ISSN (e): 2250 – 3005 || Volume, 09 || Issue, 6 || June– 2019 ||
International Journal of Computational Engineering Research (IJCER)

Maintaining Privacy and Overcoming Duplication Overhead in Cloud Computing

Zarka Khan, Mr. Anwar Ahmed Sheikh
Student, Department of Computer Science and Engineering,Integral University, Lucknow, India.
Assistant Professor, Department of Computer Science and Engineering,Integral University, Lucknow, India.
Corresponding Author; Zarka Khan

ABSTRACT
The rapid growth of technology has changed many things around us. If we talk about social networking sites a few years back it was just used as a medium for communication for making new friends etc. If we compare today's current situation from years back we'll notice a huge difference in terms of how these sites maintain their data. Nowadays social networking sites stores there data online. Memory efficiency its management and deduplication overhead are some important challenges in online storage. In our proposed system the main aim is to maintain memory efficiently by not storing the same images again in cloud. There are many deduplication algorithms that are in use. In this paper, a novel approach is proposed which will use bit by bit technique for image comparison on string matching pattern and the comparison will be done in a particular zone. RSA will be used for encryption and key generation technique.
Keyword- *Cloud Storage, Cloud Computing, Data Redundancy, Data Deduplication, RSA Algorithm, MPODO.*

Date of Submission: 26-05-2019 Date of acceptance:08-06-2019

I. INTRODUCTION

Social media are the sites where trillions of users connect with each other and share their contents with each other. Due to great success of social media sites as number of users connecting with it are increasing at a huge rate and large number of data is shared every second. As data is increasing day by day on social networking sites it has become important to provide users with easy environment so that users can share their contents more efficiently and memory can be used efficiently avoiding data redundancy. Cloud technology is changing the way of seeing storage needs and computational powers. Cloud technology in social sites is providing more flexibility to share data as well as it uses deduplication technique to overcome data redundancy and maintain memory efficiently. Once data deduplication is done it can reduce the communication and storage overheads in the cloud storage services. There are many data deduplication schemes but they can either resist brute-force attacks or ensure the efficiency and data availability, but cannot satisfy both the conditions. There is no such existing scheme which achieves accountability, in the sense of reducing duplicate information discloser. Henceforth, three-tier-cross-domain architecture is used that can maintain privacy and overcome duplication overhead in cloud storage (hereafter referred to as MPODO). MPODO achieves deduplication along with both privacy-preserving and data availability, and resists brute-force attacks. Although, data deduplication provides a lot of benefits, there are some security and privacy concerns as the sensitive data is susceptible to many kinds of attacks. So, to overcome these issues a referential key generation technique called RSA is also included in this approach. The most important thing to be taken into consideration is accountability to offer better privacy assurance in comparison to other existing schemes. MPODO can outperform existing competing schemes, in terms of computation, communication and storage overheads. The time complexity of duplicate search over cloud in MPODO is logarithmic. In this the comparison of the image is done in a particular zone.

1.1. Cloud Computing

In cloud user is not aware about where there data is going to be stored. Cloud is basically defined as a term which can be stored anywhere in the world. Cloud Computing refers to accessing, configuring and manipulating the software and hardware resources remotely. It provides user with online infrastructure, data storage and application. Cloud supports pay per use facility and platform independency, as software need not be installed

locally on the user's PC. Data stored should be encrypted form to provide security. Cloud computing is as important as water and air which is changing trend very rapidly day by day. It is providing more flexibility to share data on social sites and is changing the way of seeing storage need.

1.2. Cloud Storage

Cloud storage can be defined as the storage of data online in cloud. It provides facilities like large reliability and accessibility; rapid deployment; security for data backup, archival and disaster recovery purposes; and the overall cost is decreased but of the pay per use facility and can easily manage and maintain expensive hardware. A cloud storage system needs one data server connected to the internet. A client which is a computer user who has subscribed to a cloud storage service sends the copies of image to the data centre over the internet which records the information. Whenever the client wants to the information, they access the data server through a Web-based interface. The server then either sends the files back to the client or allows the client to access and make changes in the files on the server itself. Cloud storage is not just only for those clients who are running out of storage space; it's also used to create backups of data. There are large numbers of cloud storage providers on the Web, and more are popping up all the time. Apart from providing storage, the amount of storage each company offers to its clients is also growing rapidly. Cloud storage is increasing rapidly in our big data driven society and from all the survey that has been conducted it is concluded that mostly 75% of the data are identical due to which the data redundancy in cloud storage is more than 90%. Data deduplication techniques has been used to overcome this problem but the major problem arises when the data stored on cloud is in encrypted form because identical data encrypted by different user have different cipher text. to overcome this problem we can use MPODO which can preserve privacy and data availability along with resisting brute force attack. The privacy of the user can be maintained by minimizing information leakage to minimum and only the cloud storage provider who manages the deduplicaton knows it.

1.2.1. Types of Cloud Storage

Personal Cloud Storage:

Personal Cloud Storage is also known as Mobile Cloud Storage, it is a subset of public cloud storage. It provides user with the facility of storing data in cloud and also user can access data from anywhere. Across multiple devices it provides sharing capabilities and also provides data syncing. It is a local network-attached storage device and it provides a platform for storing data, music, photos, videos etc... An example of Personal Cloud Storage is Apple's iCloud.

Public Cloud Storage:

In this the user have no control over the nature of storage infrastructure as the data and files of the user are stored within the premises of the company that offers cloud storage services. It can be accessed online by any authorized user and requires little administrative control. Many different attacks and data hijacking can be experienced because the infrastructure is shared and also the resources that are used. Public Cloud Storage companies ensure that the servers used are of high performance.

Private Cloud Storage

It is also known as internal cloud storage. It runs on dedicated infrastructure in data centre and the users only pay for the storage capacity they need as does offer a pay-as-you-go model. It is similar to that of public cloud storage, it provides flexibility, scalability and usability of the storage architecture. It is not publicly accessible unlike public cloud storage and is owned by a single organization and its authorized partner.

Hybrid Cloud Storage

It combines the functionality of public and private cloud storage models to provide storage services. The services provided by hybrid cloud storage can be accessed using a web services API framework or cloud applications. It manages storage that uses both off-site and local resources. Some most popular hybrid cloud storage providers are Amazon Web Services, IBM, Microsoft, Cisco, etc.. In this the critical data resides in the enterprise private cloud while other data is stored and can easily be accessed from a public cloud storage provider.

1.2.2. Benefits of Cloud Storage

- Files can easily be accessed by user from anywhere via internet connection
- Most of the cloud provider provides 128 or 256 bits AES Encryption.
- Instead of e-mailing files to individual user can send a link to recipients through email.
- Cloud Storage can be used as a backup plan for our data. The data can be accessed via internet from the remote location.

- Cloud Storage costs about 3 cents per gigabyte to store data internally and it does not require any internal power when information is to be stored remotely.

1.3. Data Deduplication

It is a data compression technique that eliminates duplicate copies of similar repeating data. This technique is mostly used in cloud server to reduce space in server. Data is encrypted using different encryption technique before storing it on server. Along with saving the data storage it also reduces the amount of bandwidth of data transfer. Data deduplication is also called as Intelligent Compression it reduces the amount of data is to be stored. Data deduplication works by eliminating the repeated data and it stores only the first instance of any data. If the user tries to store the similar data again then only the reference id is created to originally store the data rather than storing the replica.

1.3.1. Benefits of Data Deduplication

- Storage space requirement decreases.
- Reduced amount of bandwidth for data transfer.
- Network efficiency increases.
- Overall cost of storage is reduced.

II. LITERATURE SURVEY

The following papers are surveyed in the following section along with its merit and demerits:

Rongzhi Wang(2017) [1] in his paper "Research on data security technology based on cloud storage" proposed a secure storage scheme based on Tornado codes (DSBT) by combining the technique of symmetric encryption and erasure codes. Based on Cassandra this paper implements a secure cloud storage prototype. It focuses mainly on how timely users are informed that there data is intact; how to restore data if it is not in good condition; how efficiently the key can be implemented to solve the problems in the cloud storage environment.

Sandip [2] in his paper "Utilization of data deduplication towards improvement of primary storage performance in cloud" suggested a overall performance oriented I/O deduplication, referred to as POD rather using capability oriented I/O deduplication that uses iDedup to enhance the I/O performance of primary storage in cloud. POD uses methods that can improve the overall performance overhead deduplication .The deduplication method used is select-Dedupe and for data fragmentation and memory control scheme it uses iCache. The overall work is done only on potential memory saving.

Dhanaraj[3] in his paper " Improving the Availability and Reducing Redundancy using Deduplication of Cloud Storage System" proposed system consists of various modules like data owner, file verification, file versioning and hybrid redundancy (HyRD), data owner module registers new users, file verification module checks whether the same file already exist or not. For this the algorithm used is MD5 to check the hash code generated by each file. In file versioning we create a new file with unique version number and then hybrid redundancy is used for storing the data into cloud storage system

Kamrul and Ragib [4] in the paper "Verifiable Data Redundancy in the Cloud", defines a model that uses distinct copies to store in the server as replicas and provides deterministic verification process which eliminates the risk of manipulation. It uses ElGamal decoder to decrypt the unique file. The proposed scheme can verify the level of data redundancy in the untrusted remote server. This system eliminates the client side storage because users do not require extra information for verification. But the VDRP scheme doubles the storage complexity in the server.

Goel [5] in the paper "Cloud Computing Based Social Media Model" proposed a model that can deal with data redundancy with the help of cloud computing in connecting several media sites with advertising companies. In order to avoid data redundancy the system shares data only to company's cloud server from where all social media sites can fetch the data by applying different algorithm. Uploading data on cloud server will reduce the overhead of sharing same data on different sites which will provide flexibility, scalibility and also leads to minimization of energy.

Keerthi and Reddy [6] in there paper "Encrypted Data Management with Deduplication in Cloud Computing" proposed a novel dynamic secret key generation protocol and a new data user authentication protocol. This scheme saves CSP storage space since it stores only one copy of the save data and the storage based data deduplication reduces the amount of storage needed for a given set of files. But this kind of duplication wastes networking resources and complicates data management.

Zheng, Mingjun, Yuxiang and Athanasios [7] in paper "Encrypted data Management with Deduplication in Cloud Computing" proposed a scheme that involves three system entities: data owner, data holder and CSP. For symmetric encryption it uses AES, for PKC it uses RSA and CP-ABE for data deduplication. Storing deduplication functional records could occupy some storage memory but this cost is minimal in comparison to the cost of storing a large volume of duplication data and it also supports many duplication instances and a huge

volume of duplicated data.

Harnik [8] in his paper "Side Channels in Cloud Services: Deduplication in Cloud Storage" proposes a multi user and a cross-user deduplication technique with a trust-based security mechanism which eliminates data redundancy in shared data environment but the drawback of this system is that it does not take security of sensitive data into view taking in account the various security attacks and vulnerabilities which occur inward and outward in cloud storage.

Satya and Krishna.B [9] in there paper "Eliminating Redundancy in cloud- Databases using Authorized Hybrid Cloud Approach" proposed a special encryption technique called convergent key encryption. The paper focuses mainly on two issues i.e. only authorized person can access the data and hybrid cloud approach for user's data. They have assumed that S-CSP is always online and has abundant storage capacity and storage power. They have taken into consideration only file level deduplication for simplicity.

Amanpreet and Sonia[10] in there paper "An Efficient Framework and Techniques of Data Deduplication in Cloud Computing" have proposed data deduplication technique along with securing technique to perform secure deduplication. The system compromises of four layers interface layer, chunk layer in thus layer for chunks hash values are computed by using MD5 algorithm and other includes algorithm to segment and upload file. This technique performs only approximate deduplication, therefore the deduplication efficiency is comparatively low as some duplicate chunks may be found across different groups.

Bellare, Sriram and Thomas [11] in paper "Message-Locked Encryption and Secure Deduplication" proposed a scheme for encryption called as Message Locked Encryption (MLE) where the encryption and decryption key is derived from the message itself. The cipher text is generated using the encryption algorithm that uses the key derived from the message. Cipher text is mapped to the tag to check the duplicates in the server. There is no storage overhead because the key is of fixed size. But the drawback of this system is that it is susceptible to brute-force attack.

Yan, Wang, Li and Vasilakos [12] in paper "Encrypted Data Management with Deduplication in Cloud Computing" proposed a scheme for deduplication which provides secure access using ABE (attribute based encryption). SHA1 is used for hash function by the system, for deduplication it uses CP_ABE, for symmetric encryption it uses AES and RSA for public key Cryptography. It supports data deletion, updation and deduplication with low operational and implementation cost as third party is not involved for key generation. Drawback of this scheme is that it takes more time for key generation.

Junbeom, Dongyoung, Youngjoo and Kyungtae [13] in there paper "Secure Data Deduplication with Dynamic Ownership Management in Cloud Storage" defined a deduplication scheme for encrypted data that has dynamic ownership management capability that uses randomized convergent encryption. It uses MD5 for key and token generation and AES with electronic code book algorithm for encrypting and decrypting a file. The system is secure against the chosen-plaintext attack, collosion attack and poison attack. It supports forward and backward secrecy of outsourced data. The disadvantage of the scheme is when a data loss attack occurs the system cannot recover the original data as all duplicates are removed from the cloud service provider.

Alzain, Ben and Eric [14] in paper "A New Approach Using Redundancy Technique to Improve Security in Cloud Computing" proposed a new model called (MCDB) which uses multi-clouds instead of single cloud service provider. It also uses the Shamir's secret sharing techniques and it also uses TMR (triple modular redundancy) with sequential method to improve the system reliability and enhance privacy and security.

Shobana, Shantha, Sridevi and Leelavathy [15] in there paper "De-Duplication of Data in Cloud" the data is encrypted using AES (Advance Encryption Standard) before being stored. System uses a effective user authentication using fingerprint feature extraction, image based authentication during file upload/download, eliminating repetition of data in cloud server and is implemented through multiple cloud storage. The convergent encryption technique is used to encrypt the data before outsourcing. It is extensively used in cloud storage to save bandwidth and minimize the storage space. The drawback is that it uses multiple cloud storage and splits the files and database of a specific user before storing.

III. CONCLUSION

It is clear that although the use of cloud computing has increased rapidly, but security, privacy and storage is considered as a major issue in cloud computing environment. This paper focused on issues related to duplication and redundancy of data in cloud along with security and privacy aspect in cloud. The purpose of this work is to propose a new model MPODO which uses RSA algorithm for key generation and encryption. For data redundancy we use the data deduplication technique by compressing both image and comparing Discrete Cosine Transform (DCT) is applied to each block and then each block is subjected to quantization for compression and string matching technique is used for comparison. The paper discusses the process of deduplication. The aim of proposed model is to maintain privacy and overcome deduplication overhead in cloud storage. In addition it also addresses data availability issue and resists brute-force attacks.

REFERENCES

[1]. Rongzhi Wang (2017) "Research on data security technology based on cloud storage" in 13th Global Congress on Manufacturing and Management, GCMM 2016.

[2]. P.Sai Sandip1, P.Rajeshwari2, Dr.G.Vishnu Murthy3 (2017)" Utilization of Data Deduplication towards Improvement of Primary Storage Performance in Cloud" in International Journal of Innovative Research in Science, Engineering and Technology.

[3]. Dhanaraj Suresh Patil, R. V. Mane, V.R.Ghorpade "Improving the Availability and Reducing Redundancy using Deduplication of Cloud Storage System" 978-1-5386-4008-1/17/$31.00 ©2017 IEEE

[4]. Mohammad Kamrul Islam and Ragib Hasan (2016) "Verifiable Data Redundancy in the Cloud" in 2016 IEEE International Conferences on Big Data and Cloud Computing (BDCloud), Social Computing and Networking (SocialCom), Sustainable Computing and Communications (SustainCom) 978- 1-5090-3936-4/16 $31.00 © 2016 IEEE DOI 10.1109/BDCloud-SocialCom-SustainCom.2016.1629.

[5]. Khilan Goel and Ishant Goel (2016) "Cloud Computing based Social Media Model" in ISSA 2011: Information Security South Africa, (2016).

[6]. S KEERTHI 1*, MADHAVA REDDY A 2* "ENCRYPTED DATA MANAGEMENT WITH DEDUPLICATION IN CLOUD COMPUTING", et al, International Journal of Research Sciences and Advanced Engineering [IJRSAE]TM Volume 2, Issue 20, PP: 26 - 30, OCT - DEC' 2017.

[7]. Zheng Yan, Mingjun Wang, and Yuxiang Li, Xidian University, China Athanasios V. Vasilakos "Encrypted Data Management with Deduplication in Cloud Computing" IEEE Cloud Computing,Volume:3,Issue:2,Mar. Apr.2016

[8]. Harnik, D. ,IBM Haifa Res. Lab., Haifa, Israel Pinkas, B. ; Shulman-Peleg, A. Side Channels in Cloud Services: Deduplication in Cloud Storage,Volume:8 Issue:6 Date Nov.-Dec. 2010

[9]. Satya Sudheer Varma Surimalla1, Krishna. B2(2015) "Eliminating Redundancy in Cloud –Databases Using Authorized Hybrid Cloud Approach" in International Journal of Computer Science and Information Technology Research ISSN 2348-120X (online) Vol. 3, Issue 3, pp: (355-358), Month: July- September 2015.

[10]. Amanpreet Kaur, Sonia Sharma "An Efficient Framework and Techniques of Data Deduplication in Cloud Computing" IJCST Vol. 8, Issue 2, April - June 2017

[11]. Mihir Bellare1, Sriram Keelveedhi1, and Thomas Ristenpart2 T. Johansson and P. Nguyen "Message-Locked Encryption and Secure Deduplication" (Eds.): EUROCRYPT 2013, LNCS 7881, pp. 296–312, 2013.c_International Association for Cryptologic Research 2013

[12]. Zheng Yan, Mingjun Wang, Yuxiang Li and Athanasios V. Vasilakos, "Encrypted Data Management with Deduplication in Cloud Computing," IEEE Cloud Computing, March-April 2016, pp. 29-35.

[13]. Junbeom Hur, Dongyoung Koo_, Youngjoo Shinz and Kyungtae Kang "Secure Data Deduplication with Dynamic Ownership Management in Cloud Storage" (Extended Abstract) 2017 IEEE 33rd International Conference on Data Engineering

[14]. Mohammed A. AlZain, Ben Soh and Eric Pardede "A New Approach Using Redundancy Technique to Improve Security in Cloud Computing" researchgate.net,publication,232733273

[15]. R. SHOBANA*, K. SHANTHA SHALINI, S. LEELAVATHY and V. SRIDEVI "DE-DUPLICATION OF DATA IN CLOUD" Int. J. Chem. Sci.: 14(4), 2016, 2933-2938 ISSN 0972-768X

[16]. Jin Li, Yan Kit Li, Xiaofeng Chen, Patrick P.C. Lee, and Wenjing Lou "A Hybrid Cloud Approach for Secure Authorized Deduplication" IEEE TRANSACTIONS ON PARALLEL AND DISTRIBUTED SYSTEMS, VOL. 26, NO. 5, MAY 2015

[17]. Manikantan U.V.1, Prof.Mahesh G.2 (2015) "A Survey on Data Deduplication in Cloud Storage Environment" in International Journal of Scientific Research Engineering & Technology (IJSRET), ISSN 2278 – 0882 Volume 4, Issue 4, April 2015.

[18]. Ali E. Taki El_Deen, El-Sayed A. El-Badawy, Sameh N. Gobran " Digital Image Encryption Based on RSA Algorithm" IOSR Journal of Electronics and Communication Engineering (IOSR-JECE) e-ISSN: 2278-2834,p- ISSN: 2278-8735.Volume 9, Issue 1, Ver. IV (Jan. 2014), PP 69-73

[19]. S. Supriya and Dr. S. Mythili (2017) "A STUDY ON DATA DEDUPLICATION IN CLOUD COMPUTING" in Volume 8, No. 8, September-October 2017 International Journal of Advanced Research in Computer Science.

[20]. Zhan Wang, Kun Suny, Jiwu Jing, Sushil Jajodia "Verification of Data Redundancy in Cloud Storage" CloudComputing'13, May 8, 2013, Hangzhou, China. Copyright 2013 ACM 978-1-4503-2067-2/13/05 ...$15.00.

Zarka Khan" Maintaining Privacy and Overcoming Duplication Overhead in Cloud Computing" International Journal of Computational Engineering Research (IJCER), vol. 09, no. 6, 2019, pp 08-12