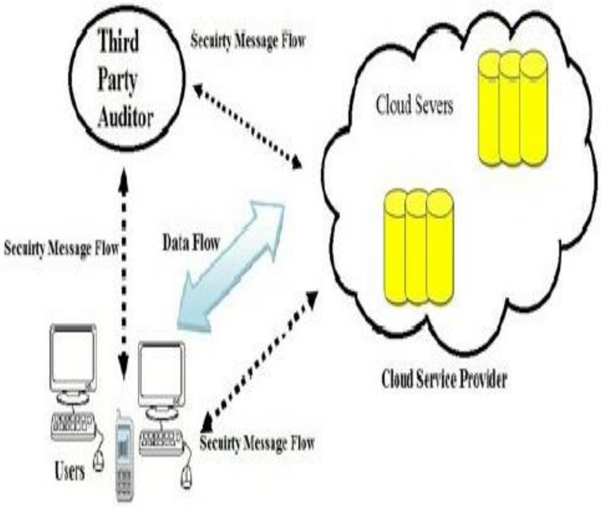

Figure 2: Cloud Data Storage Architecture

B. System Models

In above system model, the data owner first outsources the encrypted data files into cloud servers via Cloud Service Provider (CSP). Once data moves to the cloud, he has no control over it. This is lack of control of data raises privacy issue in the cloud, even if CSP provides some standard security mechanism to protect the data form attackers, still it is hacking (M & M, 2012). Therefore, we need an efficient and secure mechanism to protect the privacy of sensitive outsourced data in the cloud. In our scheme, we consider the efficient and secure ranked keyword search over encrypted data as follows: the search result should returns the files according to certain ranked relevance criteria to improve file retrieval accuracy for users without prior knowledge on the file collection. However, the cloud server should learn nothing about the index and data as they exhibit significant sensitive information against keyword privacy. To reduce bandwidth, the CSP sends only top-k most relevant files to the users inserted keywords.

C. Threat Model

In threat model, we are considering mainly two types threats, which are disturbing the outsourced data in the cloud:

Internal Attacks: which are initiated by malicious insiders: Cloud users, malicious third party user (either cloud provider or customer organizations) are self-interested to accesses the data or disclose the data stored in the cloud. They also alter or modify the data.

External Attacks : which are initiated by unauthorized outsiders we assume that external attackers can compromise all storage servers, so that they can intentionally access the owner's data.

D. System Goals

In order to address the privacy of sensitive data stored in the cloud, we propose an efficient and secure privacy-preserving approach with following goals:

Privacy Preserving: to ensure that there is no way for unauthorized parties and malicious insiders to access the sensitive data content from the cloud.

Index Privacy: The search index or the query index does not leak any information about the corresponding keywords

Efficiency: The above goals should be achieved with less computation and communication overhead.

Data Integrity: Detect the modifications or deletions of data and maintain the consistency of data.

IV. BASIC REQUIREMENTS AND EVALUTION 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)

A. Security Evaluation:

Blockless Verification: The auditor can verify data blocks and need not to retrieve all audited data blocks in the cloud storage service.

Stateless Verification: The auditor need not to maintain and update data situation because data situation is maintained by the client and cloud storage service together.

Batch Auditing : The auditor can verify the data of different clients at the same time because the auditor can be delegated by a lot of clients.

Dynamic Data: The data owner can insert, modify and delete data blocks in the cloud storage service because their data can be continuously updated at any time.

Privacy Presenting: The auditor cannot get knowledge which is the delegated data from the response of the cloud storage service.

B. Performance Evaluation:

Computing Cost: In order to achieve an efficient public auditing we will analyze the client TPA and cloud storage service cost on the computing resources.

Storage Cost: Because the client will upload data to the cloud storage service without the local copy of data files, we will analyze the client TPA and cloud storage service cost on the storage spaces.

CONCLUSION AND FUTURE WORK

The users data is stored in the cloud storage service, it brings users data security issues. We have designed a general construction of secure cloud storage protocol based on any secure network coding protocol. We first created an index for file collection and stored both index and file collection in the cloud. The authorized user creates a trapdoor and sends it to the server. For future development with big data generation. Therefore, it Will be a major challenge how to efficiently verify data integrity in big data. However, this scheme must also satisfy basic requirements. Construct new efficient secure cloud storage protocol based on the current work and existing protocols in the secure network coding area. Finally, dynamic data operation and ranked keyword search over the encrypted big data in cloud.

ACKNOWLEDGEMENT

First of all, I am glad to thank THE LORD ALMIGHTY for giving me the spirit in completing this paper. I would thank my 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 and Secure 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 International Conference On data Engineering, (pp. 1156-67). Washington.

[13] Mell, P. M., & Grance, T. (2011). The nist definition of cloud computing,” Technical Report. SP 800-145.

[14] Shacham, H., & Waters, B. (2008). Compact proofs of retrievability. Proceedings of the 14th International Conference on the Theory and Application of Cryptology and Information Security (ASIACRYPT'08), (pp. 90–107). Melbourne, Australia.

[15] 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.

[16] 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.