A Survey on Various Cloud Storage Schemes for Preserving Privacy and Preventing Data Loss

Ashwitha C Thomas
Department of CSE
SJEC
Mangaluru, India

Dr. Sridevi Saralaya
Department of CSE
SJEC
Mangaluru, India

***Abstract*— Cloud storage is provided as a service where data is managed, maintained and also made globally available to the service requesters. The data uploaded by requesters is managed by Cloud Service Provider (CSP). This leads to separation of ownership and management of data. CSP has free access to such data which is also vulnerable to attack from outsiders. Data stored in cloud is also susceptible to cloud server failure. These circumstances lead to risk of information leakage and data loss. In this paper we review the various methods which address privacy issues and data loss in Cloud computing.**

***Keywords*— Privacy, encryption, cloud service provider, partitioning, cloud sever, fog server**

I. INTRODUCTION

There are a lot of companies which provide cloud Storage as a Service (SaaS), such as Dropbox, Google Drive, iCloud etc. [1]. These companies provide large capacity of storage and various services related to other popular applications such as whatsap, email, iTunes, photo editors etc. which in turn leads to their success in attracting numerous subscribers. In the cloud storage provided as a service there are a lot of security problems which includes data interruption, malicious insiders, denial of service, data loss, data leakage and issues related to shared technology [2]. The privacy problem is more significant among these security issues. Following are the security requirements to reduce the threats to some extent:

A. Confidentiality

Confidentiality can be preserved by preventing illegal access to data [3]. Data uploaded to cloud is out of its owner's control. Therefore it must be taken care so that only an authorized user can access the data and others including the service provider has to be considered as an invalid user. Data owners must be able to utilize all possible services without any data leakage.

B. Access Controllability

Data owners must be able to restrict access to their data which is uploaded into cloud [4]. Unauthorized users must not be allowed to access the data without permissions. Every user should be granted with different access privileges that are appropriate for any data stored in cloud.

C. Integrity

The owners must be able to store their data in cloud trustworthily so that their data is not modified or deleted [5]. They expect to get their complete data when needed without being tampered. In case of any damages, system should be able to recover the data without loss.

D. Encryption

The data stored in cloud has to be encrypted so that the attacker cannot read the data [6]. An authorized user can decrypt the data using the key when needed. Encryption helps protecting privacy of data to a great extent. Identity–Based Encryption (IBE), Attribute-Based Encryption (ABE) and Fully Homomorphic Encryption are the encryption techniques used in cloud computing [7].

Any measure taken up to ensure security of data stored in cloud fails when there is an attack from malicious insider. A malicious insider is an employee of service provider who misuses his/her position to gain illegal access on data [8]. The reason for attack can be any of the following:

- To steal data-by stealing valuable data which sometimes costs millions of dollars, the attacker can generate revenue, for example, WikiLeaks.
- To create controversy- creating controversy and gaining popularity is another motive for attack.
- As a revenge-an employee who has any dissatisfaction with the organization can also get revenge by hacking the server.
- To help – an employee can hack the server and help the organization to identify any security flaws in the system.
- Curiosity- some hackers may not really want to break security rules but are just curious to learn something about the organization.

To ensure security in cloud computing, there are different management controls that are associated with any cloud service. These management controls can be categorized as follows [9]:

- Deterrent Control-which alerts the attackers about the consequences they might have to face if they proceed with the attack.
- Preventive Control-which makes the system less vulnerable to attack by mechanisms including proper authentication.
- Detective Control-which detects the attack and triggers some action to address the issue.
- Corrective Control-which limits the amount of damage by taking necessary actions.

Various authors have discussed privacy leakage issues in cloud storage [10, 11, 12]. The data uploaded by the user to the cloud is managed by CSP. As a result, the physical storage of a user's data is under the control of CSP [13]. The CSP can freely access and search the data stored in the cloud.

Meanwhile the hackers can attack the CSP server to obtain the user's data. The above two cases leads to the danger of information leakage and data loss respectively. Following are a few possible authentication attacks in cloud computing [14]:

- Brute force attack- where all possible combinations of password are applied to break it.
- Dictionary attack-where password is matched with words that are most occurring.
- Shoulder surfing-where the attacker observes user's typing patterns and finds the credentials.
- Phishing attack-where the user is redirected to a fake web page to find the password.
- Key loggers-is a software which keeps track of user activities through cache history.

Traditional secure cloud storage solutions for the above problems usually focus on access restrictions or data encryption. Section II of this paper includes a literature survey on various approaches for addressing privacy and data loss. Section III describes architecture of an efficient three layer cloud storage scheme followed by a comparison of various approaches in section IV and conclusion in section V.

II. LITERATURE SURVEY

Protecting privacy and preventing data loss are the two main issues to be addressed in cloud computing. A lot of research is conducted every year on this domain to come up with the best solution. Most of them are based on different types of encryption schemes. Encryption can address privacy leakage to some extent, but fails when attacked by a person within service provider's domain. Data partitioning is a widely used approach against data loss. This literature survey includes a walkthrough on a few works which addresses privacy and data loss in cloud computing.

A. Privacy

1) Effective Privacy Protection Scheme for Cloud Computing

Chuang et al. proposed an Effective Privacy Protection Scheme (EPPS) for cloud computing to satisfy the user demand privacy requirement and maintain system performance simultaneously [15]. At first, the security degree and performance of various encryption algorithms along with the user's privacy level were analyzed. Then, a suitable security composition is obtained based on the analysis. Composition includes encryption algorithms such as AES, RC4 and Blowfish. Simulation results showed that EPPS fulfils user-demand privacy along with maintaining system performance in different cloud environment. The efficiency analysis of this approach shows that worst security loss here is 46%.

2) Data privacy protection using multiple cloud storages

Zhang et al. proposed a cloud storage privacy protection method involving Bit Split (BS) and Bit Combination (BC) [16]. First, data is split into bits and reassembled to form multiple part-files before being uploaded into various cloud servers. While downloading, these part-files are combined to form original file. This method protects privacy of data along with providing performance improvement when compared with traditional encryption and decryption methods. As the data stored in cloud is not encrypted, information leakage takes place when there is an attack.

3) Privacy protection based access control scheme in cloud-based services

Fan et al. presented an access control system with privilege separation based on privacy protection (PS-ACS) [17]. Users are logically divided into personal domain and public domain. Read and write access permissions are set to users in personal domain. For improved access efficiency, Key-Aggregate Encryption (KAE) is exploited to implement permission for read access. Improved Attribute-based Signature (IABS) is used to determine the users' write access. A hierarchical attribute-based encryption is applied for the users of public domain which addresses single point failure issues. Users' private keys are managed by authorized agencies and there are chances that they can break the trust.

4) Time-series pattern based effective noise generation for privacy protection on cloud

Zhang et al. developed a novel time-series pattern based noise generation strategy for privacy protection on cloud [18]. Initially, probability of occurrences of fluctuations in existing noise that can threat privacy of a user, is investigated. Later, time intervals are generated dynamically using cluster based algorithm. Based on the time intervals generated, corresponding probability fluctuations are determined and forecasting algorithm based on time-series pattern is proposed. Finally, a noise generation strategy can be obtained based on forecasting algorithm. The collaboration of multiple malicious service providers can cause threat to privacy.

5) The modification of RSA algorithm to adapt fully homomorphic encryption algorithm in cloud computing

Sha and Zhu designed an encryption system to achieve fully homomorphic encryption [19]. The system at first identifies if the values of the keys generated during encryption contains prime number. Later a new cryptosystem is constructed by combining it with Pascal's triangle theorem, RSA algorithm model and inductive methods. This cryptosystem meets homomorphic computation of operations such as addition and multiplication on cipher texts. Thus fully homomorphic encryption in cloud computing is satisfied. Less computational efficiency and large key size are the limitations of this approach.

6) Homomorphic cloud computing scheme based on hybrid homomorphic encryption

Song and Wang have developed a hybrid cloud computing scheme which is based on an additively homomorphic Paillier algorithm and a multiplicative homomorphic RSA encryption algorithm [20]. Calculation requests made by the customer is interpreted as a combination of basic sum and product operations and the operands. For each request, an encryption decryption machine which is running on private cloud performs encryption process. The cipher text obtained is then uploaded into the public cloud. In the public cloud all calculations takes place without the knowledge of exact data. Simulation results showed that the scheme is efficient and practical. The drawback of this approach is it supports limited operations.

7) A Privacy-Preserving KNN Classification Algorithm Using Yao's Garbled Circuit on Cloud Computing

Kim et al. proposed a secure and efficient KNN classification algorithm on the encrypted databases [21]. The algorithm focuses at preserving query privacy, data privacy, data access patterns and the resulting class labels from the cloud. Encrypted index scheme is also adopted in this algorithm which helps to improve the performance.

B. Data Loss

1) Using multiple clouds for addressing data loss

Sidharth and Basawraj proposed a multi-cloud storage scheme for preventing data loss in cloud computing [22]. The data is replicated to form multiple copies and each copy is stored in different cloud servers. In this case if the data is lost from one server, it can be recovered from other servers. This approach is not a good choice when concerned with storage optimization.

2) A managed data loss prevention using POC framework

Sharma et al. proposed a Data Loss Prevention (DLP) scheme using a Proof-of-Concept (POC) framework [23]. The system consists of a DLP core which includes functions to address data loss at various levels including storage, network and endpoint and file level. It also includes a DLP manager to handle policy enforcement and security settings. The limitation of this approach is that the encrypted data and data hidden within images cannot be read.

3) Data loss prevention using agents

Carolin and Somasundaram proposed an approach to recover the lost data from the cloud servers [24]. A cloud manager is responsible for managing the virtualization and handling faults. The data is recovered by using Erasure code algorithm, using which the data is initially split into multiple parts, encrypted and stored in data servers. Any changes made to this data can be determined by Artificial Intelligence methods where agents are used. It is inferred from the efficiency analysis that complete data recovery cannot be achieved.

A three layer cloud storage scheme for privacy protection is proposed recently by Wang et al. and its system architecture is discussed in next section [25].

III. SYSTEM ARCHITECTURE

In Three Layer Privacy Preserving Cloud Storage Scheme data is not completely put into cloud server. Instead, it is divided into parts and distributed into cloud server, fog server and the local machine. Data being partitioned and stored separately improves privacy protection as the attacker cannot get the complete data. Also, during any failure with cloud server the data is not completely lost as there is a part of data which is not stored in cloud. Scope of this approach is that a lot of organizations or individuals depending on cloud for storing their data are benefited. Following are the objectives of this scheme:

- Addressing privacy and security of data by distributing the encoded data blocks across three distinct layers.
- Avoiding the data loss to some extent by saving a part of data stored locally when the cloud server fails.

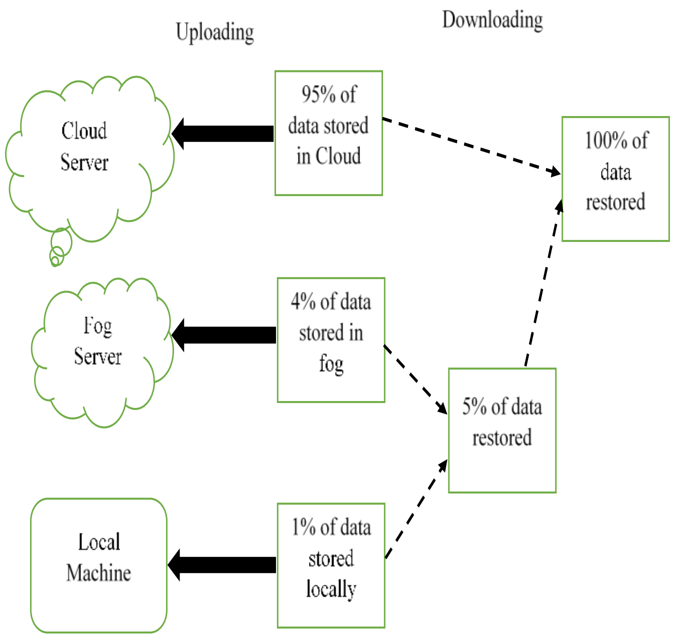

Fig. 1. Basic Architecture of Three-Layer Privacy Preserving Cloud Storage Scheme

Figure 1 shows the architecture of this system. When uploading, data is partitioned and distributed such that major portion is stored in cloud, and a small percentage of it is in fog server and local machine as shown in the figure. Proportion of storage is based on users' allocation strategy. While downloading, these partitions are retrieved and combined together to form the complete original data. The methodology followed is as follows:

- To upload, the data is split into multiple blocks, hash transformed and encoded by using Hash-Solomon algorithm.
- A small portion of the data (1%) is kept in local machine and 99% is sent to fog server.
- In the fog server, 4% of data is retained and 95% is sent to cloud server.
- While downloading, the data in cloud server is fetched and combined with the data in fog server which is then brought together to append to the data in local machine.
- The encoded data is decoded and rearranged using Hash Solomon decoding to get the complete data requested by the user.

IV. COMPARISON OF DIFFERENT APPRAOCHES FOR PRESERVING PRIVACY AND PREVENTING DATA LOSS

TABLE I. COMPARISON OF VARIOUS APPROACHES

Author(s) & Ref.	**Approach**	**Pros**	**Cons**
Chuang et al. [15]	Effective Privacy Protection Scheme for Cloud Computing	Satisfies the user demand privacy requirement. Maintains system performance.	Risk of attack from service provider's domain.

Zhang et al. [16]	Data privacy protection using multiple cloud storages	Addresses both privacy and data loss issues.	Data stored is not encrypted.
Fan et al. [17]	Privacy protection based access control scheme in cloud-based services	Addresses single point failure. Better privacy protection.	Users’ private keys are managed by a third party.
Zhang et al. [18]	Time-series pattern based effective noise generation for privacy protection on cloud	Preserves privacy to a great extent.	Collaboration of multiple malicious CSPs becomes a threat.
Sha and Zhu[19]	The modification of RSA algorithm to adapt fully homomorphic encryption algorithm in cloud computing	Provides homomorphic encryption.	Less efficiency in computation. Key size is very large.
Song and Wang[20]	Homomorphic cloud computing scheme based on hybrid homomorphic encryption	All computation takes place on the cipher texts.	Addition and multiplication are the only operations supported.
Kim et al. [21]	A Privacy-Preserving kNN Classification Algorithm Using Yao's Garbled Circuit on Cloud Computing	Addresses the privacy of databases. Better performance when compared to other classification algorithms.	Risk of attack from service provider’s domain.
Sidharth and Basavraj [22]	Multiple clouds for addressing data loss	Data loss is addressed.	Redundancy of data.
Sharma et al. [23]	A managed data loss prevention using POC framework	Ensures data loss prevention by providing maximum security.	Encrypted data and data hidden within images cannot be read.
Carolin and Somasundaram [24]	Data loss prevention using agents.	Lost data can be partially recovered.	Efficiency analysis shows that complete data recovery cannot be achieved.
Wang et al. [25]	Three-Layer Privacy Preserving Cloud Storage Scheme Based on Computational Intelligence in Fog Computing	Provides better privacy as the data is both encoded and distributed. Addresses data loss to some extents.	Data cannot be fully protected against loss in case of cloud server failure.

V. CONCLUSION

Cloud storage is a very convenient technology that helps the users to store huge data. However, the data stored in cloud is under the control of service providers and there are chances of the data being misused or getting lost. It is inferred from the survey that most common approaches used for preserving privacy and addressing data loss are based on encryptions, restricting access permissions or using multiple cloud servers. The privacy provided by using different encryption methods may not be helpful when there is an attack from inside. In Three Layer Storage Scheme, the attackers cannot get complete data, as the data is split and stored in separate layers. As the data stored is encoded and hash transformed, more privacy can be ensured. Therefore, this approach helps users both in preserving privacy and protecting at least a part of their data in case of cloud failure.

REFERENCES

[1] K.V. K Kumar, “Software as a service for efficient cloud computing”, International Journal of Research in Engineering and Technology, p.10, 2014.

[2] C. Linda Hepsiba and J.G.R.Sathiaseelan, “Security issues in service models of cloud computing”, International Journal of Computer Science and Mobile Comuting, pp.610-615, 2016.

[3] M. R. Gawande and A. S. Kapse, “Analysis of data confidentiality techniques”, International Journal of Computer Science and Mobile Computing, pp.169-175, March 2014.

[4] S. Sudha and V. Arora, “Identity and access management in cloud computing”, International Journal for Research in Applied Science and Engineering Technology, pp. 2321-9653, July 2014.

[5] N. Thakur and A. K. Sharma, “Data integrity techniques in cloud computing: an analysis ", International Journals of Advanced Research inComputer Science and Software Engineering, pp. 2277-128X, August 2017.

[6] L. G. G Patidar, “A survey on cryptographic security over cloud”, International Journal of Advanced Research in Computer Engineering & Technology (IJARCET), Volume 4 Issue 3, March 2015.

[7] K. Alshafee, “Encryption techniques in the cloud”, nternational Journal Of Scientific & Engineering Research, pp. 2229-5518, July-2016.

[8] A. Mahajan and S. Sharma, “The malicious insiders threat in the cloud”, International Journal of Engineering Research and General Science, pp.245-256,March 2015.

[9] B. Mahesh, “Data security and security controls in cloud computing”, International Journal of Advances in Electronics and Computer Science, pp.2393-2835, 2016.

[10] R. Masood and N. Pandey, “Information leakage prevention in cloud computing”, International Journal of Engineering Research and Applications, pp.58-61, November 2014.

[11] Y. Sun, J. Zhang, Y. Xiong and G. Zhu, “Data security and privacy in cloud computing”, International Journal of Distributed Sensor Networks, p.190903, July 2014.

[12] R. R. Kanthe and R. C. Patel, “Data Security and Privacy Protection Issues in Cloud Computing”, International Journal of Computer Science and Information Technology Research, pp. 1130-1134, June 2015.

[13] R. Purohit, “Comparative Analysis of few Cloud Service Providers Considering Their Distinctive Properties”, International Journal of Advanced Research in Computer Science, pp. 1;8(5), May 2017.

[14] A. Singh and D. M. Shrivastava, “Overview of attacks on cloud computing”, International Journal of Engineering and Innovative Technology, pp. 1(4), April 2012.

[15] I.H. Chuang, S.H. Li, K.C. Huang and Y.H. Kuo, “An effective privacy protection scheme for cloud computing”, In Advanced Communication Technology (ICACT), 2011 13th International Conference, IEEE, pp. 260-265, February 2011.

[16] Z. Wei, S. Xinwei and X.Tao ,” Data privacy protection using multiple cloud storages", In Mechatronic Sciences, Electric Engineering and Computer (MEC), Proceedings 2013 International Conference, IEEE, pp. 1768-1772, Dec 2013.

[17] K. Fan, Q. Tian, J. Wang, H. Li and Y. Yang, “Privacy protection based access control scheme in cloud-based services”, China Communications, pp.61-71, Jan 2017.

[18] G. Zhang, X. Liu and Y. Yang, "Time-series pattern based effective noise generation for privacy protection on cloud", IEEE Transactions on Computers, pp.1456-1469, May 2015.

[19] P. Sha and Z. Zhu, “The modification of RSA algorithm to adapt fully homomorphic encryption algorithm in cloud computing", In Cloud Computing and Intelligence Systems (CCIS), 2016 4th International Conference on, IEEE, pp. 388-392, Aug 2016.

[20] X. Song and Y. Wang, "Homomorphic cloud computing scheme based on hybrid homomorphic encryption", In Computer and Communications (ICCC), 2017 3rd IEEE International Conference on, IEEE, pp. 2450-2453, Dec 2017.

[21] H.J. Kim, H.I. Kim and J.W. Chang, "A Privacy-Preserving KNN Classification Algorithm Using Yao's Garbled Circuit on Cloud Computing" , In Cloud Computing (CLOUD), 2017 IEEE 10th International Conference on, IEEE, pp. 766-769, Jun 2017.

[22] G. Sidharth and D. Baswaraj, "Cloud Computing Security from Single to Multi-Clouds" , International Journal of Computer Science and Mobile Computing, pp.57-61, October 2013.

[23] D. H. Sharma, C. A Dhote and M. M Potey, " Managed data loss prevention security service in cloud" , Second International Conference on Future Networks, 2016.

[24] S.P. Carolin and M. Somasundaram, “Data loss protection and data security using agents for cloud environment" , International Conference on Computing Technologies and Intelligent Data Engineering, pp. 1-5,Jan 2016.

[25] T. Wang , J. Zhou, X. Chen , G. Wang , A. Liu , and Y. Li, “A three-layer privacy preserving cloud storage scheme based on computational intelligence in fog computing”, IEEE Transactions on Emerging Topics in Computational Intelligence, pp.3-12, Feb 2018.

[26] R.P. Padhy, M.R. Patra and S.C. Satapathy, "Cloud computing: security issues and research challenges", International Journal of Computer Science and Information Technology & Security (IJCSITS), 1(2), pp.136-146, Dec 2011.