International Journal of Scientific Engineering and Applied Science (IJSEAS) – Volume-2, Issue-3, March 2016
ISSN: 2395-3470
www.ijseas.com

REPLICATION APPROACH IN CLOUD COMPUTING ON DATA STORAGE SECURITY

Loheswaran.K[1], Anbarasu.S[2]
[1]Associate Professor, [2] Assistant Professor, Department of CSE,
Sasurie College of Engineering, Vijayamangalam.

Abstract:

Cloud storage enables users to remotely store their data and enjoy the on-demand high quality cloud applications without the burden of local hardware and software management. Organizations and individuals worldwide are evaluating and experimenting the possibilities of cloud-based computing. Organizations examine cloud computing as a simple and flexible model for outsourcing the management and maintenance of IT-infrastructure, whereas individuals experience cloud as a realm of services. As the cloud-based data storage services have evolved to meet the requirements of modern data management, they have gained wide interest from both organizations and individuals. However, the interest is restrained, because the basic concept of cloud-based storage services has not gained complete trust from either side. Several concerns have risen about storing data into a completely unknown grid. If an attacker chooses to attack a specific client, then he can aim at a fixed cloud provider, try to have access to the client's information. This makes an easy job of the attackers, both inside and outside attackers get the benefit of using data mining to a great extent. Inside attackers refer to malicious employees at a cloud provider. Thus single data mart storage architecture is the biggest security threat concerning data mining on cloud, so in this paper present the secure replication approach that encrypt and replicate the data in distributed data mart storage system. This approach involves the encryption, replication and storage of data.

1. INTRODUCTION

Cloud computing is technology that provides the different services at very low cost. The different client stores data on Cloud storage. Cloud computing provides storage for storing the information and provides the security of that information. Cloud service models are infrastructure as a service, platform as a service and software as a service.

Cloud services are provided by different famous organizations like Google, Amazon and Microsoft etc. By using these services the client avoid the cost of buying extra resources. Cloud services provide the high computation capacity at low cost. The various data analysis techniques which are used for extracting valuable information from a large volume of data. These different techniques are used by Cloud service provider like Google uses the technique for identifying the user behavior on the basis of search behavior.

In previous trend data to store on a single cloud the attacker applies an attack on it and accesses the information which is stored by the client on Cloud storage. If the client is an organization related to healthcare, shopping etc then there is big loss of information access by attackers, so distributed environment handles such kind of problem.

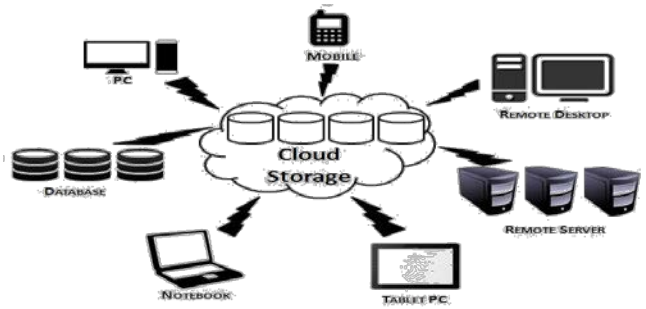

Fig.1.1 cloud storage

The previous figure indicates cloud storage stores the information from different devices and these devices can access the information on demand at any time. Here purposes a secure approach that replicates the client's data and store on different data mart. Before replication, full copy of encrypted information stores on data warehouse for increasing the availability of information. It will increase the reliability and privacy of data.

A.SaaS (Software as a Service)

Software as a Service is a considerable change as we see software. There is no expenditure capital, only service cost. Software just like the processing power and storage is seen as a utility that clients can pay for only as needed. The goal is to centralize administrative tasks while improving scalability and workloads.

B.PaaS (Platform as a Service)

It offers a platform to clients for different purposes. For example, Windows Azure offers a platform to developers to build, test, and host applications that can be accessed by the end users. The end users may or may not know that the application is hosted on the cloud. The storage space for user data may be increased or decreased as per the requirement of the applications. As with the SaaS, users do not need to build the platform. Users just pay a nominal fee for using the service.

C .IaaS (Infrastructure as a Service)

It offers infrastructure on demand. The infrastructure can be anything from storage servers to applications to operating systems. Office 365 offers a combination of these infrastructure and falls under this category. With Office 365, user can get plenty of applications along with storage space. Buying infrastructure or renting it out in traditional models can be very expensive. When users opt for IaaS, they save a lot on expenses, space, and personnel required to set up and maintain the infrastructure. The cloud service provider takes care of setting up and maintaining the infrastructure. They just pay a fee to use it as per their requirements.

II.LITERATURE SURVEY

Cloud Computing

Cloud computing is a very young concept and there is no consensus on a formal definition at the time of writing most experts agree that the cloud computing is a buzz which encompasses a variety of services. Other focus on the business model which is typically a pay –as –you-go services.

The following definition approaches cloud computing from a broad conceptual level:

Cloud computing represents a broad array of web –based s services aimed at allowing users to obtain a wide range of functional capabilities on a "pay-as-you-go" basics that previously required tremendous hardware/software investments and professional skills to acquire. Cloud computing is the realization of the earlier ideals of utility computing without the technical complexities or complicated deployment worries.

Although most definition do not use such generalized concepts, these generalizations are often implied as a base for other definitions. This makes the definition above highly applicable. as an addendum to the definition above, these key technical concept are often associated with (bt not requires of) cloud computing : instantaneous and on-demand resource scalability, parallel and distributed computing ,and virtualization.

Uses of Cloud Computing

Various authors have proposed three different tiers of systems employed by cloud service provider .These tiers make up the different levels of technologies used in cloud computing.

Infrastructure as a Service (IaaS)

This level represents the most computational and storage (e.g., Microsoft, Google, Amazon) manage a vast set of computational and storage resources. Depending on the provider, end users may have direct access of the hardware resources or access to a set of virtual resources.

Clouds typically utilize virtual resources and grid applications typically have direct access to hardware. Application and service built upon virtual resources sets are not hardware dependent and can be deployed seamlessly across different cloud platforms. This service is best representing by services like Amazon EC2, a virtual machine platform.

Platform as a Service (PaaS)

At the next level of services are presented to users as a software /application platform instead of hardware. Typically this layer consists of application frameworks that make up the basis of the SaaS layer describe next. The Google APP Engine and Microsoft Azure both offer a large set of programming tools at this level.

Software as a Service (SaaS)

This is the highest level of services provided by cloud platforms. This level provides applications that end users interact with. Examples include Google Docs, Microsoft Office live, Google Maps and Face book.

III.RELATED WORK

The proposed algorithm for encryption of sensitive data for C-governance along with their proper decryption to the authorized user by using of hadamard matrix. The proposed a practical security model based on key security considerations by looking at a number of infrastructure aspects of Cloud Computing such as SaaS, Utility, Web, Platform and Managed Services, Service commerce platforms and Internet Integration. It have introduced a set of security protocols for ensuring the privacy and legal compliance of customer data in cloud computing architectures. Jaeger and Schiffman have proposed in their paper to improve security of cloud architecture by building

"verifiable base systems". It presented information risk management framework for better understanding critical areas of focus in cloud computing

.

IV.STORAGE AS A SERVICE (SaaS)

One of the primary uses of cloud computing is for data storage. With cloud storage, data is stored on multiple third-party servers, rather than on the dedicated servers used in traditional networked data storage. When storing data, the user sees a virtual server that is, it appears as if the data is stored in a particular place with a specific name. But that place doesn't exist in reality. It's just a pseudonym used to reference virtual space carved out of the cloud. In reality, the user's data could be stored on any one or more of the computers used to create the cloud. The actual storage location may even differ from day to day or even minute to minute, as the cloud dynamically manages available storage space. But even though the location is virtual, the user sees a static location for his data and it can actually manage his storage space as if it were connected to his own personal computer.

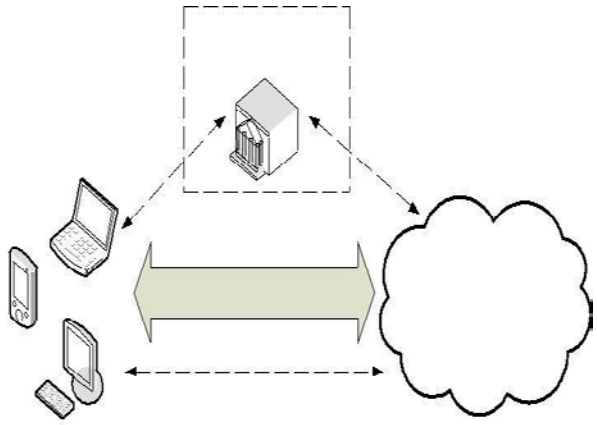

Fig. 4.1 Cloud data storage architecture

Cloud storage has both financial and security advantages. Financially, virtual resources in the cloud are typically cheaper than dedicated physical resources connected to a personal computer or network. As for security, data stored in the cloud is secure from accidental erasure or hardware crashes, because it is duplicated across multiple physical machines; since multiple copies of the data are kept continually, the cloud continues to function as normal even if one or more machines go offline. If one machine crashes, the data is replicated on other machines in the cloud. In this paper, we have mainly focused on storage as a service in the cloud computing.

V.TYPES OF INFORMATION

For any organization the data is classified into the following categories.

A. Public

Information that is similar to unclassified information. All information of a company that does not fit into any of the next categories can be considered public. While its unauthorized disclosure may be against policy, it is not expected to impact seriously or adversely the organization, its employees and/or its customers.

B.Sensitive

This is the information that requires a higher level of security than public data. This information is protected from a loss of confidentiality as well as from a loss of integrity due to an unauthorized alteration. This classification applies to information that requires special precautions to ensure the integrity of the information by protecting it from unauthorized modification or deletion. It is information that requires a higher than normal assurance of accuracy and completeness.

C.Private

This classification applies to personal information that is intended for use within the organization. Its unauthorized disclosure could seriously and adversely impact the organization and its employees. For example, salary levels and medical information are considered private.

D.Confidential

This classification applies to the most sensitive business information that is intended strictly for use within the organization. Its unauthorized disclosure could seriously and adversely impact the organization, its stockholders, its business partners, and its customers.

This information is exempted from disclosure under the provisions of the Freedom of Information Act or other applicable federal laws or regulations. For example, information about new product development, trade secrets, and merger negotiations is considered confidential.

Cloud Security Guidance

As consumers transition their applications and data to use cloud computing, it is critically important that the level of security provided in the cloud environment be equal to or better than the security provided by their traditional IT environment. Failure to ensure appropriate security protection could ultimately result in higher costs and potential loss of business thus eliminating any of the potential benefits of cloud computing.

This section provides a prescriptive series of steps that should be taken by cloud consumers to evaluate and manage the security of their cloud environment with the goal of mitigating risk and delivering an appropriate level of support. The following steps are discussed in detail:

1. Ensure effective governance, risk and compliance processes exist
2. Audit operational and business processes
3. Manage people, roles and identities
4. Ensure proper protection of data and information
5. Enforce privacy policies
6. Assess the security provisions for cloud applications
7. Ensure cloud networks and connections are secure
8. Evaluate security controls on physical infrastructure and facilities
9. Manage security terms in the cloud SLA
10. Understand the security requirements of the exit process

Requirements and best practices are highlighted for each step. In addition, each step takes into account the realities of today's cloud computing landscape and postulates how this space is likely to evolve in the future, including the important role that standards will play to improve interoperability and comparability across providers.

VI. PROBLEM STATEMENT

From the perspective of data security, which has always been an important aspect of quality of service, Cloud Computing inevitably poses new challenging security threats.

1. Firstly, traditional cryptographic primitives for the purpose of data security protection cannot be directly adopted due to the user's loss control of data under Cloud Computing.

Therefore, verification of correct data storage in the cloud must be conducted without explicit knowledge of the whole data. Considering various kinds of data for each user stored in the cloud and the demand of long term continuous assurance of their data safety, the problem of verifying correctness of data storage in the cloud becomes even more challenging.

2. Secondly, Cloud Computing is not just a third party data warehouse. The data stored in the cloud may be frequently updated by the users, including insertion, deletion, modification, appending, reordering, etc. To ensure storage correctness under dynamic data update is hence of paramount importance.

3. Third is the deployment of cloud computing, it is powered by data centers running in a simultaneous cooperated and distributed manner. Individual user's data is redundantly stored in multiple physical locations to further reduce e data integrity threats.

Notation and Preliminaries

- f– The data file to be stored. We assume that F can be denoted as a matrix of m equal-sized data vectors, each consisting of l blocks. Data blocks are all well represented as elements in Galois Field GF (2p) for p = 8 or 16.
- A – The dispersal matrix used for Reed-Solomon coding.
- G – The encoded file matrix, which includes a set of n = m + k vectors, each consisting of l blocks
- f key (•) – pseudorandom function (PRF), which is defined as f : {0, 1}∗ × key → GF (2p).
- φ key (•) – pseudorandom permutation (PRP), which is defined as φ : {0, 1}log2 (ℓ) × key → {0, 1}log2 (ℓ)
- ver – a version number bound with the index for individual blocks, which records the times the block has been modified. Initially we assume ver is 0 for all data blocks.

Ensuring cloud data storage

In cloud data storage system, users store their data in the cloud and no longer possess the data locally. Thus, the correctness and availability of the data files being stored on the distributed cloud servers must be guaranteed.

Our main scheme for ensuring cloud data storage is presented in this section. The first part of the section is devoted to a review of basic tools from coding theory that is needed in our scheme for file distribution across cloud servers. Then, the homomorphic token is introduced. The token computation function we are considering belongs to a family of universal hash function, chosen to preserve the homomorphic properties, which can be perfectly integrated with the verification of erasure-coded data. Subsequently, it is shown how to derive a challenge-response protocol for verifying the storage correctness as well as identifying misbehaving servers. Finally the procedure for file retrieval and error recovery based on erasure-correcting code is also outlined.

File distribution preparation

It is well known that erasure-correcting code may be used to tolerate multiple failures in distributed storage systems. In cloud data storage, we rely on this technique to disperse the data file F redundantly across a set of n = m + k distributed servers. An (m, k) Reed-Solomon erasure-correcting code is used to create k redundancy parity vectors from m data vectors in such a way that the original m data vectors can be reconstructed from any m out of the m + k data and parity vectors.

By placing each of the m+k vectors on a different server, the original data file can survive the failure of any k of the m + k servers without any data loss, with a space overhead of k/m. For support of efficient sequential I/O to the original file, our file layout is systematic, i.e., the unmodified m data file vectors together with k parity vectors is distributed across m + k different servers.

Let **F** = (F1 , F2 , . . . , Fm) and Fi = (f1i , f2i, . . . , fli)T (i ∈ {1, . . . , m}). Here T (shorthand for transpose) de-notes that each Fi is represented as a column vector, and l denotes data vector size in blocks. All these blocks are elements of GF (2p). The systematic layout with parity vectors is achieved with the information dispersal matrix A, derived from a an m×(m+k) Vander monde matrix :

1 1 … 1 1 … 1

β1 β2 … βm βm+1 … βn

B1m-1 β2m-1 … βmm-1 … βm+1m-1 βnm-1

where βj (j ∈ {1, . . . , n}) are distinct elements randomly Picked from GF (2p).After a sequence of elementary row transformations,

the desired matrix A can be written as 1 0 . . . 0 P11 P12 . . . P1k0 1 . . . 0 P21 P22 . . . Pmk A = (I|P)0 0 . . . 1 Pm1 Pm2 . . . Pmk

Where I is a m × m identity matrix and P is the secret parity generation matrix with size m × k. Note that A is derived from a Vander monde matrix, thus it has the property that any m out of the m + k columns form an invertible matrix.

By multiplying **F** by **A**, the user obtains the encoded file: G = F · A = (G(1) , G(2) , . . . , G(m) , G(m+1) , . . . , G(n)) = (F1, F2 , . . . , Fm , G(m+1) , . . . , G(n)),Where G(j) = (g1(j), g2(j), . . . , gl(j))T (j ∈ {1,…,n }).noticed, the multiplication reproduces the original data file vectors of **F** and the remaining part (G(m+1) , . . . , G(n)) are k parity vectors generated based on **F**.

VII. CONCLUSION

In this paper, we investigated the problem of data security in cloud data storage, which is essentially a distributed storage system. To ensure the correctness of users' data in cloud data storage, we proposed an effective and flexible distributed scheme with explicit dynamic data support, including block update, delete, and append. We rely on erasure-correcting code in the file distribution preparation to provide redundancy parity vectors and guarantee the data dependability.

By utilizing the homomorphism token with distributed verification of erasure coded data, our scheme achieves the integration of storage correctness insurance and data error localization, i.e., whenever data corruption has been detected during the storage correctness verification across the distributed servers, we can almost guarantee the simultaneous identification of the misbehaving server(s). Through detailed security and performance analysis, we show that our scheme is highly efficient and resilient to Byzantine failure, malicious data modification attack, and even server colluding attacks. We believe that data storage security in Cloud Computing, an area full of challenges and of paramount importance, is still in its infancy now, and many research problems are yet to be identified. We envision several possible directions for future research on this area. The most promising one we believe is a model in which public verifiability is enforced. Public verifiability, and supported in allows TPA to audit the cloud data storage without demanding users' time, feasibility or resources. An interesting question in this model is if we can construct a scheme to achieve both public verifiability and storage correctness assurance of dynamic data. Besides, along with our research on dynamic cloud data storage, we also plan to investigate the problem of fine-grained data error localization.

REFERENCES

[1] Amazon.com, "Amazon Web Services (AWS)," Online at http://aws. amazon.com, 2008.

[2] N. Gohring, "Amazon's S3 down for several hours"

[3] A. Juels and J. Burton S. Kaliski, "PORs: Proofs of Retrievability for Large Files," *Proc. of CCS '07*, pp. 584–597, 2007.

[4] H. Shacham and B. Waters, "Compact Proofs of Retrievability," *Proc.of Asiacrypt '08*, Dec. 2008.

[5] K. D. Bowers, A. Juels, and A. Oprea, "Proofs of Retrievability: Theory and Implementation," Cryptology ePrint Archive, Report 2008/175, 2008, http://eprint.iacr.org/.

[6] G. Ateniese, R. Burns, R. Curtmola, J. Herring, L. Kissner, Z. Peterson, and D. Song, "Provable Data Possession at Un trusted Stores," *Proc. Of CCS '07*, pp. 598–609, 2007.

[7] G. Ateniese, R. D. Pietro, L. V. Mancini, and G. Tsudik, "Scalable and Efficient Provable Data Possession," *Proc. of SecureComm '08*, pp. 1–10, 2008.

[8] T. S. J. Schwarz and E. L. Miller, "Store, Forget, and Check: Using Algebraic Signatures to Check Remotely Administered Storage," *Proc.of ICDCS '06*, pp. 12–12, 2006.

[9] M. Lillibridge, S. Elnikety, A. Birrell, M. Burrows, and M. Isard,"A Cooperative Internet Backup Scheme," *Proc. of the 2003 USENIX Annual Technical Conference (General Track)*, pp. 29–41, 2003.

[10] K. D. Bowers, A. Juels, and A. Oprea, "HAIL: A High-Availability and Integrity Layer for Cloud Storage," Cryptology ePrint Archive, Report 2008/489, 2008, http://eprint.iacr.org/.

[11] L. Carter and M. Wegman, "Universal Hash Functions," *Journal of Computer and System Sciences*, vol. 18, no. 2, pp. 143–154, 1979.

[12] J. Hendricks, G. Ganger and M. Reiter, "Verifying Distributed Erasure coded Data," *Proc. 26th ACM Symposium on Principles of Distributed Computing*, pp. 139–146, 2007.

[13] J. S. Plank and Y. Ding, "Note: Correction to the 1997 Tutorial on Reed-Solomon Coding," University of Tennessee, Tech. Rep. CS-03-504, 2003.

[14] Q. Wang, K. Ren, W. Lou, and Y. Zhang, "Dependable and Secure Sensor Data Storage with Dynamic Integrity Assurance," *Proc. of IEEE INFOCOM*, 2009.

[15] R. Curtmola, O. Khan, R. Burns, and G. Ateniese, "MR-PDP: Multiple-Replica Provable Data Possession," *Proc. of ICDCS '08*, pp.411–420,2008.

[16] D. L. G. Filho and P. S. L. M. Barreto, "Demonstrating Data Possession and Uncheatable Data Transfer," Cryptology ePrint Archive, Report 2006/150, 2006, http://eprint.iacr.org/.

[17] M. A. Shah, M. Baker, J. C. Mogul, and R. Swaminathan, "Auditing to Keep Online Storage Services Honest," *Proc. 11th USENIX Workshop on Hot Topics in Operating Systems (HOTOS '07)*, pp. 1–6, 2007.