A Review on Confidentiality of the Outsourced Data

B.Divya[1], R.Kalaiselvi[2]

[1] Department of Computer Science and Engineering, Kumaraguru College of Technology, Coimbatore-641049
[2]Assistant Professor, Department of Computer Science and Engineering, Kumaraguru College of Technology, Coimbatore-641049

Corresponding author: B.Divya , divyapriyabsm94@gmail.com

Abstract:

Cloud Computing is a paradigm which supports data storage and access at pay per use mode. The outsourced data may be sensitive and may cause undesired consequences when it is accessed by unauthorized persons. To protect such precious personal data, practically, data owner employs encryption techniques. Revealing data search pattern and access pattern also leads to problem. No information like sensitivity of data, location of data, size of data, etc., shall be revealed to anonymous. Many techniques exist to protect such information where this paper concentrates on few techniques which are employing for content confidentiality, search pattern confidentiality and access confidentiality. Several articles which deal with these concepts are discussed and best methods are explored.

Keywords: Data protection, content confidentiality, search pattern confidentiality, access confidentiality.

1. INTRODUCTION

Now a day, people enjoy the benefits of cloud by utilizing the cost effective services. Cloud services allow their customers to outsource their data in cloud where data can be accessed at anytime from anywhere using any tiny devices like mobile phones, laptops over the internet. When data are left in the cloud, users lose their control over their data. Cloud service providers are responsible for all compute security or cyber security to certain level. Cloud security is a set of policies, technologies and controls deployed to protect the data, applications and the infrastructure of cloud computing. Cloud computing security concerns all the aspects of making cloud computing secure. As many of these aspects are not unique to the cloud setting, cloud security encompasses all the topics of computing security, including the design architecture, minimization of attacks surfaces, protection from malware, and enforcement of access control.

Security concerns associated with cloud computing fall into two broad categories: security issues faced by cloud providers and security issues faced by their customers. Cloud service providers are anticipated to ensure the security for client data and applications while the users are expected to conscious enough to support their application and use strong password and authentication measures. Many tools as Cloud Access Security Brokers (CASB) exist to monitor all the activities between cloud service users and cloud applications such that they enforce security policies.

To face the issue, each enterprise maintains its own identity management system to control access to information and other applications. Service providers integrate the user identity management system into their own infrastructure using federation called SSO (Single-Sign-On) technology or a biometric identification system. It links the confidential information of the users to their biometrics and stores it in an encryption fashion. Making use of searchable encryption technique, potential attackers do not gain access to any sensitive data or even the content of the any queries.

A number of security threats are associated with cloud data services not only traditional security threats, such as network eavesdropping, illegal attack and denial of service attack, but also specific cloud computing threats, such as side channel attacks, virtualization vulnerabilities, and abuse of cloud services[1][2]. Some security requirements may limit the threats: Server Environment which prevents the unauthorized access by changing the data location by swapping or shuffling techniques; Confidentiality that means the protection of personal information. Out sourced data can be accessed only by authorized users. Others including CSPs, should not gain any information about the data; Content confidentiality, that ensures the protection of outsourced data either by using cryptography or other techniques; Access confidentiality hides the retrieval fashion of the data from unknown parties; Search Pattern confidentiality covers the search sequences of user queries.

2. Server Environment

Data are stored in cloud in encrypted format where storage may happen in a single server. Practically, data access follows the same physical location access for a particular data. So, the adversary may try to attack the data by using the access pattern by tracking the history of access sequence. In case of using single server holding the access confidentiality is seems to be difficult[3].

Aggarwal et al. [4] solves the single server problem by distributing the data to two different servers. Also they ensure the data confidentiality by swapping the data among two different servers after the data access. However, when using only two servers, swapping of data among two servers after each access reveals this swapping information 100 %. The data swapping provides the better protection, however the

deterministic reallocation in the case of two server and could then cause exposure in case of collusion. The data distributed among multiple servers may limit the data visibility. Multi-server environment improves the access confidentiality [5].

The use of three servers provides the better production. Swapping ensures that data are moved out from a server at an every access. While going from two servers to three servers provides considerably higher production guarantee, further increasing the number of servers provides limited advantage, while instead increasing the complexity of the system [6,7].

3. **Access confidentiality**

Access confidentiality ensures the data location confidentiality from hackers. Many techniques are available to promise the access confidentiality. De Capitani Di Vimercati et.al used tree structure format for data index where each node having index for easy access of the data. Index contains information about the level of the node and next access node. Data index also gets shuffled after every access of data with data shuffling. This confuses the intruder to conclude the data location or in determining the index [8].

Obfuscation is a programming technique in which code is intentionally obscured to prevent reverse engineering and deliver unclear code to anyone other than the programmer. Data obfuscation is a process of hiding original data into random data or character. Obfuscation techniques break the correspondence between the data and the location of the data. Such a dynamic allocation prevents the server observing sequence of accesses from withdrawing inferences which could compromise pattern confidentiality and even break data confidentiality [5,9].

In traditional tree index, each node in tree index contains pointer to the next level of the node for access. Adversary can easily track the path and find the next level of the node. Shadowing technique is introduced to solve the path revealing problem. In this technique, path information is protected by making observations by each server as if the server was the only one involved in all accesses [8]. The distribution of the shuffle index increases the protection of data content and access patterns by effectively preventing the servers from acquiring knowledge through the observation of all actions followed in servers [10]. From the study of access confidentiality it is concluded that swapping provides a significant guarantee even in the presence of collusion among two or even all three, of the involved servers [11,12].

4 Content confidentiality

Many encryption techniques are used for data protection. Authorization also employs a vital role in data confidentiality. Authentication is a process, where the authorized users are assigned confidential user

identification and passwords to ensure the content confidentiality. Signcryption [12] is a high performance cryptographic primitive that can perform the functions like digital signature and public key encryption. This method provides high security such as confidentiality, integrity, authentication and non-repudiation with a lower cost. Another type of authentication is biometrics. When the data need to be shared only among the authorized users, encrypted data are circulated where re-encryption and decryptions are required for maintaining content confidentiality [13]. Mandatory Access Control (MAC) method is used for security where the information is classified into unique categories and each category will be assigned a particular security degree so that data access is restricted based on the authority level. Role based security methods may be employed to ensure user authorization. Role based access control is based on individual or group of user responsibilities and role in the cloud. In case the user perform different operations, for instance, create, modify personal information. It is based on the user role and also depending on the responsibility and authority within the organization [14,15]. Attribute Based Access Control (ABSC) is also used by many industries where each user will coupled among specific set of attributes. Data proprietor assigns attributes to the distinct user. Whenever the resource not available, the owner provides the permission to access only assigned features.

Homomorphic encryption scheme is widely used to ensure the content confidentiality. Homomorphic encryption is a form of encryption that generates encrypted result where plain text can be derived by applying decryption using same key. The uniqueness of homomorphic encryption is it can employ on encrypted data. This characteristic of homomorphic algorithm supports to have several vital operations like search on encrypted data. This algorithm also supports for secure maintenance of the services that has to be protected to avoid adversary activities. Homomorphic encryption is widely used in applications like voting system, collision-resistant hash function and Private Information Retrieval (PIR) [17]. W. Lu, et al., proposed two homomorphic encryptions: fully homomorphic encryption and partially homomorphic encryption where fully homomorphic cryptosystems have great practical implications in the outsourcing of private computations, for instance in the context of cloud computing [18].

5 Search Pattern Confidentiality

Generally string matching algorithms are used to search and identify encrypted data in cloud. Various search algorithms are used such as Naive string search algorithm, Rabin-karp string search algorithm, and Two-way-string-matching algorithm. While using these algorithms, search patterns must be protected from the adversary. Leakage of such access pattern may enable potential privacy attacks against selected data items [19]. Cloud server may also infer a cloud user's activity pattern or private interest by tracking the user's access to particular data item. To strictly protect the privacy of data search pattern, the intention

of every data search operation should be hidden so that observers of the operation cannot gain any meaningful information.

PIR (Private Information Retrieval), publicly accessible databases are an indispensable resources for retrieving up-to-date information. However, accessing such databases also poses a significant risk to the privacy of the users. A curious database operator can follow the user queries and infer what the user is after. The solution of PIR problem enables the user to retrieve a desired data item, while giving each individual database on the query.

By enabling privacy preserving access, search pattern confidentiality can be achieved. RAM (ORAM) is a method that aims to enable privacy preserving access to data stored in the cloud. There is a set of clients and an untrusted server where clients store data in encrypted format. To facilitate search on encrypted data, an encryption index structure is stored in the sever along with the encrypted data. Authorized users have rights to access the trapdoor generation function which generates different cyber text for each user query. Therefore, they can generate valid trapdoors for any arbitrary keyword. This trapdoor is used in the server to search for intended keyword. As the trapdoor generation function can not be accessed by service provider, they cannot ascertain the keyword searched for. However, it is imperative to hide the corresponding keyword of the given query from an adversary. Otherwise, the adversary learns the set of documents that contains the given keyword and the set of documents that does not. A searchable encryption scheme is qualified as secure, if it satisfies the following conditions, query generation function only known to the authorized data users and the search mechanism works only in conjunction with a valid query generation function. These methods are used to preserve the search pattern confidentiality.

6 Conclusion

While the usage of cloud enhances the resource sharing, security threats are addressed with content confidentiality, search and access pattern confidentialities. With a study on the security measures certain existing techniques can be taken for consideration to restrict the data hacking. Three server environment improves the security when compare to single server or two server environment. Leading industries can prevent their data leakage by outsourcing their encrypted data in encrypted format. From these we can certainly identify that each and every access design has different technique to provide a security level. This short article aimed to deliver basics associated with confidentiality techniques as well as its technologies to enhance the cloud security. By underwent various confidentiality models, it is considered to combine trapdoor with three server model to protect the access and search pattern.

References:

[1] Jhawar, R, Piuri, V, and Samarati, P. (2012). Supporting Security Requirements for Resource Management in Cloud Computing, 2012 IEEE *15th Internaltional Conference on Computer Science and Engineering*, p. 170-177.

[2] Ryan, M.D. (2013). Cloud Computing Security: The scientific challenges and a survey of solutions, *Journal of System software*, Vol. 86(9), p. 2263-2268.

[3] Ostrovsky, R. and Skeith, W.E. (2007). A Survey of Single-Database Private Information Retrieval: Techniques and Applications In: Okamoto T., Wang X. (eds) Public Key Cryptography – PKC 2007. PKC 2007. Lecture Notes in Computer Science, Lecture Notes in Computer Science, vol 4450. Springer, Berlin, Heidelberg p. 393-411.

[4] Aggarwal , G. (2005). Two can keep a secret: A distributed architecture for secure database services, Cidr, p. 186-199.

[5] Stefanov, E. and Shi, E. (2013). Multi-cloud oblivious storage, *Computer & communications security*, p. 247-258.

[6] De Capitani, S. Foresti, S. Paraboschi, S. Pelosi, G and Samarati, P. (2015). Three-server swapping for access confidentiality, *IEEE Transactions on Cloud Computing,* Vol. 7161(C), p. 1-14.

[7] De Capitani Di Vimercati, S. Foresti, S. Paraboschi, S. Pelosi, G. and Samarati, P. (2015). Protecting access confidentiality with data distribution and swapping, Proceedings of 4th International conference on Big data cloud computing, International conference on comuting networking, Vol.1, p. 167-174.

[8] De Capitani, S. Foresti, S. Paraboschi, S. Pelosi, G. and Samarati, P. (2013). Distributed shuffling for preserving access confidentiality, *Lecture Notes in Computer Science*, Springer, Berlin, Heidelberg, Vol. 8134. p. 628-629.

[9] Arockiam, L. and Monikandan, S. (2014). Efficient cloud storage confidentiality to ensure data security, *2014 International conference on computers and communication informatics Ushering Technology*, p. 1-5.

[10] De Capitani Di Vimercati, S. Foresti, S. Paraboschi, S. Pelosi, G. and Samarati , P. (2013). Supporting concurrency and multiple indexes in private access to outsourced data, *Journal of Computer Security- research in Computer Security and Privacy: Emerging Trends*, Vol. 21(3), 452-461.

[11] Bindia, (2016). Enhancing Security through data swapping and shuffling across the servers in cloud, *International Journal of Emerging Technologies in Engineering Research*, Vol. 4(5), p. 169-171.

[12] Mivule, K. (2017). Data swapping for private Information Sharing of Web search logs, *Procedia Computer Science*, Vol. 114, p. 149-158.

[13] Li, F. Liu, B. and Hong, J. (2017). An Efficient Signcryption for data access control in cloud computing, *Journal of Computing*, Vol. 99(5), 465–479.

[14] Aluvalu, R. Muddana, L. (2015). A Survey on Enhancing Cloud Security through Access Control Models and Technologies, *Advances in Intelligent Systems and Computing*, Vol. 337,

653-664.

[15] Langaliya, C. and Aluvalu, R. (2015). Enhancing Cloud Security Through Access Control Models- A Survey, *International Journal of Computer Applications*, Vol. 112(7).

[16] Hu, V. C. Kuhn, D.R. and Ferraiolo, D.F. (2015). Attribute – based access control, Computer , Vol. 48(2), 85-88.

[17] Lu, W. Varna, A. L. and Wu, M. (2014). Confidentiality-preserving image search: A comparative study between homomorphic encryption and distance – preserving Randomization, IEEE Access, Vol. 2, 125-141.

[18] Yang, K. Zhang, J. Zhang, W. and Qiau, D. (2011). A light – weight solution to preservation of access pattern privacy in un-trusted clouds, Lecture Notes in Computer Science, Vol. 6879, 528-547.

[19] Islam, M.S. Kuzu, M. and Kantarcioglu, M. (2012). Access pattern disclosure on searchable encryption: Ramification, attack and mitigation, Network and Distributed System Security Symposium, Vol. 20, 1-15.