ISSN:2348-2079

Volume-5 Issue-1

International Journal of Intellectual Advancements and Research in Engineering Computations

Cloud data redundancy check mechanism for secure integrated public clouds

[1]T. Dhanur Bavidhira, [2]C. Selvi

PG Scholar[1], Assistant Professor (Sl.Gr)[2]

Velalar College of Engineering and Technology[1,2]

dhanur14@gmail.com

Abstract **- Cloud computing is an evolving technology that has attracted more number of users in the past decade because of its striking features of data accessing from anywhere at any time, low data maintenance and pay per use. Also, it has brought the minds of IT giants and clients from different fields to keep their information in the cloud storage. Due to this cloud is experiencing huge amount of data which has led to the cropping of new issues in storage availability. To overcome the above problem the proposed system uses a novel mechanism called Data Redundancy Check (DRC) mechanism by which single-instance of data are stored without compromising on the security of data in the cloud. This system integrates the multiple public clouds as a single integrated system which gives an assurance in data storage integrity. The DRC mechanism is applied to the data before they are going to be stored in integrated multiple public clouds environment in order to avoid the redundant copies of data. This demonstrates that storing single instance of data based on the content can discard the redundant copies of data in the multiple public clouds efficiently. This proposed system gives nominal overhead on comparing to the existing operations.**

Index Terms— Cloud Computing, Data Redundancy, Integrated Public Clouds, Secure Storage

I. INTRODUCTION

In the present day cloud computing has become an integral part in most of the people's lives. Moving to the cloud, running in the cloud, storing in the cloud, accessing from the cloud have become the buzz words in networking and computing technologies which seems like everything is happening "in the cloud". Some realizes life would be very unusual in devoid of cloud. Without it there would be no Gmail, Facebook and Twitter. With the advancements in the network and computing technologies huge amounts of data are being produced. At the same time preserving data confidentiality, integrity and availability has become a mandate. As the data are growing at higher rate and doubling in short duration, management of these data has now become an intelligent task. Research shows that 75% of digital world is a copy and 90% of data are duplicated in backup data sets. Handling these duplicate data in a proper way can sufficiently increase the storage space in the cloud servers. This will lead to reduction in the cost of hardware and can enlarge the space for storing in cloud. Only way to administrate these redundant data are deduplication. During deduplication there are possibilities for compromising in the confidentiality, integrity and availability of data. Hence secure deduplication is the best possible way to avoid the redundant data without negotiating the confidentiality, integrity and availability of data. It is an intelligent compression technique that eliminates the multiple instances of data for optimizing the storage in cloud.

A. Motivation

In cloud, the number of users is increasing day by day due to which there created a big need for storage space. The International Data Corporation has predicted that by the year 2020 the digital data will exceed 44 Zeta Bytes [15]. This clearly shows that how difficult is to manage the growing needs of the storage. Data Deduplication is a well-known technique for getting rid of redundant copies of data and keeping a single instance of data in storage. It also helps in reducing the storage requirement and network utilization for the users with the constant rise in digital data [16]. The main idea behind in this is that deletion of redundant copies of data and storing only one copy of the data in the cloud. Though this technique appears to be simple and familiar, their real time applicability in the cloud domain seems to be more tedious. Whenever new advancement takes place in the cloud the key attributes of data for cloud storage has to be preserved. Hence secure deduplication is the best possible approach. It reduces the storage space and bandwidth for data uploading.

B. Related work

More and more clients would like to store their data to PCS (public cloud servers) along with the rapid development of cloud computing. New security problems have to be solved in order to help more clients process their data in public cloud. When the client is restricted to access PCS, he will delegate its proxy to process his data and upload them. On the other hand, remote data integrity checking is also an important security problem in public cloud storage. It makes the clients check whether their outsourced data is kept intact without downloading the whole data. From the security problems, a novel proxy-oriented data uploading and remote data integrity checking model is proposed [2] for identity-based public key cryptography: IDPUIC (identity-based proxy-oriented data uploading and remote data integrity checking in public cloud). A formal definition is given for system model and security model. Then, a concrete ID-PUIC protocol is designed by using the bilinear pairings. The proposed ID-PUIC protocol is provably secure based on the hardness of CDH (computational Diffie-Hellman) problem. Based on the original client’s authorization, the proposed ID-PUIC protocol can realize private remote data integrity checking, delegated remote data integrity checking and public remote data integrity checking.

In this paper [8] the proposed system addresses the above said issues by using the concept of data anonymity for anonymous matching of data access request. It guarantees that user get shared access authority of other user’s data without revealing the identity. The proposed system also provides the data auditing functionality to protect the integrity of users shared data.

A new flexible, automated and log based RDPC model has been proposed as ID Based ADIC. [6] ID Based ADIC is Identity Based Automated Data Integrity Checking model for single-cloud storage. The proposed model is based on bilinear pairings and RDPC technique. The approach eliminates certification management with the help of Identity management and additionally provides log management towards data integrity. The model makes client independent from initiating verification request and keeping the track of previous records which reduces client’s time. The principle concept here is to make data integrity checking as a painless job for clients.

The sharing of cloud computing resources like storage, services and applications with other tenants is very risky as they accidentally get other tenants information. Multi-tenancies are considered as an important feature in cloud computing resource utilization. Hence Mohamed

Al Morsy et al., [5] has delivered a system that secures multi-tenancy by the segregation of cloud users.

In this paper [3], proxy provable data possession (PPDP) system is designed to address the above said issues. Based on the bilinear pairing technique an efficient PPDP protocol is designed for the system. Through security analysis and performance analysis, the proposed PPDP protocol is provably secure and efficient.

Garima and Naveen [10], "Triple Security of Data in Cloud Computing", proposed a system for enhancing security in cloud by applying three algorithms namely: DSA (Digital Signature Algorithm), AES (Advanced Encryption Standard) and Steganography. For data encryption, DSA is done for authentication followed by AES for encryption and finally applied Steganography for concealing data within audio file to have maximum security. By applying the algorithms in the reverse order the receiver can decrypt the data but the issues found here is high complexity in time since the algorithms are applied one after the other.

Access control and user authentication is very important in cloud environment. Some related access control schemes has been proposed to solve the security problems in cloud, however, the high computation cost is a crucial factor in practical use. In this paper [4] a novel lightweight identity authentication-based access control scheme is proposed for cloud, due to the adoption of authorized agency to assist the authentication and key distribution. This scheme is proved to be more efficient and practical.

Shirole and Sanjay [9], "Data Confidentiality in Cloud Computing with Blowfish Algorithm", proposes a system with OTP (One-Time Password) for authentication and Blowfish algorithm for encryption. This provides reliability and easiness in storing secure data. Here ciphered data is uploaded in the cloud by applying encryption on the plain text and whenever the data is needed, it is obtained from the cloud and is stored on the system in the plain format. Hence this protects the data internally.

According to Keiko David, Eduardo and Eduardo [7], "An analysis of security issues for cloud computing", this paper discusses the various threat areas that require high level of security. External data storage, using public internet for communication, multi-tenancy, data integration are some of the major risk areas that has been dealt in detail.

In this paper [11], an efficient mutual verifiable provable data possession scheme is proposed which utilizes Diffie- Hellman shared key to construct the homomorphic authenticator. The system designs a Private Verifier which is trusted, stateless and independent of the cloud storage service. It is designated to assess and expose risk of cloud storage services upon request. After verification, Private Verifier can also act as a client to modify the data blocks and retag them. The presented scheme is very efficient compared with the previous PDP schemes, since the bilinear operation is not required.

C. Contributions

In public cloud, this paper focuses on the Cloud Data Redundancy Check Mechanism for Secure Integrated Public Clouds. By using redundancy check mechanism, our proposed DRC mechanism is efficient since the duplicate data are eliminated. DRC is a novel redundancy check mechanism to avoid duplicate data in the integrated public cloud. The formal system model for DRC mechanism is given. Based on the uploading of client's data our mechanism can realize redundancy checking.

D. Paper Organization

The paper is organized below. The formal system model of DRC mechanism is given in Section II. The concrete mechanism and prototype are presented in Section III. At the end of the paper, the conclusion is given in Section IV.

II. SYSTEM MODEL OF DRC

In this section, we give the system model of DRC mechanism. A DRC mechanism consists of four different entities which are described below:

1. Client Access

The Client Access is the first module where the registration or login for entering into the cloud takes place. During the time of registration client will be provided with a key generated by a public key generator based on their identity. At the time of login the clients use their identity based key for accessing into the cloud. Based on the privileges, the client can either view the file or download the file. The client may or may not be a data owner. The data owner has all privileges like downloading the file, sharing data access to any user and uploading the file.

2. Proxy

The second module is Proxy. It is considered as a delegate of Public Cloud Server (PCS). The proxy carries out most of the important works of PCS. This helps the public cloud server to be free from heavy network congestion and traffic. Initially the proxy verifies the authenticity of the cloud users. The primary role of the proxy is to prevent the uploading of multiple copies of data to the public cloud server. It is done by generating a tag for each file to be uploaded in the cloud. It ensures that no two file can have same tag value. When the client tries to upload a file to the PCS, the proxy generates a tag based on the content of the file. It checks for the existence of the same tag. If the file exists in the cloud then there will be a matching tag hence the proxy will not upload the file to the cloud. Instead, it will give the access permission for the client to that existing file in the cloud. Suppose the file to be uploaded has no matching tag then the proxy considers that the file to be uploaded is a unique file and it will be uploaded in the public cloud server. By doing this, storing of multiple copies of data can be avoided completely to the public cloud server.

3. Multiple Public Clouds

The entire client's data are stored in the Public Cloud Server in an encrypted format. Multiple public clouds mean that integrating many public clouds into a single integrate public cloud. When the client wants to download their file they can directly access the cloud for file downloading. Similarly the clients can share their file access to other clients in the cloud. By doing this the storage complexity in the cloud can be well addressed. It also helps in reducing the computational overhead in the cloud.

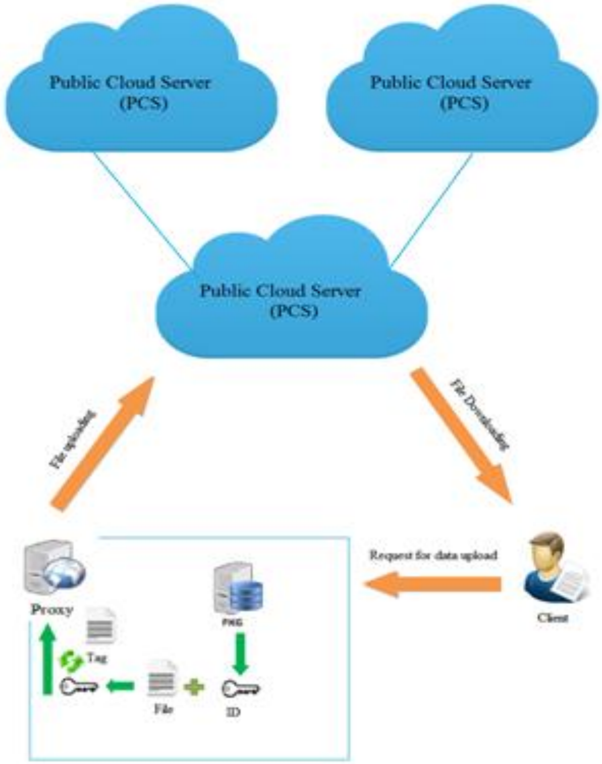

Fig1. Architecture of DRC mechanism

III. THE PROPOSED DRC MECHANISM

In this section, we propose an efficient DRC mechanism for secure data uploading without duplicates and storage service in the public clouds. Hashing technique makes data redundancy check method a practical one. Our mechanism is built on the hashing technique. We first review the hashing technique. Then, the concrete DRC mechanism is designed from the hashing technique.

A. Hashing technique

The key idea is using the hashing technique for each file data. Many hashing algorithms are available for this purpose. Hashing algorithm like SHA- 1 and MD5 are most commonly used hashing algorithm. The result of hashing algorithm is in the form of cryptographic hash form. It is the basis for identifying all the duplicate records in the cloud storage. The benefit of this technique is to have quick execution with low computation and overhead. It shows high efficiency and works well for whole file duplication. This is mainly due to the predominant availability of whole file duplicate records in the storage.

B. Our Concrete DRC mechanism

This DRC mechanism consists of four main procedures: Client authentication, Token generation, Tag generation and File uploading. The DRC mechanism is initiated when the client wants to access the cloud server and the major part of the work is carried out when the client wants to upload the data to the cloud server. The figure2 explains the working of DRC mechanism. The process gets started when the client enters into the public cloud server. First the client has to register himself in the cloud in order to have the authenticity in accessing. Once the client is registered the proxy will generate an identity based key of the client for ensuring higher security to get into the cloud. Based on the interest of the client, he can upload, download or share the file. If the client wants to upload a file into the cloud the DRC mechanism will generate a tag based on the content of the file that the user is trying to upload. Once the tag is generated for the file that is to be uploaded, then our mechanism will look for the existence of the generated tag in the cloud. It is done by checking the generated file tag against the list other file tags that are already in the cloud. The matching of file tag in the cloud makes us to understand that the file to be uploaded is already present in the cloud. On knowing this, the so called file to be uploaded will be considered as a duplicate file and that file will not be uploaded to the cloud. Instead the client will be made to access the existing file in the cloud. By which the duplicate file is avoided completely in the cloud storage. Suppose the file that is to be uploaded doesn't have any matching file tag in the cloud storage then that file is considered to be a distinctive file and then the file will be uploaded.File tag verification is done only when the file is to be uploaded into the cloud.

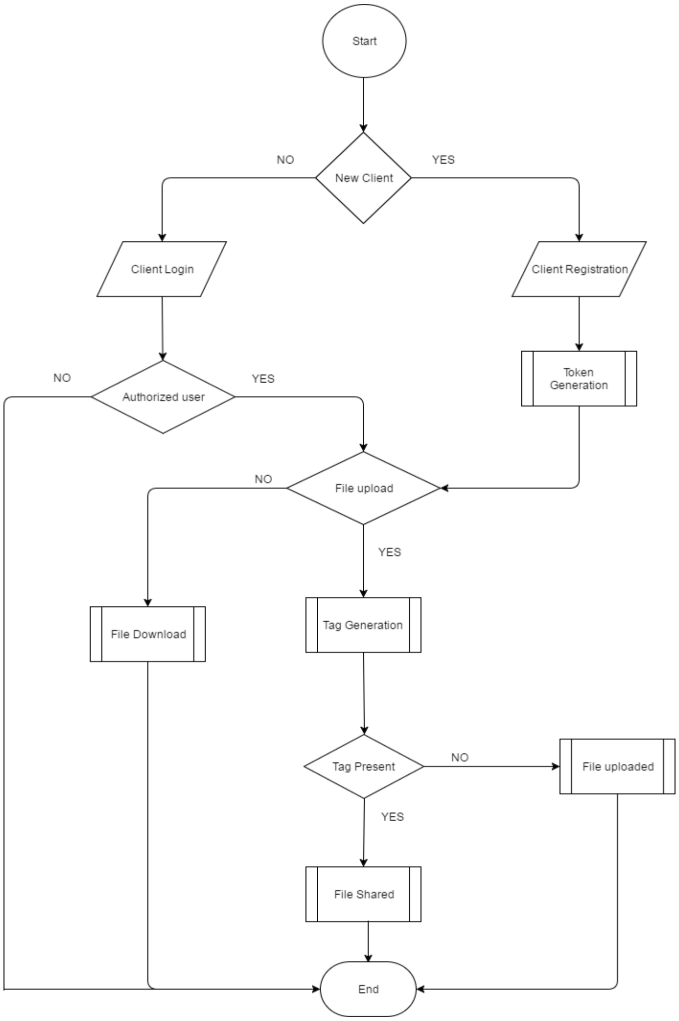

Fig2. Flowchart of DRC mechanism

IV. CONCLUSION

Motivated by the presence of numerous replicated digital data has made this paper to propose a novel idea for secure deduplication in cloud storage. The paper formalizes DRC's system model. Then, concrete DRC mechanism is designed by using the hashing technique. The proposed mechanism identifies and eliminates redundant data among multiple users over multiple integrated clouds without affecting the security of the data. This provides higher duplicate detection and elimination rate thereby decreasing the utilization of storage space and the cost of hardware in cloud. This approach examines the presence of multiple copies of data in cloud before uploading it. Due to which the bandwidth and time of the user is saved at a higher level.

V.REFERENCES

[1] Lan Zhou, Vijay Varadharajan and Michael Hitchens, "Achieving Secure Role-Based Access Control on Encrypted Data in Cloud Storage", IEEE Transactions on Information Forensics and Security, vol. 8, no. 12, 2013.

[2] HuaqunWang, Debiao He and Shaohua Tang, "Identity-Based Proxy-Oriented Data Uploading and Remote Data Integrity Checking in Public Cloud", IEEE Transactions on Information Forensics and Security, vol. 11, no. 6, 2016.

[3] Xuefeng Liu, Yuqing Zhang, Boyang Wang and Jingbo yan, "Mona: Secure Multi-Owner Data Sharing for Dynamic Groups in the Cloud", IEEE Transactions on Parallel and Distributed System, vol. 24, no. 6, 2013.

[4] Cong Wang, Sherman S.M. Chow, Qian Wang, Kui Ren and Wenjing Lou, "Privacy-Preserving Public Auditing for Secure Cloud Storage", IEEE transactions on computers, vol. 62, no. 2, 2013.

[5] Younis A.Younis, Madjid Merabti and Kashif Kifayat, "Secure Cloud Computing for Critical Infrastructure: A Survey", International Journal of Computer Applications, 2013.

[6] Thapliyal, Meenakshi, Hardwari Lal Mandoria, and Neha Garg, "Data Security Analysis in Cloud Environment: A Review", International Journal of Innovations & Advancement in Computer Science, vol. 2, no. 1, 2014.

[7] KeikoHashizume, David G Rosado, Eduardo Fernández-Medina & Eduardo B Fernandez, "An Analysis of Security Issues for Cloud Computing", Journal of Internet Services & Applications, 2013.

[8] Singla and Jasmeet Singh, "Cloud Data Security using Authentication and Encryption Technique", Global Journal of Computer Science and Technology, 2013.

[9] Subhash and Shirole Bajirao, "Data Confidentiality in Cloud Computing with Blowfish Algorithm", International Journal of Emerging Trends in Science and Technology, 2014.

[10] Saini, Garima and Naveen Sharma, "Triple Security of Data in Cloud Computing", International Journal of Computer Science & Information Technologies, 2014.

[11] R. Saranya,G. Indra and Dr. N. Sankar Ram, "Data Compression Technique to Eliminate Duplicates in Cloud Computing", International Journal for Scientific Research & Development (IJSRD), Vol. 3, Issue 03, 2015.

[12] S.Sajithabanu and Dr.E.George Prakash Raj, "Data Storage Security in Cloud", International Journal of Computer Science and Technology, vol. 2, no. 4, 2011.

[13] Yukun Zhou, Dan Feng, Wen Xia, Min Fu, Fangting Huang, Yucheng Zhang, Chunguang Li, "SecDep: A User-Aware Efficient Fine-Grained Secure Deduplication Scheme with Multi-Level Key Management", IEEE Mass Storage Systems and Technologies (MSST) 31st Symposium, 2013.

[14] Cong Wang, Sherman S.M.Chow, Qian Wang, Kui Ren and Wenjing Lou, "Privacy Preserving Public Auditing for Secure Cloud Storage", IEEE Transactions on Computers, vol. 62, no.2, 2013.

[15] "The Digital Universe of Opportunities: Rich Data and the Increasing Value of the Internet of Things", http://www.emc.com/leadership/digital-

universe/2014iview/executive-summary.htm, April 2014, EMC Digital Universe with Research & Analysis by IDC.

[16] Jin Li, Xiaofeng Chen, Xinyi Huang, Shaohua Tang, Yang Xiang, Mohammad Hassan and Abdulhameed Alelaiwi, “Secure Distributed Deduplication Systems with Improved Reliability”, IEEE Transactions on Computers Volume: PP, Year – 2015.