Survey of various Data Duplication Method and Security Algorithm Issue in Cloud Computing

Mr. Liladhar M.Kuwar[1] Prof. Kapil Vyas[2]
[1]M.E. Student [2]Assistant Professor
[1,2]Department of Computer Science & Engineering
[1,2]BM College of Technology, Indore (M.P.)

Abstract— Internet and network application are growing very fast and use of cloud computing has increased rapidly in today's world. Cloud computing provides many advantages such as reducing cost, on-line, on-demand services and accessibility of data. Cloud computing defined different computing idea which includes huge number of peripheral device connected with real time communication like internet in computer networking. Now days due to benefited in storage technology and networking, larger function of data is being maintained in digital form. Similar data storage in disk is responsible for unnecessary storage space of data. Data de-duplication is effective way to reduce the storage space and optimize the utilization of storage space. Duplication records appear when uniform storage technique is used in data base. Identification of duplicate data is time consuming. In this paper we concentrate about data de-duplication technique and security mechanism because Organizations are transferring important information to the Cloud that increases concern over security of data. Cryptography is common approach to protect the sensitive information in Cloud. Encryption algorithm plays a main role in information security system. In this survey paper we examine bout various security algorithm comparison.

Key words: Cloud Computing, Data De-Duplication, Security Algorithm

I. INTRODUCTION

Cloud Computing is technology that interact between businesses and users to use application without an installation. Users and businesses can access the information and files at any computer system having an internet connection. Cloud computing utilizes both central remote servers and internet to manage the data and applications with use of internet technology. With a lot of benefit of cloud such as scalability, accessibility, cost saving world user tend to shift their data to cloud storage.

Cloud storage providers uses different techniques to improve storage efficiency and one of leading technique employed by them is de-duplication, which claims to be saving 90 to 95% of storage [1],[2].

Cloud computing providers offer their services according to three fundamental models Infrastructure-as-a Service (IaaS), Platform-as-a-Service (PaaS), and Software as-a-Service (SaaS) where IaaS is the most basic and each higher model abstracts from the details of the lower models. PaaS in the cloud is defined as a set of software and product development tools hosted on the provider's infrastructure. Developers create applications on the provider's platform over the Internet. PaaS providers may use APIs, website portals or gateway software installed on the customer's computer. In SaaS cloud model, the vendor supplies the hardware infrastructure, the software product and interacts with the user through a front-end portal. IaaS is to start, stop, access and Data de-duplication is a process where only one copy of the duplicate data is stored on the server. This reduces the storage space and cost of maintaining storage space.

Data once deployed to cloud servers, its beyond the security premises of the data owner, thus most of them prefer to outsource their in an encrypted format. Data encryption by data owners eliminates cloud service providers chance of de-duplicating it since encryption and de-duplication techniques have conflicting strategies, i.e., data encryption with a -key converts data into an unidentifiable format called cipher text thus encrypting, even the same data, with different keys may result in different cipher texts, making de-duplication less feasible .However, performing encryption is essential to make data secure, at the same time, performing de-duplication is essential for achieving optimized storage. Therefore, de-duplication and encryption need to work in hand to hand to ensure secure and optimized storage. Various techniques and approaches used for de-duplication over encrypted data are studied in this paper.

De-duplication is an effective technique to optimize the utilization of storage space. Data de-duplication is a technique which stores the data only once which means that the same data cannot be stored in the cloud storage area. Data de-duplication is used to reduce the storage space in the cloud and to efficiently use the bandwidth for uploading and downloading the data from the cloud storage area. For cloud provider it is very helpful because you can de-duplicate what you store. Due to reduction in cost it is being more popular. This paper will briefly describe Data De-duplication and give a comprehensive survey.

Secure de-duplication is one of the most complimentary and arising challenge. Although the convergent key encryption. Security in cloud computing is major issue, as user often store sensitive information in cloud storage provider but these providers may be entrusted. Encryption algorithms play a main role in information security systems. Several issues as scalability, security, performance, integrity etc. are discussed so far. This paper surveys recent research related to cloud security and data duplication solution. This paper introduces a better data duplication technique for cloud. It is found in research that security is major issue in cloud. This work aims to promote the use of data in cloud and reduce security risks that affect the cloud computing user. The way to secure the data using different compression and encryption algorithms and to hide its location from the users that stores and retrieves it.

II. BACKGROUNDS

A. De-duplication

Now day data duplication is rapidly growing technique use in data backup storage without redundancy. It is very important in unique data management De-duplication technique is used for technique for removing redundant data. Instead of keeping multiple data copies with the same content, de-duplication eliminates redundant data by keeping only one physical copy and referring other redundant data to that copy [9].

B. Granularity based de-duplication

De-duplication can be categorized as file level de-duplication and block level de-duplication based on granularity

- File level de-duplication In File level only unique copy of file will be stored and duplicate copy will be discarded. File level de-duplication takes into account the entire file, thus even small update or append makes the file different from previous version of it and thereby reducing de-duplication ratio.
- Block level de-duplication data blocks are considered for de-duplication. De-duplication can further categorized based on location of de-duplication i.e., as client side de-duplication and as source side de-duplication. De-duplication is widely is used various applications like backup, metadata management, primary storage, etc. for storage optimization [3] Block level de-duplication. In Block level each file is divided in the blocks and then only unique block will be stored. Length of divided block can be fixed or variable. Level of de-duplication in block level is more than file level de-duplication that means de-duplication ratio is high in block level de-duplication.

There are disadvantages and advantages to each approach. De-duplication can also be applied at byte level. The differences lie in the amount of reduction each produces and the time each approach takes to determine what's unique.

Table 1 Comparison of data duplication methods: Data De-Duplication approaches are comprised according to cost, throughput, Used Bandwidth, ration of de-duplication and required storage basics of file level and block level method[13].

Approach	Cost	Throughput	Used Bandwidth	De-duplication ratio	Required Storage
File Level	Deduplication	Low High	Low	Low	Medium
Block level Deduplication	High	Low	Low	High	Less
Source based Deduplication	Relatively Low	Medium	Low	Medium	Medium
Target based Deduplication	High	Medium	High	Medium	Medium

Table 1: Data Deduplication Method

C. Convergent Encryption

Convergent encryption [4], is an encryption approach that support de-duplication. With convergent encryption, encryption key is generated out of hash of plain text. Convergent Encryption [9]. In convergent encryption secret key will be derived as hash value of plaintext. So the same plaintext will lead to the same cipher text. In addition tag is also derived to detect the duplicate.

D. Proof of Ownership

De-duplication works performed by computing cryptographic hash function over data and use of hash value to determine similar data. Identification of duplicate data then but pointer to file ownership is updated thus saving storage and bandwidth and new data is not uploaded. When it comes to client side de-duplication, hash values of data are computed at client and send for duplicate check. An attacker, who gains access to hash value of a data which not authorized to him/her, may claim de-duplication of file and thereby gaining access to the file. To defend such an attack, a Proof Of Ownership (PoW) has been proposed in [6], and various works like[6][7] etc. adapted this method. PoW works as an interactive algorithm between two parties - a prover and verifier to prove the ownership of the file. Verifier computes a short value of data M whereas, a prover need to compute short value of M and send it to verifier for claiming ownership of M [5],

Data De-duplication gives benefits but security and data confidentiality is still major issues. So, usual way to provide security is encryption. But there is confliction between data deduplication and encryption. Because it is possible that same plaintexts may lead to different ciphertexts. If one can realize data de-duplication on ciphertexts, the cloud server must be able to identify all of the ciphertexts of the same plaintext]5]. One more thing needs to be there in data deduplication is authorized deduplication in which users would have set of privileges because in many of applications differential authorized duplication is needed. User can not check duplicate out of his privilege set. For example any role based application may require authorized de-duplication [13].

III. DATA SECURITY ISSUES IN CLOUD COMPUTING

Data de-duplication is major issue in cloud computing. We can't be used data de duplication technique alone in cloud, because there is often need of data security. So data de--duplication and convergent encryption work in collaboration such that, data de-duplication is possible with security of data. But convergent encryption does not provide much security, as it can be susceptible to guessing and brute force attacks.

Enhance data de-duplication process and security. In order to protect the user's information from reveal, Siani Pearson [10]. put forward design principles in design process of cloud computing services to ensure that user's message and business information would not leaked out. It includes Transmit and store user's information as little as possible. After systemic analysis, the cloud computing applications will collect and store the most necessary information only. Security measures will be adopted to prevent unauthorized access, copying, using or modifying

personal information[10]. Achieve user's control to the greatest degree. Firstly, it is necessary to allow the user to control the most critical and important personal information. Secondly, it is available to manage personal information by a trusted third party. Allow users to make choice. Users have the right to select the use of personal information. Besides, they can join or leave freely. Make clear and limit the purpose of use of data. Personal information must be used and handled by Secure communication over the network done by security algorithm, encryption algorithm plays a role of fundamental tool for protecting the data[11]. Encryption algorithm converts the data into scrambled form by using "the key" and only user have the key to decrypt the data. In Symmetric key encryption, only one key is used to encrypt and decrypt the data. Another technique is using asymmetric key encryption; two keys- private and public keys are used. Public key is used for encryption and private key is used for decryption. Fig shows some of the symmetric & asymmetric algorithms[12]. Symmetric Encryption . In Symmetric encryption common key will be used to encrypt or decrypt the information and key generation algorithm that generates k using security parameter .The symmetric encryption algorithm that takes the secret k and message M and then outputs the Ciphertext C; and M .symmetric decryption algorithm that takes the secret k and Cipher text C and then outputs the original message M.

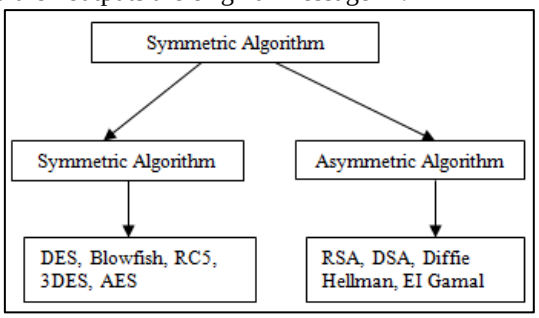

Fig. 1: Security Algorithms in Cloud computing

A. Comparison of Existing Algorithms on the basis of different parameters

Characteristics	DES	Blowfish	RC5	3-DES	AES
Developed	1977	1993	1994	1998	2000
Block Size	64	64	32 ,64 or 128	64	128, 192 or 256
Key Length	56	32-448	MAX2040	112,168	128, 192 or 256
Security	Proven Inadequate	Considered Secure	Considered Secure	Considered Secure	Considered Secure
Speed	Very slow	Fast	Slow	Slow	Very fast

Table 2: Comparison of Existing Algorithms on the basis of different parameters

IV. Conclusion

Cloud Computing is interest of topic for research. De-duplication is a method available in cloud storage for saving bandwidth and storage capacity. As in the starting of paper it is very clear that data de-duplication is the one of effective method to handle duplicate data but, de-duplication is less feasible with encrypted data. Data security issue is known to better secure data.

This paper would be helpful to new researcher who wants to research on secure data de-duplication. In this paper various methods are discussed where de-duplication methods are carried out on encrypted data in a large storage area. Security methods studied here which is a simple approach that makes de-duplication compatible with encrypted data. In this paper we study both security and de-duplication of data across storage areas. A strategy needs to study for data duplication and secure transmition over cloud computing envoirment.in future we work for a new security approach for secure data transmition and de duplication mechanism using one of the security algorithm .

Reference

[1] OpenDedup. OpenDedup, Global inline deduplication for Block Storage and Files. [online] 2010 Available from: http://opendedup.org/index.php.
[2] J. Douceur, A. Adya, W. Bolosky, D. Simon, and M. Theimer. "Reclaiming space from duplicate files in a serverless distributed file system. In Distributed Computing Systems", 2002. Proceedings. 22nd International Conference on, pages 617{624. IEEE
[3] Dutch T Meyer and William J Bolosky."A study of practical Deduplication". ACM Transactions on Storage (TOS), 7(4):14, 2012
[4] J. Douceur, A. Adya, W. Bolosky, D. Simon, and M. Theimer. "Reclaiming space from duplicate files in a serverless distributed file system. In Distributed Computing Systems", 2002. Proceedings. 22nd International Conference on, pages 617{624. IEEE, 2002.
[5] Dropbox http://www.dropbox.com.
[6] AmazonS3 http://aws/amazon.com/s3s
[7] GoogleDrive http://www.drive.google.com.
[8] RuWei Huang, Si Yu Wei Zhuang and Xia]oLin XiaoL in Gui," Design Of Privacy-Preserving Cloud Storage Framework "2010 Ninth International Conference on Gried and Cloud Computing .
[9] 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.
[10] Balachandra Reddy Kandukuri, Rama Krishna Paturi and Dr. AtanuRakshit, "Cloud security issues" In Services Computing, 2009. IEEE International Conference on, page 517520, 2009
[11] David Pointcheval, "Asymmetric Cryptography and Practical Security", International Journal of Security and Its Applications, Volume 4,2002
[12] Kashish Goyal, Supriya Kinger" Modified Caesar Cipher for Better Security Enhancement" International Journal of Computer Applications (0975 – 8887) Volume 73– No.3, July 2013.
[13] Riddhi Movaliya ,Harshal Shah "A survey of secure data duplication" International Journal of Computer Applications (0975 – 8887) Volume 138 – No.11, March 2016
[14] Randeep Kaur1 ,Supriya Kinger2 "Analysis of security algorithm in cloud" International Journal of Application

or Innovation in Engineering & Management (IJAIEM) Volume 3, Issue 3, March 2014

[15] Jia Yu,Kui Ren,Cong Wang, Vijay Varadharajan " Enabling Cloud Storage Auditing With Key-Exposure Resistance " IEEE Transaction on information and security ,Vol. 10,No. 6, June 2015.

