International Journal for Research in Engineering Application & Management (IJREAM)
ISSN : 2494-9150 Vol-02, Issue 03, June 2016.

Secure Data Compression Technique over Hybrid Cloud Approach

[1]Prajakta Patil, [2]Prof. B. R. Nandwalkar

[1]M.E Student, [2]Assistant Professor, [1,2]Department of Computer Engineering, Late G. N. Sapkal College of Engineering, Anjaneri, Nasik, SavitribaiPhule Pune University, Maharashtra, India.

[1]*pprajakta25@gmail.com*, [2]*nandwalkar.bhushan@gmail.com*

Abstract - The use of Cloud storage is increasing day-by- day. So the management of such ever increasing data is critical challenge to cloud storage provider. As a solution to this problem we are using data de-duplication technique which eliminates the redundant data copies and keeps only unique copies of data. So, by using such data compression technique we remove duplicate copies from cloud and by this we can improve the storage utilization. Here in this paper we are considering only file level de-duplication and block level de-duplication is considered for future work. We proposed a convergent encryption technique which provides confidentiality to sensitive data and also supports de-duplication. We address problem that is authorized de-duplication with hybrid cloud approach. We use multiple clouds for storage to reduce the security risk. The data of data owners is stored on the public cloud while all operations regarding that data is managed with private cloud. For the security purpose we are using the encryption techniques to encrypt the data stored on public clouds.

Keywords *-- Authorization, Authorized duplicate check, confidentiality, Data access, De-duplication, Hybrid cloud.*

I. INTRODUCTION

Cloud computing is becoming more and more popular day-by-day and the number of cloud users are also increasing rapidly. It provides a low-cost, and scalable infrastructure for managing data and storage. Its rapid adoption for storage is accompanied by increasing data so the management of such data is needed. Data de-duplication gives a solution to this problem it keeps only unique data copies in the cloud storage and removes all the redundant data copies from cloud storage. Data de-duplication improves storage utilization and saves bandwidth.

Data de-duplication [14] is based on different granularities: It can be whole file that is file level de-duplication or block of particular file that is block level de-duplication. Though data de-duplication gives many benefits security and privacy problems arises as users data is susceptible to attacks.

Traditional encryption techniques provides data confidentiality but these are incompatible with de-duplication because in traditional encryption different users will lead to different encryption keys and which will lead to different cypher text makes data de-duplication impossible. The convergent encryption technique is been proposed to provide data confidentiality as well as data de-duplication in cloud. In convergent encryption technique convergent keys are derived from data copy itself by applying cryptographic hash function on the data copy, so the identical data copies will have same cypher text which makes de-duplication possible.

To prevent the unauthorized access of data the users have to run the "proof of ownership" protocol with the cloud service provider. It is preventing the unauthorised data access. The proposed system supports differential authorization duplicate check, each user is assigned with different privileges at the time of system initialization. We are using hybrid cloud architecture [1] [7] which consist public cloud and private cloud. Data owner will send their data to public cloud while all the data operations are managed by the private cloud. We promote use of multi-clouds for its ability of reducing security risk which affect the cloud storage user.

The paper focuses on issues that are related to cloud storage:
- **Authorized access guarantee:** Only the user who is having the privileges to access the authorized data can access secure the file.
- **Hybrid clouds:** Public and private clouds for security of users data.

Data de-duplication problem is being resolved by using the differential privileges for each user in the cloud computing, hybrid cloud architecture for security of users sensitive data is also considered which consist public and private cloud. The

proposed system is enhanced in the security as multiple clouds are used for storage. So the sensitive data of cloud users is protected and securely sent to the cloud storage.

II. LITERATURE REVIEW

Hybrid cloud architecture provides more security to the cloud data. An organization could work efficiently using both private and public clouds.

Kamara et al. developed virtual private storage service [12] it is based on the cryptographic technique that provides security to cloud storage. That is confidentiality, integrity and non-repudiation to data stored on cloud storage.

Wang et al. proposed a system that provides secure and efficient access to outsourced data [5]. The end user requests for data access, then the data owner send back an encryption key and data access certificate to the end user, then the end user will send that access certificate to the data storage provider and data storage provider send the encrypted data to the end user. System is incompatible with de-duplication.

Bellare et al. Developed a new technique of cryptography [10] in which it overcomes the drawback of traditional encryption techniques called Massage Locked Encryption. The key for message encryption and decryption will be derived from the message data itself. This technique gives number of encryption keys.

Douceur et al. proposed a mechanism called Farsite [15] it is a distributed file storage system and it provides a security and reliability to encrypted replicas of file on the multiple desktop machines. System results in number of encryption keys.

Users are using the PoW[8] [4] protocol to prove their identity. Haveli et al. developed Proof of Ownership [8] protocol for the de-duplication system. The user will efficiently prove to cloud storage provider that it owns that file without uploading the file to cloud storage. It is based on Markel tree.

Bugiel et al. Developed a twin cloud architecture [11] that provides a secure outsourcing of data and arbitrary computations of untrusted commodity cloud. The end user can communicate with private cloud which encrypts the data and verifies the stored data and then performs operation on untrusted commodity cloud.

III. EXISTING SYSTEM

In the existing system author provided different methods to solve problem of de-duplication in hybrid cloud environment using differential privileges. Here they have considered hybrid cloud architecture consisting public and private cloud. In system data owners outsource their data to the public cloud and all the operations regarding data are managed with private cloud.

They have conducted experiment on 3 different machine which is equipped with Intel core-2-quad-2.66GHz Quad Core CPU, 4GB RAM and having Ubuntu 12.04 operating system. 1Gbps Ethernet connection is used to connect the machines.

They break down the overall process into six steps. That is tag generation, token generation, duplicate check, share token generation, encryption and transfer.

III. RESULT AND DISCUSSION

The critical challenge of cloud storage system is the management of increasing volume of data. So, to solve such problem data de-duplication techniques are used. The de-duplication techniques removes redundant data copy and will not store the duplicate data on cloud storage, it will store only the physical data and provide a pointer to the redundant to that stored copy. Thus such de-duplication techniques improve the storage utilization and bandwidth of the cloud storage.

The diagram shown below can describe the overall predicted architecture of our system. We introduce the use of multi-clouds which reduces the security risk. We also introduce hybrid clouds for data security and authorized data access to prevent unauthorized data access.

A. Objectives and findings for the proposed system:

- To protect confidentiality of the sensitive data and also supporting de-duplication.
- To propose a new de-duplication system that supports differential authorization and duplicate check.
- To promote the use of multiple-clouds because of its ability of reducing security risk.
- Protects users data

B. Basic modules of system:

a) User Module: In this module, users are authenticated and has the secure access to the detail stored on the system. But before access or search operation of the details user must have the account in the system otherwise they must register first.

b) De-duplication System: To support de-duplication, the tag of file F is determined using file F and the privilege set P. To support authorized data access, a secret key kp is bounded to a privilege p for generating a file token.

c) Duplicate Check Token Security:

i) Unforgeability of duplicate-check token: Two types of adversaries are considered, external and internal adversary. The external adversary could be viewed like an internal adversary who is not having any privilege p. If user is having privilege p, it is required that adversary should not forge and output any valid duplicate token using any other privilege p' on

F, where p does not match with p'. Further, it is also required that if adversary does not makes a request of file token with its own privilege p from private cloud , it should not forge and output any valid duplicate token with privilege p on any file F that is been queried.

- Send Key: Once the request for the key was received, the sender can send key or he can reject it. With the key and the request id, which has been generated at time of sending request for key, receiver is able to decrypt the message.

C) Block Diagram :

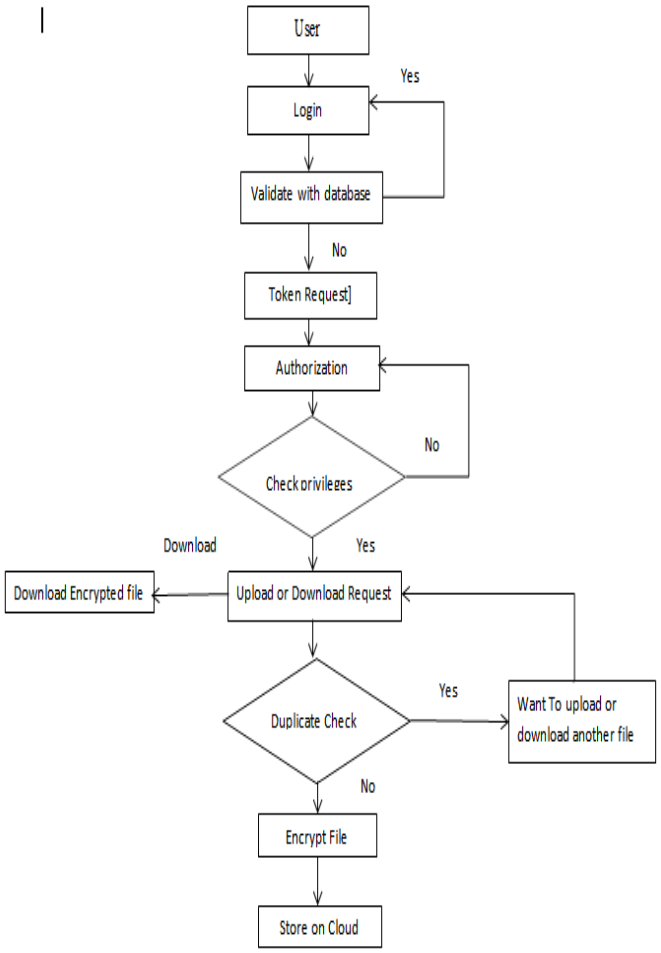

Fig. 1 Architecture for Secure Authorized De-duplication

D. Flow of Proposed System:

- File Uploading: User who wants to upload data to the cloud will compute a tag of a file and will send it to private cloud. Private cloud will check for the duplicate copy with S-CSP. If duplicate is found, user will run PoW protocol to prove its identity. If it is passed then user will get a pointer to the stored file for future use.

If no duplicate is found the user will compute encrypted data and store it on public cloud.

- File Downloading: The user who wants to download file, first sends request with file name to S-CSP. Then S-CSP checks whether user is authenticate user for file F. If it is failed, S-CSP will send abort signal. Otherwise, S-CSP sends back cipher text to the user. User will recover the original data from that Cipher text.

E. Algorithms Used:

- File Uploading:

BEGIN

Step-1 Read File

Step-2 Cloud server generates a token and checks for duplication

Step-3 Sends a response back to user whether file is already exist or not

Step 4 If file already does not exist

4.1 Display 'No duplication found'

Step-5 If the file exists

5.1 Display the details of user who uploads that file

END

F. File Downloading:

BEGIN

Step-1 Read File

Step-2 Cloud server generates a token and checks for eligibility

Step-3 Sends a response back to user whether user is eligible or not

Step 4-If user is eligible

4.1 User can download encrypted file

Step-5 If the user is not eligible

5.1 user receives abort signal

END

V. RESULT AND ANALYSIS

We conduct experiment on a machine equipped with Intel core i3 processor 2.13Ghz CPU, 4GB RAM and has windows 7 64 bit operating system, Google drive as cloud space for storage. The results are based on speed of the internet connection used. With different speed connection values obtained may vary. So, the results depend on connection speed.

Here in fig 2. we input a file of different file sizes to existing as well as proposed system and observe the time taken by the both the system, Where we consider the file sizes ranging from 2mb to 20mb.We noticed that proposed system takes less time for different file size input. And note down the average time taken by both the systems.

The fig 3.shown below shows the time taken by the Proposed and existing system for the calculation of different stages in de-duplication, Where we consider the file sizes ranging from 2mb to 20mb.Where we uploads 30 unique 5mb files. From the fig 3 we can see that results remain constant. Here we do token checking with hash table.

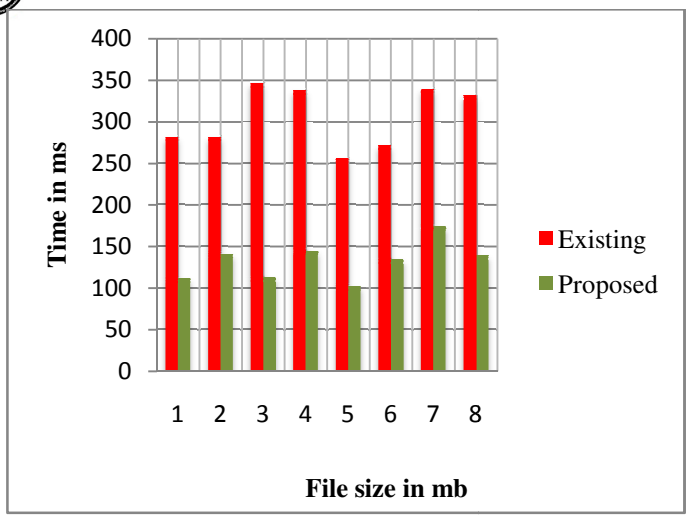

Fig 2. Time breakdown for proposed and existing system for different file sizes(mb)

Fig.3 time breakdown for proposed and existing system for no. Of stored files

File Size	Existing	Proposed	Diff.
1	281.5	111.9	169.6
2	282.16	140.66	141.5
3	346.166	113.055	213.11
4	337.83	144.55	193.28
5	255.75	102.75	153
6	271.66	134.83	136.83
7	339.5	174.41	165.09
8	332	139.5	192.5
11	402.166	156.833	245.33
13	303	159.166	143.83
15	325	145.33	179.67
16	304.66	159.166	145.49
19	300	204.166	95.83

Table 1.. Time breakdown for proposed and existing system for different file sizes (mb)

From the above table we can see the difference generated for the results of different file sizes in the proposed and existing system. The result generated is dependent on the speed of the internet connections we used

VI. CONCLUSION AND FUTURE SCOPE

The proposed Authorized de-duplication System provides protection to the data from the attacker, by generating Tag and running Proof of Ownership (PoW). Our proposed system shows that it incurs minimal overhead compared with previous de-duplication system in a hybrid approach. The de-duplication system provides a way to secure file by sending file Tag. Proposed system is totally different from previous de-duplication systems, the differential privileges of users is the enhanced thought in duplicate check beside the data itself. Block level de-duplication is considered as future work.

REFERENCES

[1]Jin Li, Yan Kit Li, Xiaofeng Chen, Patrick P. C. Lee, Wenjing Lou. *A Hybrid Cloud Approach For Secure Authorized De-duplication.* In IEEE Transactions on Parallel and Distributed Systems, 2014.

[2] J. Li, X. Chen, M. Li, J. Li, P. Lee, and W. Lou. *Secure de-duplication with efficient and reliable convergent key management*. In IEEE Transactions on Parallel and Distributed Systems, 2013.

[3] W. K. Ng, Y. Wen, and H. Zhu. *Private data de-duplication protocols in cloud storage.* In S. Ossowski and P. Lecca, editors, Proceedings of the 27th Annual ACM Symposium on Applied Computing, pages 441–446. ACM, 2012.

[4] R. D. Pietro and A. Sorniotti. *Boosting efficiency and security in proof of ownership for de-duplication*. In H. Y. Youm and Y. Won, editors, ACM Symposium on Information, Computer and Communications Security, pages 81–82. ACM, 2012.

[5] W. Wang, Z. Li, R. Owens, and B. Bhargava, ''*Secure and Efficient Access to Outsourced Data*,'' in Proc. ACM CCSW, Nov. 2009, pp. 55-66.

[6] M. Bellare, S. Keelveedhi, and T. Ristenpart.*Dupless: Server aided encryption for de duplicated storage*. In USENIX Security Symposium, 2013.

[7] K. Zhang, X. Zhou, Y. Chen, X.Wang, and Y. Ruan. *Sedic: privacy aware data intensive computing on hybrid clouds*. In Proceedings of the 18th ACM conference on Computer and communications security, CCS'11, pages 515–526, New York, NY, USA, 2011. ACM.

[8] S. Halevi, D. Harnik, B. Pinkas, and A. Shulman-Peleg. *Proofs of ownership in remote storage systems.* In Y. Chen, G. Danezis, and V. Shmatikov, editors, ACM Conference on Computer and Communications Security, pages 491–500. ACM, 2011.

[9] Z. Wilcox-O'Hearn and B. Warner. *Tahoe: the least-authority filesystem.* In Proc. of ACM StorageSS, 2008

[10] M. Bellare, S. Keelveedhi, and T. Ristenpart. *Message-locked encryption and secure de-duplication.* In EUROCRYPT, pages 296–312, 2013.

[11] S. Bugiel, S. Nurnberger, A. Sadeghi, and T. Schneider. *Twin clouds: An architecture for secure cloud computing*. In Workshop on Cryptography and Security in Clouds (WCSC 2011), 2011.

[12] S. Kamara and K. Lauter, ''*Cryptographic Cloud Storage,'' in Proc. Financial Cryptography: Workshop Real-Life Cryptograph*. Protocols Standardization, 2010, pp. 136-149.

[13] R. Geambasu, T. Kohno, A. Levy, and H.M. Levy, ''*Vanish: Increasing Data Privacy with Self-Destructing Data*,'' in Proc. USENIX Security Symp., Aug. 2009, pp. 316-299..

[14] M. W. Storer, K. Greenan, D. D. E. Long, and E. L. Miller. *Secure data de-duplication*. In Proc. of StorageSS, 2008.

[15] J. R. Douceur, A. Adya, W. J. Bolosky, D. Simon, and M. Theimer. *Reclaiming space from duplicate files in a server less distributed file system*. In ICDCS, pages 617–624, 2002

