Volume 4, Issue 7, July 2014 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

Providing Security for Data Storage in Cloud through Third Party Auditor

Ananda S. Hiremath
Department of CSE
B.L.D.E.A's CET,
Bijapur , India

Shivaputra S. Panchal
Department of CSE
B.L.D.E.A's CET,
Bijapur , India

Shriharsha S. Veni
Department of CSE
B.L.D.E.A's CET,
Bijapur India

Abstract— In the Cloud computing environment understanding the Security, Privacy and Trust challenges is to advise on policy and other interventions which should be considered in order to ensure that the users of cloud environments are offered appropriate protections, and to underpin a world-leading cloud ecosystem. To ensure the security, privacy and trust we propose publicly auditable data storage. With public auditability, a trusted third party can be delegated as an external audit party to assess the risk of outsourced data on behalf of owner as and when needed. Such an auditing service not only helps save data owners' computation resources but also promise potential cost savings for businesses by offering remote, scalable computing resources. We describe approaches and system requirements that should be brought into consideration, and draw challenges that need to be determined for providing data storage security in cloud computing using the third party auditor.

Keywords— security; privacy; trust; third party auditor; authentication

I. INTRODUCTION

Cloud computing is a model for enabling convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, services) that can be rapidly provisioned and released with minimal management effort or service provider interaction [1].
The literature identifies four different broad service models for cloud computing [2]:

A. *Software as a Service (SaaS)*, where applications are hosted and delivered online via a web browser offering traditional desktop functionality.
B. *Platform as a Service (PaaS)*, where the cloud provides the software platform for systems (as opposed to just software).
C. *Infrastructure as a Service (IaaS)*, where a set of virtualized resources, such as storage and computing capacity, are hosted in the cloud; customers deploy and run their own software stacks to obtain services.
D. *Hardware as a Service (HaaS)*, where the cloud provides access to dedicated firmware via the Internet.

Cloud computing has been visualized as the next generation architecture of the IT enterprise due to its long list of unprecedented advantages in IT:

On demand self-service: A consumer can unilaterally provision computing capabilities, such as server time and network storage, as needed automatically without requiring human interaction with each service's provider.
***Ubiquitous network access*:** Capabilities are available over the network and accessed through standard mechanisms that promote use by heterogeneous thin or thick client platforms.
Location-independent resource pooling: There is a sense of location independence in that the customer generally has no control or knowledge over the exact location of the provided resources but may be able to specify location at a higher level of abstraction (e.g., country, state, or data center). Examples of resources include processing, storage, memory, virtual machines and network bandwidth.
Rapid resource elasticity: Based on the user need he can extend his storage space in cloud which is easier than extending the hardware resources on his own system.
Usage-based pricing: Owner needs to pay for what he uses, no need to buy in advance and no contract.
Transference of risk: Here Owners transfers his risk of storing and maintaining the data to TPA i.e., Third Party Auditor [2]. Before dealing with security, privacy and trust in the cloud, it is important to define these terms because their usage can change radically differ in different contexts. Security concerns the confidentiality, availability and integrity of data or information. Security may also include authentication and non-repudiation [3].

Privacy concerns the expression of or adherence to various legal and illegal norms. In this context this is often understood as compliance with data protection regulations regarding the right to private life. This is often understood as compliance with data protection regulations. Although it would be highly complex to map cloud issues onto the full

privacy and personal data protection regulatory architectures, the globally accepted privacy principles give a useful frame: consent, purpose restriction, legitimacy, transparency, data security and data subject participation [4]. Trust revolves around 'assurance' and confidence that people, data, entities, information or processes will function or behave in expected ways. Trust may be human to human, machine to machine (e.g., handshake protocols negotiated within certain protocols), human to machine (e.g., when a consumer reviews a digital signature advisory notice on a web site) or machine to human (eg, when a system relies on user input and instructions without extensive verification). At a deeper level, trust might be regarded as a consequence of progress towards security or privacy objectives.

II. PROBLEM STATEMENT

The network architecture for providing data security in cloud is as shown in figure1.

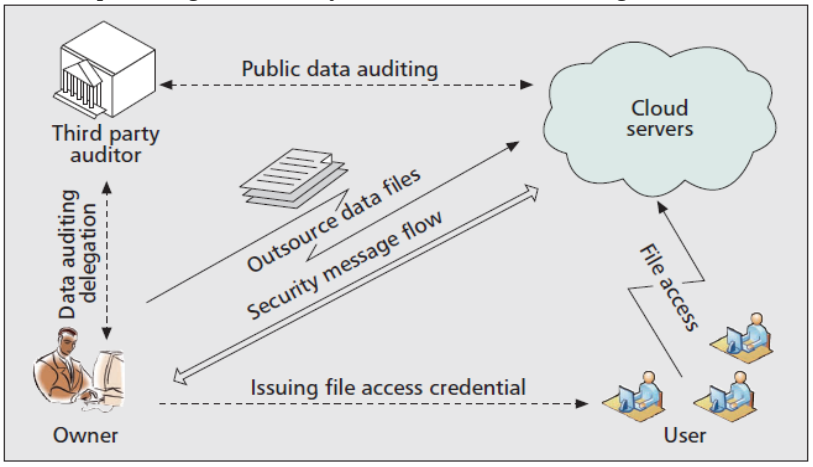

Figure 1: The cloud data storage architecture.

The different network entities can be identified as follows:

Owner: an entity, who has data to be stored in the cloud and relies on the cloud for data storage and computation, can be either enterprise or individual customers.

Cloud Server (CS): an entity, which is managed by cloud service provider (CSP) to provide data storage service and has significant storage space and computation resources (we will not differentiate CS and CSP hereafter.).

Third Party Auditor (TPA): an optional TPA, who has expertise and capabilities that users may not have, is trusted to assess and expose risk of cloud storage services on behalf of the owner upon request.

User: an entity, who is authorized to access the data stored in the cloud by taking permission of the respective owners of the data [5].

Security can be added at different layers and granularity levels in a cloud storage system. There are several strategies for providing the data storage security. One approach is to directly encrypt individual blocks in a cloud, for instance,

III. DATA SPLITTING TECHNIQUE

It is an approach to protecting sensitive data from unauthorized access by encrypting the data and storing different portions of a file on different servers. When the split is accessed, the parts are retrieved combined and decrypted. An unauthorized person would need to know the location of the servers containing the parts be able to get access to all servers, know what data to combine and how to decode it.

The security of data splitting algorithm is related to key length. Furthermore, it also increases exponentially with the increase in the number of data blocks. However, traditional data protection methods usually adopt symmetrical encryption, such as DES, the security of which merely depends on key length. For a computer capable of processing one million instructions within a second, with the same key length, the decoding time of the splitting encryption method significantly increases with the increase in the number of data blocks, whereas the decoding time of the symmetrical encryption algorithm remains almost the same. In practical analysis, the key length of the algorithm is usually determined as 129 bits. The number of data blocks is 16. Its security is 8 times higher than that of traditional methods, and its reliability is 50 times higher.

Secure data storage in cloud computing is realized on the basis of a distributed system. After reaching the cloud, data can be randomly stored in any one or more servers. According to characteristics of the storage mode, each server in the distributed system can be abstracted as a storage node.

Suppose there are m servers in the system, written as: $S = \{s_1, s_2, \ldots, s_m\}$.

Suppose the plaintext data set is d. The k equations based on the splitting algorithm is applied to data set d to generate k ($k<m$) data, written as: $\{d_1, d_2, \ldots, d_k\}$ = Partition (d)

in which Partition ( ) is the data splitting algorithm. The generated data blocks are then split, and k servers are randomly chosen out of m servers, which can be expressed as the following formula: $\{ d_1 , d_2 ,.., d_k \}$ = map(S),

where $S = \{s_1, s_2,..., s_m\}$.

The data restoration process can be expressed as $d_1, d_2,\ldots, d_k \bmod p$

where p is a large prime number.

The core of the secure storage strategy is its data splitting algorithm, which is an extension of fundamental theories of k equations in algebra, n congruence surplus principle in elementary number theory, key sharing and online data storage algorithm , through which data splitting storage is realized. The safety of the strategy mainly depends on two aspects. First, is the difficulty of decoding the data splitting algorithm. The second, is that because storage servers are randomly chosen after data splitting, encrypted data cannot be completely obtained by attacking one or more servers, making decoding even more difficult. In addition, the strategy has inherent advantages in its fault tolerance compared with traditional data protection methods. In cloud computing, no assumptions on the robustness of any node in the distributed system can be made. Various unexpected factors can all result in temporary inaccessibility of some nodes or permanent inaccessibility of data. In such a case, traditional data protection means are often powerless. The secure strategy ensures that data can be restored even when some nodes fail, which considerably improves system reliability.

E. *Algorithm of the Secure strategy:*

Data splitting algorithm in cryptography, it is much more convenience for constructing an isomorphic quotient ring as a complex field than algebraic operation when with the same structure. We construct an isomorphic quotient ring with the same structure as complex field Z_p (where p is a large prime number), and a k congruence equation expressed as:

$$x^k + \sum_{i=1}^{k-1} a_{k-i} x^{k-i} + d \equiv 0 \bmod p \quad (1)$$

where d Z_p is the data to be split, $0 \le a_i \le p-1$, and $0 \le d \le p-1$

(Note: d here can be -d). According to the fundamental principle of k equations in algebra, Equation (1) has k roots. These roots are expressed as:

$$x^k + \sum_{i=1}^{k-1} a_{k-i} x^{k-i} + d \equiv 0 \bmod p$$

And $\{r_1, r_2, \ldots\ldots, r_k\} \subseteq C$ (C is a set of complex numbers), the Equation (1) can be rewritten as:

$$\prod_{i=1}^{k} (x - r_i) \equiv 0 \bmod p \quad (2)$$

where $1 \le r_i \le p-1$. These r_i are data blocks generated after the splitting.

Equations (1) and (2) show that d is independent of variable x. Therefore, the following can be generated:

$$\prod_{i=1}^{k} r_i \equiv d \bmod p \quad (3)$$

F. *High efficiency of the algorithm:*

High efficiency of the algorithm is illustrated through two aspects: data splitting and storage process and data restoration process. In the data splitting and storage process, splitting algorithm applied to data set d generates k blocks of data r_1, r_2,....., r_k. Then, these data blocks are stored in a randomly chosen server. In addition, coefficient a_i is stored as backup information. The process mainly includes the following operations:

1. k-1 numbers $r_1, r_2, \ldots, r_{k-1}$ are randomly chosen within the finite field Z_p.
2. $r_k = d\,(r_1 . r_2 \ldots\ldots, r_{k-1})^{-1} \bmod p$ is calculated.
3. Coefficients $a_1, \ldots, a_{k-1}$ is calculated by constructing polynomial p(k), in which p(k) is shown in the following:

$$p(k) = (x - r_1)(x - r_2)\ldots(x - r_{k-1}) \bmod p = x^k + a_{k-1}x^{k-1} + a_{k-2}x^{k-2} + \ldots + a_1 x + a_0 \bmod p$$

From the calculation process above, we can infer that k multiplications, one modular inversion, and the multiplication of p (k) polynomial to k times are needed for the algorithm to generate k blocks of data. Therefore, the time complexity is O (k). For data decoding and restoration, the user retrieves data of each block $R=\{r_1, r_2, \ldots, r_k\}$ from relevant servers according to the locally-stored data position index, and then obtains the plaintext data by calculating

$d = r_1 . r_2 \ldots r_k \bmod p$.

Clearly, time complexity of the decoding process, the same as that of the encryption process, is O(k).

Therefore, execution efficiency of the algorithm, whether in an encryption or decoding process, is rather high—much higher than in asymmetrical encryption algorithm.

IV. USING THE THIRD PARTY AUDITOR

Allowing the public auditability for cloud data storage security is of a very high risk and of a critical importance so that data owners can switch to an external audit party to check the integrity of outsourced data when needed. To securely introduce an effective third party auditor (TPA), the fundamental requirements the TPA should satisfy are as follows:

1) TPA should be able to efficiently audit the outsourced cloud data storage without demanding the local copy of data, and introduce no additional on-line burden to the cloud user i.e. the data owner; 2) The third party auditing process should bring in no new vulnerabilities towards owner data privacy.

To fully ensure the data security and save the cloud users' computation resources, it is of critical importance to enable public auditability for cloud data storage so that the data owners may relax by transferring their risk of data maintenance

to a third party auditor (TPA), who is expert in maintaining the security of data and has capabilities that the data owners do not, to audit the outsourced data when needed. Based on the audit result, TPA could release an audit report, which would not only help data owners to evaluate the risk of their subscribed cloud data services, but also it is beneficial for the cloud service provider to improve their cloud based service platform. In a word, enabling risk of auditing will play an important role for this nascent cloud economy to become fully established; where data owners and users will need ways to assess risk and gain trust in Cloud. (pppafssicc) The notion of public auditability has been proposed in the context of ensuring remotely stored data integrity under different systems and security models. Public auditability allows an external party, in addition to the data owner himself, to verify the correctness of remotely stored data. However, most of these schemes do not support the privacy protection of users' data against external auditors, i.e., they may potentially reveal user data information to the auditors. This drawback greatly affects the security of the outsourced data in Cloud storage. From the perspective of protecting data privacy, the data owners, who own the data and rely on TPA just for the storage security of their data, do not want this auditing process introducing new vulnerabilities of unauthorized information leakage towards their data security. There is a further demand for the outsourced data not to be leaked to external parties. Exploiting data encryption before outsourcing is one way to mitigate this privacy concern. Without a properly designed auditing protocol, encryption itself cannot prevent data from "flowing away" towards external parties during the auditing process. Thus, it does not completely solve the problem of protecting data privacy but just reduces it to the one of managing the encryption keys. Unauthorized data leakage still remains a problem due to the potential exposure of encryption keys.

A. ***Noteworthy points for Public Auditing---***We briefly elaborate a set of suggested properties below that satisfy the design principle.

B. ***Minimize Auditing Overhead*** — The overhead imposed on the cloud server by the auditing process must not outweigh its benefits. Such overhead may include both the I/O cost for data access and the bandwidth cost for data transfer. Any extra online burden on a data owner should also be as low as possible. Ideally, after auditing delegation, the data owner should just enjoy the cloud storage service while being worry-free about storage auditing correctness.

C. ***Protect Data Privacy*** — Data privacy protection has always been an important aspect of a service level agreement for cloud storage services. Thus, the implementation of a public auditing protocol should not violate the owner's data privacy. In other words a TPA should be able to efficiently audit the cloud data storage without demanding a local copy of data or even learning the data content.

D. ***Support Data Dynamics*** — As a cloud storage service is not just a data warehouse, owners are subject to dynamically updating their data via various application purposes. The design of auditing protocol should incorporate this important feature of data dynamics in Cloud Computing.

E. ***Support Batch Auditing*** — The prevalence of large-scale cloud storage service further demands auditing efficiency. When receiving multiple auditing tasks from different owners' delegations, a TPA should still be able to handle them in a fast yet cost-effective fashion. This property could essentially enable the scalability of a public auditing service even under a storage cloud with a large number of data owners.
Some of the core set of requirements to be met by a security monitoring system for clouds are as following,

1. ***Effectiveness:*** the system should be able to detect most kinds of attacks and integrity violations.

2. ***Precision:*** the system should be able to (ideally) avoid false-positives; that is, mistakenly detecting malware attacks where authorized activities are taking place.

3. ***Transparency:*** the system should minimize visibility; that is: cloud provider, data owners, and potential intruders should not be able to detect the presence of the monitoring system.

4. ***Non-subvert ability:*** the host system, cloud infrastructure should be protected from attacks proceeding from a compromised guest and it should not be possible to disable or alter the monitoring system itself.

5. ***Deploy ability:*** the system should be deployable on the vast majority of available cloud middleware and HW/SW configurations.

6. ***Dynamic:*** *Reaction:* the system should detect an intrusion attempt over a cloud component and, if required by the security policy, it should take appropriate actions against the attempt and against the compromised guest and/or notify remote middleware security-management components.

7. ***Accountability:*** the system should not interfere with cloud and cloud application actions, but collect data and snapshots to enforce accountability policies.

V. RELIABILITY

The reliability of the secure data storage strategy depends on the backup data coefficients. When one or more nodes cannot be accessed, the secure strategy can ensure that the data will be restored as long as one of the k nodes can be accessed. However, traditional data storage methods require all the data in the k nodes to be retrieved. The more blocks the data are split into, the poorer the reliability of traditional data storage is. Reliability of the splitting storage strategy to that of traditional data protection methods increases exponentially, with the increase in the number of data splitting blocks. Therefore, the secure storage strategy has tremendous advantages in terms of reliability.

VI. RELATED WORK

We first consider public auditability in their defined "provable data possession" (PDP) model for ensuring possession of data files on untrusted storages. Their scheme utilizes the RSA-based homomorphic authenticators for auditing outsourced data and suggests randomly sampling a few blocks of the file. However, the public auditability in their scheme demands the linear combination of sampled blocks exposed to external auditor. When used directly, their protocol is not provably privacy preserving, and thus may leak user data information to the auditor. It describe a "proof of retrievability" (PoR) model, where spot-checking and error correcting codes are used to ensure both "possession" and "retrievability" of data files on remote archive service systems. However, the number of audit challenges a user can perform is a fixed priori, and public auditability is not supported in their main scheme. Although they describe a straight forward Merkle-tree construction for public PoRs, this approach only works with encrypted data. The design an improved PoR scheme built from BLS signatures with full proofs of security in the security model defined in. Similar to the construction in, they use publicly verifiable homomorphic authenticators that are built from provably secure BLS signatures. Based on the elegant BLS construction, public retrievability is achieved. Again, their approach does not support privacy-preserving auditing for the same reason as propose allowing a TPA to keep online storage honest by first encrypting the data then sending a number of pre-computed symmetric-keyed hashes over the encrypted data to the auditor. The auditor verifies both the integrity of the data file and the server's possession of a previously committed decryption key.

This scheme only works for encrypted files and it suffers from the auditor statefulness and bounded usage, which may potentially bring in on-line burden to users when the keyed hashes are used up. In other related work, the purpose of a partially dynamic version of the prior PDP scheme that uses only symmetric key cryptography. However, the system imposes a priori bound on the number of audits and does not support public auditability. Wang et al. consider a similar support for partial dynamic data storage in distributed scenario. The proposed challenge-response protocol can both determine the data correctness and locate possible errors. In a subsequent work, Wang et al, propose to combine BLS based homomorphic authenticator with MHT to support both public auditability and fully data dynamics. Almost simultaneously, Erway et al. developed a skip lists based scheme to enable provable data possession with fully dynamics support. However, all their protocol requires the linear combination of sampled blocks just as, and thus does not support privacy-preserving auditing on user's outsourced data. While all above schemes provide methods for efficient auditing and provable assurance on the correctness of remotely stored data, none of them meet all the requirements for privacy-preserving public auditing in Cloud Computing, as supported in our result. More importantly, none of these schemes consider batch auditing, which will greatly reduce the computation cost on the TPA when coping with large number of audit delegations.

VII. CONCLUSION

Cloud computing has been envisioned as the next-generation architecture of enterprise IT. In contrast to traditional enterprise IT solutions, where the IT services are under proper physical, logical, and personnel controls, cloud computing moves the application software and databases to servers in large data centers on the Internet, where the management of the data and services are not fully trustworthy. This unique attribute raises many new security challenges in areas such as software and data security, recovery, and privacy, as well as legal issues in areas such as regulatory compliance and auditing, all of which have not been well understood. In this article we focus on cloud data storage security. We first present network architecture for effectively describing, developing, and evaluating secure data storage problems. We then suggest a set of systematically and cryptographically desirable properties for public auditing services of dependable cloud data storage security to become a reality. Through in-depth analysis, some existing data storage security building blocks are examined. The pros and cons of their practical implications in the context of cloud computing are summarized. Further challenging issues for public auditing services that need to be focused on are discussed too. We believe security in cloud computing, an area full of challenges and of paramount importance, is still in its infancy now but will attract enormous amounts of research effort for many years to come.

ACKNOWLEDGMENT

We acknowledge the contribution of our students Shantkumar H. S., Snehit B. G. and Veeresh T. B. of Department of Computer Science and Engineering, B.L.D.E.A's College of Engineering and Technology, Bijapur for their support in implementation of this work.

REFERENCES

[1] C. Wang, Q. Wang, K. Ren, and W. Lou, "Ensuring data storage security in cloud computing", in *Proc. of IWQoS'09*, July 2009.

[2] Anita Kumari Nanda, Brojo Kishore Mishra, "Privacy and Security issues in Cloud Computing", International Journal of Advanced Computer Research, Volume-2, Number-4, Issue-6, December-2012.

[3] Qian Wang, Cong Wang, Kui Ren, Wenjing Lou, Jin Li, "Enabling Public Auditability and Data Dynamics for Storage Security in Cloud Computing", IEEE Transactions On Parallel And Distributed Systems, Vol. 22, No. 5, 2011.

[4] Qian Wang, Cong Wang, Kui Ren, Wenjing Lou, Jin Li, "Toward Publicly Auditable Secure Cloud Data Storage Services", IEEE Network, 2010.

[5] Manoja Nuthaki, "Cloud Data Integrity and Security through third Party Auditing", International Journal of Advanced Research in Computer Science and Software Engineering, Volume 4, Issue 2, February 2014, pp. 1148-1151.

[6] Cong Wang, Qian Wang, Kui Ren, Wenjing Lou, "Privacy Preserving Public Auditing for Data Storage Security in Cloud Computing", 2010.

[7] Hongwei Li, Yuanshun Dai, Ling Tian, Haomiao Yang, "Identity-Based Authentication for Cloud Computing", CloudCom 2009.

[8] Danwei Chen, Yanjun He, "A Study on Secure Data Storage Strategy in Cloud Computing", Journal of Convergence Information Technology, Volume 5, Number 7, September 2010.

[9] Amazon.com, "Amazon web services (aws)," Online at http://aws.amazon.com/, 2009.

[10] Sun Microsystems, Inc., "Building customer trust in cloud computing with transparent security," Online at https://www.sun.com/offers/details/sun transparency.xml, November 2009.

[11] M. Arrington, "Gmail disaster: Reports of mass email deletions," Online at http://www.techcrunch.com/2006/12/ 28/gmail-disasterreports-of-mass-email-deletions/, December 2006.

[12] J. Kincaid, "MediaMax/TheLinkup Closes Its Doors", Online at http://www.techcrunch.com/2008/07/10/ mediamaxthelinkup-closes-its-doors/, July 2008.

[13] Amazon.com, "Amazon s3 availability event: July 20, 2008,"Online at http://status.aws.amazon.com/s3-20080720.html, July 2008.

[14] Pearson, S.; (2009), "Taking account of privacy when designing cloud computing services", 5071532 searchabstract CLOUD '09. ICSE Workshop on Software Engineering Challenges of Cloud Computing, 2009. pp 44, 23-23 May 2009.

[15] Chang, L, Ti ; Chin L; Chang, A.Y.; Chun J, C;(2010), " Information security issue of enterprises adopting the application of cloud computing", IEEE 2010 Sixth International Conference on Networked Computing and Advanced Information Management (NCM),pp 645, 16-18 Aug. 2010.

