A Novel Approach for Data Sharing in The Cloud

Rishika K K[1] & Damodaran V[2]
[1]*Department of Computer Engineering, School of Engineering, CUSAT, Kochi, Kerala, India.*
[2]*Associate Professor, Division of Computer Engineering, School of Engineering, CUSAT.*
[1]*rishikakk@cusat.ac.in*
[2]*damodar@cusat.ac.in*

Abstract: *Cloud computing has opened up infinite possibilities for businesses and other related organizations in the last ten years in conjunction with the development in areas of Bigdata, IoT, etc. Parallel to this enormous growth there arise security aspects as well. This paper proposes a reliable yet strong method for attaining data security while sharing data among the entities in the cloud environment without any reduction in the data utility. It also proposes a fool proof method to identify any insider who has caused data leakage.*

Keywords*:* Cryptography, Watermarking, Timestamp, Data security, Cloud computing.

1. INTRODUCTION

Cloud computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources that can be rapidly provisioned and released with minimal management effort or service provider interaction [1]. In the cloud paradigm, a pool of computing resources like hardware, CPU, storage, software, etc. are managed by cloud service providers. They offer services in the form of infrastructure, platform, and software to the consumers, as per their requirement, over the Internet. The features of the cloud services like no initial capital investment, pay-per-use, fast deployment, rapid provisioning, scalability, low-cost disaster recovery, etc. resulted in its rapid adoption by enterprises.

As the possibilities that the cloud computing opens up in endless so are the chances of data breaches. Cloud security, also known as cloud computing security, consists of a set of policies, controls, procedures and technologies that work together to protect cloud-based systems, data and infrastructure. To prevent data leakage various methods are made use of to design data leakage prevention systems in organizations. Preventing data leaks through strict security policies and access rights is widely implemented by many organizations. Conversely, limiting the sharing of information in order to maintain security results in reduced data utility which may affect the performance of the organization. So, it is also necessary to keep in check that these measures don't inversely affect the usability and availability of data to the legitimate user in time of need.

This paper proposes a novel approach for secure data sharing in the cloud by using a hybrid combination of cryptography and watermarking based on the sensitivity of the data stored in the cloud. It also proposes a method to detect the person who is responsible for leaking the data in a public domain by making use of a unique identifying information, created using a combination of the details during the request for that particular file, about the client who requests the data. Thus, it provisions data leakage detection.

2. LITERATURE REVIEW

Vaquero et al. [2] described issues related to IaaS: virtualization, network, and physical domains. They reviewed threats associated with these three areas and finds access control and encryption techniques as the major solutions for ensuring the security. Even though

access control mechanisms can be used for security purposes they degrade the utility of the services.

Kao et.al.[3] proposed a cryptographic method involving user-centric key management scheme. It secures the keys by storing private keys on user's mobile devices and presents through 2D barcode to control the information leakage.

A content based fingerprinting approach is presented to preserve the information from unauthorized revelation [4]. Leakage is detected by extracting the patterns from the document and then matching these with the outgoing documents. The benefit of the scheme is that it does not reveal the entire sensitive document to the intermediate party, instead, only a set of keywords are exposed. A major disadvantage of the approach is that it becomes unable in detecting the leakage even when slight changes are performed in transmitting data.

To identify the guilty agent, [5] preferred to choose probabilistic approach which implements a variety of data distribution strategies. This improves distributor's chances of identifying leaker. The limitation of the probability method is that it cannot identify the exact leaker responsible for leaking the data.

Nagle et.al.[6] proposed, an improved DES algorithm that uses 112-bit key size for data encode and decode. For data encoding purpose two keys are used. 128-bit input of DES algorithm is divided into two parts. That two parts are executed at a same time. DES algorithm has one weakness of less key size. 3DES algorithm needs large amount of time for encryption and decryption. Improved DES algorithm has capability to provide better performance as compared to DES and 3DES.

Jeeva et.al.[8] provided comparative analysis of encryption algorithms. Comparisons were based on parameters such as encryption ratio, speed, tunability, power consumption, hardware software implementations and key length. Conclusively among the symmetric encryption techniques, the Advanced Encryption standard AES is reviewed as more preferable solution because of less energy consumption and buffer usage, and decrease in encryption and decryption time.

3. PROPOSED SYSTEM

Addressing the security issues in cloud data sharing environment we propose a layer-based security scheme which provides enough protection to data based on the level of sensitivity to which it belongs to. The attacks that can occur in this model are: a) To lose the confidentiality of data by leaking it. b) To make the system incapable of identifying the malicious entity.

Accordingly, data in the server can be classified as public(P), private(P') and confidential(C) as per the content in them. In our system the entities in the cloud environment are the cloud administrator, data owner and the client. The client can request the data which he wants to look up and simultaneously his details viz. client id and timestamp along with the location of the document in the server, for which he has requested is stored in a Request Management Table (RMT) maintained by the cloud server. A Data Management Table (DMT) is also maintained which meets the purpose of storing the information regarding the location of various chunks of data, this is explained in detail in the coming sections. This comes needy to track any malicious activity. The various levels of security measure adopted are level-1 for public data, level-2 for private data and level-3 for confidential data which are discussed below. Various cryptographic methods authentication and watermarking schemes are combined in a hybrid way to

ensure maximum security. A detailed system architecture is shown in the Figure 1 given below.

Level-1: Public data is made accessible to the client without much cumbersome steps. The user needs to just login with his registered username and password and request for the document. Therefore, only the authentication phase is done in level-1 which makes the - data available to the authentic user without any delay and thus providing maximum data utility.

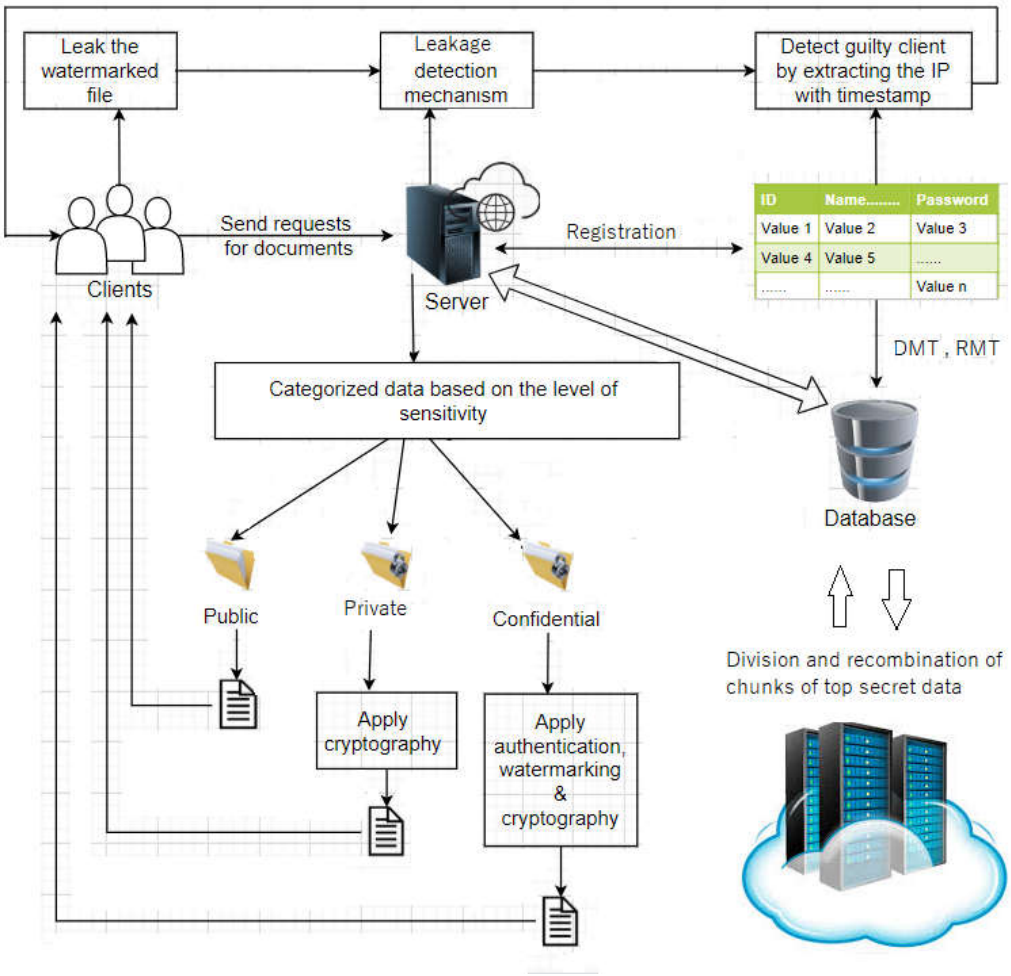

Figure 1: System Architecture

Level-2: Private data has more stronger protection mechanisms. As for preserving the integrity of the data and maintaining data security it is first encrypted using AES algorithm. As AES is symmetric key algorithm the same key is used for encryption and decryption of the file. This key is further encrypted using RSA and sent along with the encrypted file to the receiving client.

The RSA private key is made available to the client as a QR-code sent to his registered email. In the process of decryption, the client has to make use of the RSA public key and the QR-coded private key decrypt the AES private key first, which should in turn be used to decrypt the received file. This mechanism thus ensures the security of data in transit on

its way to the client. The encryption process of this level is illustrated in the Figure 2 given below.

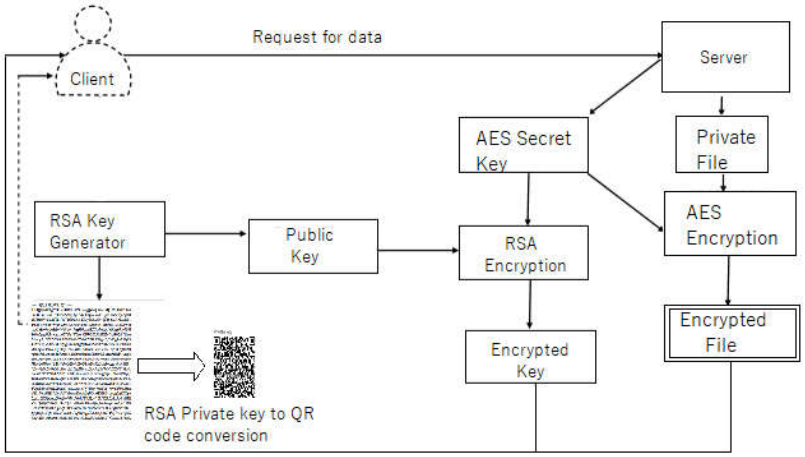

Figure 2: Encryption phase of Level-2

Level-3: Confidential data can be divided into smaller pieces and stored at different locations in different servers, also a hash value of each of these chunks can be taken and stored along with the id for the corresponding chunk and its location details in the Data Management Table (DMT) maintained by the cloud administrator. At the time of request for the file these chunks can be recombined and encrypted before sending the entire document to the client. Before recombining the chunks, their hash values can be compared to the ones stored in the DMT so as to ensure the integrity. Any change in the hash value indicates the loss of integrity and points at some security breach. The encryption mechanisms adopted in the level-2 is similarly done in the level-3 also, in addition to which an invisible watermark is also embedded in the requested document. The identification details of the client whose request is granted by the data owner is embedded in this watermark.

The identification information involves the client id, IP address along with the timestamp, captured when the request for the file is sent, concatenated and encrypted using Blowfish algorithm. This encrypted information lies hidden in the Confidential file as an invisible watermark which gives us the clue to identify the guilty party in case of any data leakage in public domain. A hash value of the encrypted information is also stored in the RMT.

Another feature of this level is that client is verified using two factor authentications while the file is made available to him. That is the client is required to enter an OTP sent to his registered mobile number along with the QR coded RSA private key. If the OTP entered is wrong, even if the private key scanned is legitimate, the client is denied the permission to view the content. Once the prescribed time limit is crossed the OTP expires. As the OTP is valid only for a limited time period, it ensures that only the single client who has requested the file is able to use it rather than leaking the information.

The data leakage detection phase comes needy in case any client leaks the confidential file in the public domain. In such cases the watermark is extracted to gain the encrypted identifying information. By comparing the values of IP, timestamp of the client in the watermark and their hash values with that in the RMT the client responsible for the data leakage could be found out. An equivalency of both the hash values confirms the person

guilty of this malicious activity. The result of this data leaker identification process as implemented in the prototype of this system is shown in Figure 3.

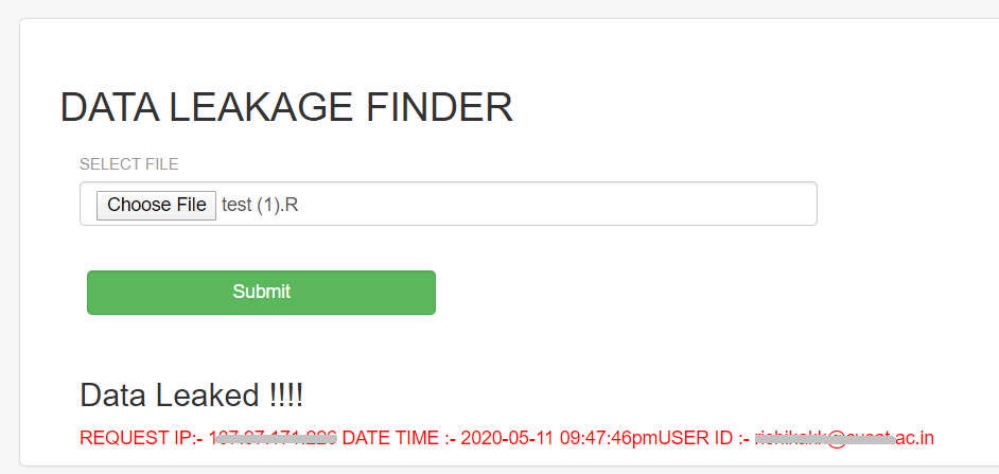

Figure 3: Data leakage detection implementation result

The experiments were performed on system equipped with an Intel® Core™ i5-8250U CPU @ 1.60 GHz 1.80GHz having 8 GB RAM. Implementation is performed using various machines with similar configurations acting as clients and servers. Our prototype system is implemented in PHP using MySQL. Combination of AES and RSA was used for encrypting the data. We construct required hash functions by using robust SHA 3 512-bit technique. Any other robust cryptography, hashing and watermarking schemes that fulfill the purpose can be used.

4. CONCLUSION

In this paper cloud data sharing related security issues are solved using cryptography and watermarking techniques. At the three layers classified according to data sensitivity, hashing, encryption and watermarking techniques are innovatively combined together in a hybrid manner to obtain a better balance between information security and utility. Data leakage detection model is very beneficial in various industries, organization, research communities or an institution who tend to share their data with any third party online. Data leakage detection system is extremely advantageous as it can offer security to the data throughout its transmission.

Acknowledgments

The authors acknowledge the students and staff of Division of Computer Engineering, CUSAT, Kochi.

REFERENCES

[1] *P. Mell and T. Grance, The NIST Definition of Cloud Computing, September, 2011, National Institute of Standards and Technology (NIST), Information Technology Laboratory.*

[2] *L.M. Vaquero, L. Rodero-Merino, D. Morán, Locking the sky: a survey on IaaS cloud security http://dx.doi.org/10.1007/s00607-010-0140-x.*

[3] *Y.-W. Kao, K.-Y. Huang, H.-Z. Gu, and S.-M. Yuan, "uCloud: a user- centric key management scheme for cloud data protection," IET Inform. Security, vol. 7, no. 2, pp. 144-154, 2013, DOI: 10.1049/ietifs.2012.0198.*

[4] Peneti, Subhashini, and B. Padmaja Rani. "Data leakage prevention system with time stamp." Information Communication and Embedded Systems (ICICES), 2016 International Conference on. IEEE, 2016.

[5] Y. Shapira, B. Shapira, and A. Shabtai, "Content-based data leakage detection using extended fingerprinting," 2013, arXiv preprint arXiv:1302.2028.

[6] M. Nagle, D. Nilesh, "The New Cryptography Algorithm with High Throughput",IEEE, ICCCI ,pages 1-5,January 2014.

[7] AL. Jeeva, Dr. V. Palanisamy and K. Kanagaram, "Comparative Analysis of Performance and Security Measuresof Some Encryption Algorithm", International Journal of Engineering Research and Applications*, vol. 2, no. 3.*

[8] N. Sharma ,A.Hasan, "A New Method Towards Encryption Schemes (Name-Based-Encryption Algorithm)",IEEE, International Conference on Reliability, Optimization and Information Technology,pages 310-313,Feb 2014.

[9] Panagiotis Papadimitriou and Hector Garcia-Molin, "Data Leakage Detection", *IEEE Transactions on Knowledge and Data Engineering*, *vol. 23, no. 1, 2011.*

[10] Mingi Zhou et al., "Security and Privacy in Cloud Computing: A Survey," Proc. 6th Int'l Conf. Semantics, Knowledge and Grids, IEEE Press, 2010, pp. 105–112.