A Survey and a Data Integrity Proofs In Cloud Storage

V.Kiruthika **B.R.Laxmi Sree**
School Of IT & Science Dr. G.R.D College Of Science Coimbatore

ABSTRACT:
With the significant advances in Information and Communications Technology (ICT) over the last half century,there is an increasingly perceived vision that computing will one day be the 5th utility (after water, electricity, gas, and telephony). This computing utility, like all other four existing utilities, will Provide the basic level of computing service that is considered essential to meet the everyday needs of the general community .To deliver this vision,a number of computing paradigms have been proposed, of which the latest one is known as Cloud computing. Hence, in this paper, we define a survey on Cloud computing and provide the architecture for creating Clouds, characteristics, deployments, and integrity proofs etc.

Keywords:
Cloud Computing, Integrity, Authentication.

1. Introduction

1.1 Cloud computing

The term “cloud”, as used in this white paper, appears to have its origins in network diagrams that represented the internet, or various parts of it, as schematic clouds. “Cloud computing” was coined for what happens when applications and services are moved into the internet “cloud.” Cloud computing is not something that suddenly appeared overnight; in some form it may trace back to a time when computer systems remotely time-shared computing resources and applications. More currently though, cloud computing refers to the many different types of services and applications being delivered in the internet cloud, and the fact that, in many cases, the devices used to access these services and applications do not require any special applications.
Many companies are delivering services from the cloud. Some notable examples are as follows:
• Google — has a private cloud that it uses for delivering many different services to its users, including email access, document applications, text translations, maps, web analytics, and much more.
• Microsoft — Has Microsoft SharePoint online service that allows for content and business intelligence tools to be moved into the cloud and Microsoft currently makes its office applications available in a cloud.
• Salesforce.com — runs its application set for its customers in a cloud, and it’s Force.com and Vmforce.com products provide developers with platforms to build customized cloud services.

The following sections note cloud and cloud computing characteristics, services models, deployment models, benefits, and challenges.

1.1.1 Characteristics

The characteristics of cloud computing include on-demand self service, broad network access, resource pooling, rapid elasticity and measured service. On-demand self service means that customers (usually organizations) can request and manage their own computing resources. Broad network access allows services to be offered over the Internet or private networks. Services can be scaled larger or smaller; and use of a service is measured and customers are billed accordingly.
Cloud computing has a variety of characteristics, with the main ones being:
• Shared Infrastructure — Uses a virtualized software model, enabling the sharing of physical services, storage, and networking capabilities. The cloud infrastructure, regardless of deployment model, seeks to make the most of the available infrastructure across a number of users.
• Dynamic Provisioning — Allows for the provision of services based on current demand requirements. This is done automatically using software automation, enabling the expansion and contraction of service capability, as needed. This dynamic scaling needs to be done while maintaining high levels of reliability and security.
Network Access — Needs to be accessed across the internet from a broad range of devices such as PCs, laptops, and mobile devices, using standards-based APIs (for example, ones based on HTTP). Deployments of services in the cloud include everything from using business applications to the latest application on the newest smartphones.
• Managed Metering — Uses metering for managing and optimizing the service and to provide reporting and billing information. In this way, consumers are billed for services according to how much they have actually used during the billing period. In short, cloud computing allows for the sharing and scalable deployment of services, as needed, from almost any location, and for which the customer can be billed based on actual usage.

1.1.2 Service Models

The cloud computing service models are Software as a Service (SaaS), Platform as a Service (PaaS) and

Manuscript received November 5, 2015
Manuscript revised November 20, 2015

Infrastructure as a Service (IaaS). In a Software as a Service model, a pre-made application, along with any required software, operating system, hardware, and network are provided. In PaaS, an operating system, hardware, and network are provided, and the customer installs or develops its own software and applications. The IaaS model provides just the hardware and network; the customer installs or develops its own operating systems, software and applications.

Once a cloud is established, how its cloud computing services are deployed in terms of business models can differ depending on requirements. The primary service models being deployed are commonly known as:

• Software as a Service (SaaS) — Consumers purchase the ability to access and use an application or service that is hosted in the cloud. A benchmark example of this is Salesforce.com, where necessary information for the interaction between the consumer and the service is hosted as part of the service in the cloud.

Also, Microsoft is expanding its involvement in this area, and as part of the cloud computing option for Microsoft® Office 2010, its Office Web Apps are available to Office volume licensing customers and Office Web App subscriptions through its cloud-based Online Services.

• Platform as a Service (PaaS) — Consumers purchase access to the platforms, enabling them to deploy their own software and applications in the cloud. The operating systems and network access are not managed by the consumer, and there might be constraints as to which applications can be deployed.

• Infrastructure as a Service (IaaS) — Consumers control and manage the systems in terms of the operating systems, applications, storage, and network connectivity, but do not themselves control the cloud infrastructure. Also known are the various subsets of these models that may be related to a particular industry or market. Communications as a Service (CaaS) is one such subset model used to describe hosted IP telephony services. Along with the move to CaaS is a shift to more IP-centric communications and more SIP trunking deployments. With IP and SIP in place, it can be as easy to have the PBX in the cloud as it is to have it on the premise. In this context, CaaS could be seen as a subset of SaaS.

1.1.3 Deployment of cloud services:

Cloud services are typically made available via a private cloud, community cloud, public cloud or hybrid cloud. Generally speaking, services provided by a public cloud are offered over the Internet and are owned and operated by a cloud provider. Some examples include services aimed at the general public, such as online photo storage services, e-mail services, or social networking sites. However, services for enterprises can also be offered in a public cloud. In a private cloud, the cloud infrastructure is operated solely for a specific organization, and is managed by the organization or a third party. In a community cloud, the service is shared by several organizations and made available only to those groups. The infrastructure may be owned and operated by the organizations or by a cloud service provider

Deploying cloud computing can differ depending on requirements, and the following four deployment models have been identified, each with specific characteristics that support the needs of the services and users of the clouds in particular ways.

• Private Cloud — the cloud infrastructure has been deployed, and is maintained and operated for a specific organization. The operation may be in-house or with a third party on the premises.

• Community Cloud — the cloud infrastructure is shared among a number of organizations with similar interests and requirements.

This may help limit the capital expenditure costs for its establishment as the costs are shared among the organizations. The operation may be in-house or with a third party on the premises.

• Public Cloud — the cloud infrastructure is available to the public on a commercial basis by a cloud service provider. This enables a consumer to develop and deploy a service in the cloud with very little financial outlay compared to the capital expenditure requirements normally associated with other deployment options.

• Hybrid Cloud — the cloud infrastructure consists of a number of clouds of any type, but the clouds have the ability through their interfaces to allow data and/or applications to be moved from one cloud to another. This can be a combination of private and public clouds that support the requirement to retain some data in an organization, and also the need to offer services in the cloud.

1.1.4 Benefits

*T*he following are some of the possible benefits for those who offer cloud computing-based services and applications:

• Cost Savings — Companies can reduce their capital expenditures and use operational expenditures for increasing their computing capabilities. This is a lower barrier to entry and also requires fewer in-house IT resources to provide system support.

• Scalability/Flexibility — Companies can start with a small deployment and grow to a large deployment fairly rapidly, and then scale back if necessary. Also, the flexibility of cloud computing allows companies to use extra resources at peak times, enabling them to satisfy consumer demands.

• Reliability — Services using multiple redundant sites can support business continuity and disaster recovery.

• Maintenance — Cloud service providers do the system maintenance, and access is through APIs that do not require application installations onto PCs, thus further reducing maintenance requirements.
• Mobile Accessible — Mobile workers have increased productivity due to systems accessible in an infrastructure available from anywhere.

1.2 Introduction of Data Storage in Cloud

Cloud Storage is a crucial service of cloud computing, that permits information house owners (owners) to maneuver data from their native computing systems to the cloud. More and additional house owners begin to store the information within the cloud. However, this new paradigm of information hosting service additionally introduces new security challenges. Data owners would worry that the information can be lost within the cloud.

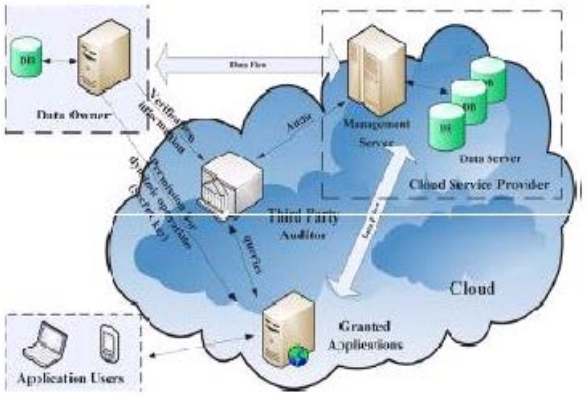

Fig. 1: Data storage in cloud

This is because information loss might happen in any infrastructure, no matter what high degree of reliable measures cloud service providers would take. Sometimes, cloud service suppliers could be dishonest. They may discard the data that haven't been accessed or seldom accessed to save the cupboard space and claim that the information area unit still correctly hold on within the cloud. Therefore, house owners have to be compelled to be convinced that the information area unit properly holds on within the cloud. Traditionally, house owners will check the information integrity based mostly on two-party storage auditing protocols. In cloud storage system, however, it is inappropriate to let either facet of cloud service providers or house owners conduct such auditing, as a result of none of them can be sure to give unbiased auditing result. During this scenario, third-party auditing could be a natural choice for the storage auditing in cloud computing. A third-party auditor (auditor) that has experience and capabilities can do a additional economical work and persuade each cloud service suppliers and house owners. For the third-party auditing in cloud storage systems, there are units many necessary needs that are projected in some previous works. The auditing protocol ought to have the subsequent properties:

1) Confidentiality. The auditing protocol ought to keep owner's information confidential against the auditor.
2) Dynamic auditing. The auditing protocol ought to support the dynamic updates of the data within the cloud.
3) Batch auditing. The auditing protocol ought to even be able to support the batch auditing for multiple house owners and multiple clouds.

Recently, many remote integrity checking protocols were projected to permit the auditor to envision the information integrity on the remote server. Table 1 provides the comparisons among some existing remote integrity checking schemes in terms of the performance, the privacy protection, the support of dynamic operations and also the batch auditing for multiple owners and multiple clouds. From Table 1, they will notice that many of them aren't privacy conserving or cannot support the information dynamic operations, in order that they cannot be applied to cloud storage systems. In [11] the authors projected a dynamic auditing protocol that may support the dynamic operations of the data on the cloud servers, however this technique might leak the data content to the auditor as a result of it needs the server to send the linear mixtures of information blocks to the auditor.

Table 1. Comparison of remote integrity checking schemes

Parameters	SPDP(interactive provable data possession)	DPDP
type of guarantee	probabilistic/Deterministic	Probabilistic
data dynamics	append only	Yes

In [12] the authors extended their dynamic auditing scheme to be privacy conserving and support the batch auditing for multiple house owners. However, as a result of the massive number of information tags, their auditing protocols might incur a heavy storage overhead on the server. In proposed work a cooperative demonstrable information possession theme that can support the batch auditing for multiple clouds and also extend it to support the dynamic auditing. However, their theme cannot support the batch auditing for multiple house owners. That's as a result of parameters for generating the information tags employed by every owner area unit completely different, and thus, they cannot mix the information tags from multiple owners to conduct the batch auditing. Another downside is that their theme needs a further sure organizer to send a commitment to the auditor throughout the multi cloud batch auditing, as a result of their theme applies the mask technique to confirm

the information privacy. However, such additional organizer isn't sensible in cloud storage systems. Moreover, each Wang's schemes and Zhu's schemes incur serious computation price of the auditor, which makes the auditor a performance bottleneck.

1.3 Overview of the Network and Cloud Security

Network Security is an organization's strategy and provisions for ensuring the security of its assets and of all network traffic. Network security is manifested in an implementation of security policy, hardware, and software. For the purposes of this discussion, the following approach is adopted in an effort to view network security in its entirety

Policy, Enforcement, Auditing, Policy

The IT Security Policy is the principle document for network security. Its goal is to outline the rules for ensuring the security of organizational assets. Employees today utilize several tools and applications to conduct business productively. Policy that is driven from the organization's culture supports these routines and focuses on the safe enablement of these tools to its employees. The enforcement and auditing procedures for any regulatory compliance an organization is required to meet must be mapped out in the policy as well.

Enforcement

Most definitions of network security are narrowed to the enforcement mechanism. Enforcement concerns analyzing all network traffic flows and should aim to preserve the confidentiality, integrity, and availability of all systems and information on the network. These three principles compose the CIA triad:

Confidentiality - involves the protection of assets from unauthorized entities

Integrity - ensuring the modification of assets is handled in a specified and authorized manner

Availability - a state of the system in which authorized users have continuous access to said assets.

2. Overview of the System

Data outsourcing to cloud storage servers is raising trend among many firms and users owing to its economic advantages. This essentially means that the owner (client) of the data moves its data to a third party cloud storage server which is supposed to - presumably for a fee - faithfully store the data with it and provide it back to the owner whenever required.

As data generation is far outpacing data storage, it proves costly for small firms to frequently update their hardware whenever additional data is created. Also maintaining the storages can be a difficult task. Storage outsourcing of data to cloud storage helps such firms by reducing the costs of storage, maintenance and personnel. It can also assure a reliable storage of important data by keeping multiple copies of the data thereby reducing the chance of losing data by hardware failures.

Storing of user data in the cloud despite its advantages has many interesting security concerns which need to be extensively investigated for making it a reliable solution to the problem of avoiding local storage of data. In this project, the problem of implementing a protocol for obtaining a proof of data possession in the cloud sometimes referred to as Proof of retrievability (POR).This problem tries to obtain and verify a proof that the data that is stored by a user at a remote data storage in the cloud is not modified by the archive and thereby the integrity of the data is assured.

Such verification systems prevent the cloud storage archives from misrepresenting or modifying the data stored at it without the consent of the data owner by using frequent checks on the storage archives. Such checks must allow the data owner to efficiently, frequently, quickly and securely verify that the cloud archive is not cheating the owner.

2.1 Objective of the System

The main objective of this project is to deal with the problem of implementing a protocol for obtaining a proof of data possession in the cloud sometimes referred to as Proof of retrievability (POR).This problem tries to obtain and verify a proof that the data that is stored by a user at a remote data storage in the cloud (called cloud storage archives or simply archives) is not modified by the archive and thereby the integrity of the data is assured. Such kinds of proofs are very much helpful in peer-to-peer storage systems, network file systems, longterm archives, web-service object stores, and database systems. Such verification systems prevent the cloud storage archives from misrepresenting or modifying the data stored at it without the consent of the data owner by using frequent checks on the storage archives. Such checks must allow the data owner to efficiently, frequently, quickly and securely verify that the cloud archive is not cheating the owner. Cheating, in this context, means that the storage archive might delete some of the data or may modify some of the data.

It must be noted that the storage server might not be malicious instead, it might be simply unreliable and lose or inadvertently corrupt the hosted data. But the data integrity schemes that are to be developed need to be equally applicable for malicious as well as unreliable cloud storage servers. Any such proofs of data possession schemes do not, by itself, protect the data from corruption by the archive. It just allows detection of tampering or deletion of a remotely located file at an unreliable cloud storage server. To ensure file robustness other kind of techniques like data redundancy across multiple systems can be maintained

3. System Study and Analysis

3.1 Existing System

As data generation is far outpacing data storage it proves costly for small firms to frequently update their hardware whenever additional data is created. Also maintaining the storages can be a difficult task. In transmitting the file across the network to the client can consume heavy bandwidths. The problem is further complicated by the fact that the owner of the data may be a small device, like a PDA (personal digital assist) or a mobile phone, which have limited CPU power, battery power and communication bandwidth.

Drawbacks of The Existing System

The main drawback of this scheme is the high resource costs it requires for the implementation.
Also computing hash value for even a moderately large data files can be computationally burdensome for some clients (PDAs, mobile phones, etc).
Data encryption is large so the disadvantage is small users with limited computational power (PDAs, mobile phones etc.).
Information system functionality remains the top priority and the security mechanisms are considered only when vulnerability gets exploited.

3.2 Proposed System

One of the important concerns that need to be addressed is to assure the customer of the integrity i.e. correctness of his data in the cloud. As the data is physically not accessible to the user the cloud should provide a way for the user to check if the integrity of his data is maintained or is compromised. In this project, a new scheme is used that gives a proof of data integrity in the cloud which the customer can employ to check the correctness of his data in the cloud. This proof can be agreed upon by both the cloud and the customer and can be incorporated in the Service level agreement (SLA). It is important to note that our proof of data integrity protocol just checks the integrity of data i.e. if the data has been illegally modified or deleted.

Advantages Of The Proposed System

* Data outsourcing to the cloud also helps in reducing the maintenance.
* It avoids local storage of data.
* By using the proposed scheme, costs of storage, maintenance and personnel will be reduced.
* It reduces the chance of losing data by hardware failures.
* Unauthorized transactions will be reduced.

4. Implementation

* Creating Cloud Storage
* Generating Authorized Code
* Verification of Integrity
* Data Validation
* Third Party Auditing

* Creating Cloud Storage

Data outsourcing to cloud storage servers is raising trend among many firms and users owing to its economic advantages. Cloud computing is an entity, which is managed by Cloud Service Provider(CSP) has significant storage space and computation resource to maintain client's data.

This essentially means that the owner (client) of the data moves their data to a third party cloud storage server and those data will be provided back to the owner whenever it is required.

* Generating Authorized Code

In this module, a security code will be generated by the admin dynamically and send it to the registered user email id .

That security code will be used by the user for login purpose.

Thus the data in the cloud can be protected from the unauthorized access.

*Verification of Integrity

After login by using security code if the user want to access the data in the cloud the user must request to the admin.

This module tries to obtain and verify a proof that the data that is stored by a user at remote data storage in the cloud (called cloud storage archives or simply archives) is not modified by the archive and thereby the integrity of the data is assured.

*Data Validation

In this module, the verifier before storing the file at the archive, preprocesses the file and appends some Meta data to the file and stores at the archive.

At the time of verification the verifier uses this Meta data to verify the integrity of the data.

*Third Party Auditing

The introduction of TPA eliminates the involvement of the client through the auditing of whether his data stored in the cloud are indeed intact, which can be important in achieving economics of scale for cloud computing.

TPA is an entity , which has expertise and capabilities that clients do not have, is trusted to assess and expose risk of cloud storage services on behalf of the clients upon request.

The client or TPA can verify the integrity of the outsourced data by challenging the server.

5. Conclusion and Scope of Future Enhancement

In this work, our proposed scheme is used to facilitate the client in getting a proof of integrity of the data that they store in the cloud storage servers with bare minimum costs and efforts. Our scheme was developed to reduce the computational and storage overhead of the client as well as to minimize the computational overhead of the cloud storage server. It also minimized the size of the proof of data integrity so as to reduce the network bandwidth consumption.

The operation of encryption of data generally consumes a large computational power. In our scheme the encrypting process is very much limited to only a fraction of the whole data thereby saving on the computational time of the client. Many of the schemes proposed earlier require the archive to perform tasks that need a lot of computational power to generate the proof of data integrity [3]. But in our scheme the archive just need to fetch and send few bits of data to the client.

The network bandwidth is also minimized as the size of the proof is comparatively very less (k+1 bits for one proof).It should be noted that our scheme applies only to static storage of data. It cannot handle to case when the data need to be dynamically changed. Hence developing on this will be a future enhancement.

References

[1] Towards Secure and Dependable Storage Services in Cloud Computing Cong Wang, Student Member, IEEE, Qian Wang, Student Member, IEEE, Kui Ren, Member, IEEE, Ning Cao, Student Member, IEEE, and Wenjing Lou, Senior Member, IEEE- 2011.

[2] Addressing cloud computing security issues, Future generation computer systems (2011) www.elsevier.com/locate/fgcs.

[3] T. Wobber, T. L. Rodeheffer, and D. B. Terry, “Policy-based access control for weakly consistent replication,” in ACM EuroSys, 2010.

[4] E. Mykletun, M. Narasimha, and G. Tsudik, “Authentication and integrity in outsourced databases,” Trans. Storage, vol. 2, no. 2, pp. 107–138, 2013.

[5] D. X. Song, D. Wagner, and A. Perrig, “Practical techniques for searches on encrypted data,” in SP ’00: Proceedings of the 2000 IEEE Symposium on Security and Privacy. Washington, DC, USA: IEEE Computer Society, 2012, p. 44.

[6] A. Juels and B. S. Kaliski, Jr., “Pors: proofs of retrievability for large files,” in CCS ’07: Proceedings of the 14th ACM conference on Computer and communications security. New York, NY, USA: ACM, 2007, pp.584–597.

[7] G. Ateniese, R. Burns, R. Curtmola, J. Herring, L. Kissner, Z. Peterson,and D. Song, “Provable data possession at untrusted stores,” in CCS ’07:Proceedings of the 14th ACM conference on Computer and communications

[8] Beginning ASP.NET 3.5 in C# 2008: From Novice to Professional, Second Edition by Matthew MacDonald.

[9] Ateniese, Giuseppe, et al. "Improved proxy re-encryption schemes with applications to secure distributed storage." ACM Transactions on Information and System Security (TISSEC) 9.1 (2006): 1-30.

[10] Lu, Rongxing, et al. "Secure provenance: the essential of bread and butter of data forensics in cloud computing." Proceedings of the 5th ACM Symposium on Information, Computer and Communications Security. ACM, 2010.

[11] Marshall D.Abrams, Harold J.podell on Cryptography.

[12] Anoop Ms, “public key cryptography Applications Algorithms and Mathematical Explanations”.