ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 7, Issue 2, February 2019

Secured Data Deduplication in Cloud Environment

S.Gayathri, P.Kowsalya, K.Mathumitha, P.N.Nagajothi

U.G. Student, Department of Computer Engineering, Saranathan College of Engineering, Trichy, Tamilnadu, India

ABSTRACT-. The most of the main issues with cloud computing have been addressed to a degree that clouds have become interesting for full commercial exploitation. Data deduplication means the process of identifying and eliminating duplicated copies of data. Client-side data compression ensures that multiple uploads of the same content only consume network bandwidth and storage space of a single upload. Compression is actively used by a number of cloud backup providers as well as various cloud services. Unfortunately,encrypted data is pseudorandom and thus cannot be deduplicated as a consequence current schemes have to entirely sacrifice either security or storage efficiency. In this system, present a scheme that permits a more fine-grained trade-off. The intuition is that outsourced data may require different levels of protection, depending on how popular it is: content shared by many users. Then present a novel idea that differentiates data according to their popularity. Based on this idea, design an encryption scheme that guarantees semantic security for unpopular data and provides weaker security and better storage and bandwidth benefits for popular data. Data deduplication is more effective and protects the data semantically over secure encryption. Finally, can use the backup recover system at the time of blocking and also analyze frequent login access system.

KEYWORDS: Data Deduplication , Security Enhancement , Backup Recovery.

I. INTRODUCTION

Cloud service provider which provide services to the user. Some of the services provided by the cloud are SAAS (Software as a Service), PAAS (platform as a Service), IAAS (Infrastructure as a Service). Due to lot of storage capacity in cloud there is a problem of storing a large number of redundant data so there is a wastage of cloud storage. There is a technique called data de-duplication which means eliminating the redundant copies of data.

There are different types of de-duplication techniques like

1) Location based deduplication
2) Content based deduplication
3) Chunk based deduplication

In location based de-duplication, the interaction between client side and server side. In client side, the file is uploaded in the cloud storage and here the server checks whether the file is duplicated or not. If its duplicated then the server sends the alert to the client. In content based, server checks the content of the file. In chunk based the file is divided into the different chunks then its encrypted followed by compression.

II. LITERATURE SURVEY

Author: Wee Keong Ng

In this paper, a new notion which we call private data deduplication protocols is introduced and formalized in the context of two-party computations. A feasible result of private data deduplication protocols has been proposed and analyzed. We have shown that the proposed private data deduplication protocol is provably secure in the simulation based framework assuming that the underlying hash function is collision-resilient, the discrete logarithm is hard and the erasure coding algorithm E can erasure up to fraction of the bits in the presence of malicious adversaries. To overcome such attacks, introduced the formalized the notion of proofs of ownership (the HHPS protocol), where a client P proves

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 7, Issue 2, February 2019

to a server S that it actually holds the data of file F and not just some short summary string, The data deduplication protocol are closely related to the proofs of retrievability and proofs of data possession but they are significantly different in the sense that the proofs of retrievability and data procession often use a pre-processing step that cannot be used in the data deduplication procedure. Specially, Harnik et al consider an attacker that is able to temporarily compromise a server machine, getting access to its internal cache, which includes the hash values for all the recently accessed files. Having obtained this piece of information, the attacker is able to download all these files, which may include confidential files of others.

III.EXISTING SYSTEM

Many systems have been developed to provide secure storage but traditional encryption techniques are not suitable for compression purpose. Deterministic encryption, in particular convergent encryption is a good candidate to achieve both confidentiality and compression but it suffers from well-known weaknesses which do not ensure protection of predictable files against dictionary attacks. Unfortunately, compression loses its effectiveness in conjunction with end-to-end encryption. End-to-end encryption in a storage system is the process by which data is encrypted at its source prior to ingress into the storage system. It is becoming an increasingly prominent requirement due to number of security incidents linked to leakage of unencrypted data and the tightening of sector-specific laws and regulations. Clearly, if semantically secure encryption is used, file compression is impossible, as no one apart from the owner of the decryption key can decide whether two cipher texts correspond to the same plain text. The design of storage efficiency functions in general and of compression functions in particular that do not lose their effectiveness in presence of end-to-end security is therefore still an open problem.

DISADVANTAGES OF EXISTING SYSTEM

- Compression checks only the length of file without considering file descriptions.
- It could not achieve secure access control under a dynamic ownership changing environment.
- Security degradation of the cloud service.

IV. PROPOSED SYSTEM

Data compression is the process by which a storage provider only stores a single copy of a file owned by several users. Compression is most rewarding when it is triggered at the client side, as it also saves upload bandwidth frequency. For these reasons, compression is a critical enabler for a number of popular and successful storage services that offer cheap, remote storage to the broad public by performing client-side compression, thus saving both the network bandwidth and storage costs. The goal of the system is to guarantee the data confidentiality without losing the advantage of compression. Confidentiality must be guaranteed for all files, including the predictable ones. The security of the whole system should not rely on the security of a single component (single point of failure), and the security level should not collapse when a single component is compromised. It considers only the server as a trusted component with respect to user authentication, access control and additional encryption. The server is not trusted with respect to the confidentiality of data stored at the cloud storage provider. Therefore, the server is not able to perform offline dictionary attacks. In threat model, the cloud storage provider is honest but curious, meaning that it carries out its tasks but might attempt to decrypt data stored by users. It also implement backup recovery scheme to recover data at the time of infrequent access. If the user doesn't login to the system means the admin will send alert to every 3 days, one week, two weeks and three weeks. Then automatically recover the data and forward it to alternate mail with mobile intimation.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

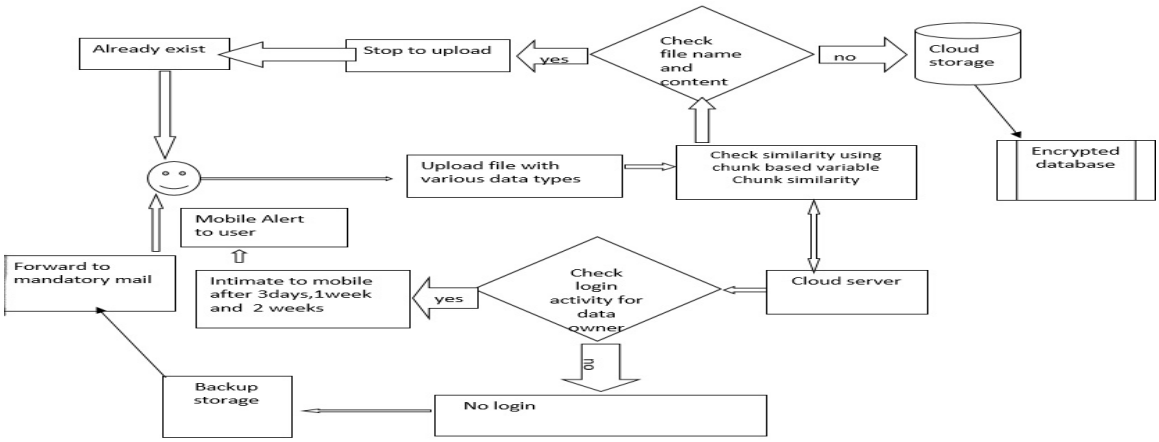
International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 7, Issue 2, February 2019

PROPOSED SYSTEM ARCHITECTURE



PROPOSED SYSTEM ADVANTAGES:

- File name and file content are analyzed
- Dynamic updation can be implemented in cloud storage
- Check all the blocks of files in terms of chunks
- Security is high and to provide data integrity to all data owner
- Proper alert is maintained by cloud server

V. MODULES DESCRIPTION

1.CLOUD STORAGE FRAMEWORK:

Cloud storage provide users and its enterprises with various capabilities to store and process their data in either privately owned, or third-party data centers that may be located far away from the world. Cloud computing refers on shared pool of resources to achieve coherence. Cloud computing means storing and access the data over internet without active interaction of the user . The term is generally used to describe data centers which is available to many users over the Internet. Huge clouds, mainly today, often have functions distributed over multiple locations from central servers. If the connection is very close to the user, it may be designated an Edge server. In this framework, it have two types of users such as data owner and data provider. The particular person or organization which legally owns a cloud service is called a cloud server. The owner of the cloud service called as cloud consumer. Cloud service provider provides the storage space to the users. Storage space can be shared by multiple data owners. Data owners can upload the files in cloud storage for future use. Cloud storage is responsible for storing the data.

2.FILE ENCRYPTION:

Encryption is the effective way to achieve data security. To read an encrypted file, the user must have access to a secret key that enables you to decrypt it. The data which is not encrypted is called plain text and encrypted data is called as cipher text. There are two main types of encryption; asymmetric encryption (also called public-key encryption) and symmetric encryption (also called private-key encryption),we can implement symmetric encryption for encrypt the data files using single key approach. The Encrypted File System, or EFS, provides an additional level of security for files and directories. Encrypted file system provides cryptographic protection of individual files on New Technology File System volumes using a public-key system. Normally, the access control to file and directory objects provided by the Windows security model is sufficient to protect unauthorized access to sensitive information. However, if a laptop that contains sensitive or confidential data which is lost, the security protection of that data may be compromised. While Encrypting the whole files, increases security. Symmetric key Encryption which uses the

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 7, Issue 2, February 2019

same key for both encryption and decryption of plain text and cipher text. The key which is used in symmetric algorithm is identical or it might be a simple transformation to go between the encryption and decryption. The key represent a shared secret between two parties that can be used to maintain a private information link. Encrypted data can be stored in cloud server.

3.SIMILARITY CHECKING:

Data compression is a specialized technique for eliminating duplicate copies of repeating data. The meaningful terms are intelligent (data) compression and single-instance (data) storage. This technique is used to increase storage utilization and can also be applied to network data transfers to reduce the number of bytes which should be sent. In the compression process, unique chunks of data, or byte patterns, are identified and stored during a process of analysis. As the analysis continues, remaining chunks are compared to the saved copy and whenever there is a match, the redundant chunk is replaced with a small reference that points to the stored chunk. In this module, can check the files using file name with file contents. Encrypted files are splited into chunks. Service provider checks the chunks at the time of uploading files. Data owner only upload original file so save storage space in cloud system. Then can compress all types of files such as text file, document file, image file and also video files.

4. ALERT SYSTEM:

It can design application for alert system for every week. After four weeks completed, if there is no access means the files are automatically sent to alternate mail and mobile which are stored at the time of registration. Server can save huge amount of storage and provide to other users.

5. BACKUP RECOVERY APPROACH:

Admin can check access time for each user login. If user login to the system means, the activity is registered in storage, and also monitored each user access. If the user access is paused more than 3 days means, admin automatically send alert to user based on registered mobile numbers. Finally if there is no access in storage system means, backup is generated. And flush the storage space and save storage for server for future use.

VI. RESULTS

Cloud computing helps to keep the data elastic in nature the user can use the cloud service as a pay per use facility and need not pay for any extra hardware. Even though the cost is reduced cloud services provide a good quality of service. Cloud storage can handle huge amount of data without any burden. The operation overhead is managed by moving the appropriate type of files to the cloud. Deduplication is good but is difficult to achieve on the encryption level. If two users are trying to upload the same data but the data is encrypted by different keys then the encrypted data is difficult to analysis for similarity. The encryption using convergent key can play an important role. In the proposed system the owner of the cloud service provides his cloud service to various users who want to use it. The users can easily upload and download their respective files from the cloud. The files are stored on the cloud in the encrypted format. The files get decrypted only when the authorized user downloads the specific file, hence security is preserved. In the backend only one original copy of each file is stored and if the same file is uploaded by many users only a link is made to the original file, which saves enormous amount of memory on the cloud. Management of data has become very common for cloud services. Cloud services prefer to focus on their core business than to manage huge amount of data getting added each day. This work studied various aspects of deduplication of data. The various needs of the cloud services such as the data management, data deduplication, and encryption were studied. The project implemented a scenario where the cloud service can deduplicate the uploaded data. The data users are provided with the proper data. The data was prevented from unwanted exposure and unauthorized access by placing a proper access control mechanism. Authorized data deduplication aims at data security to keep the data secured and avoid unauthorized access. Deduplication at encryption level saves a lot of memory and it can be utilized efficiently.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 7, Issue 2, February 2019

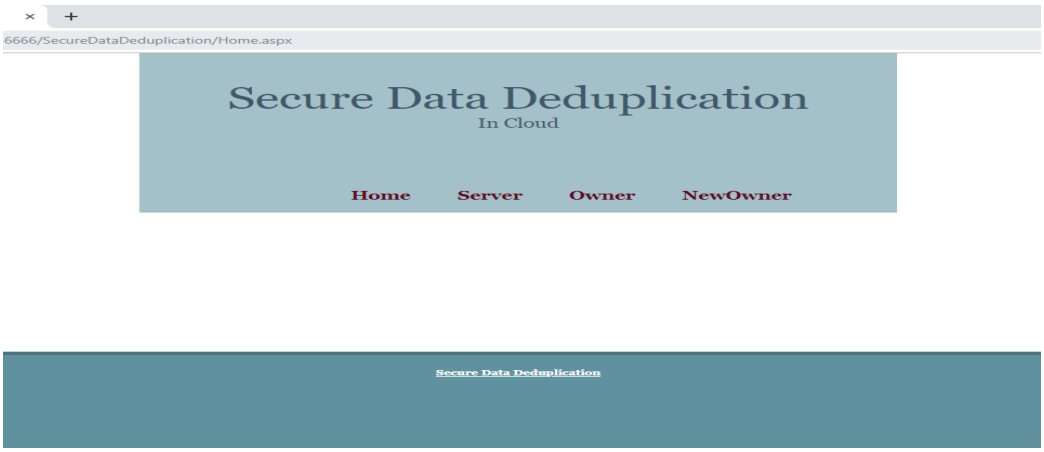


FIGURE 1: HOME PAGE

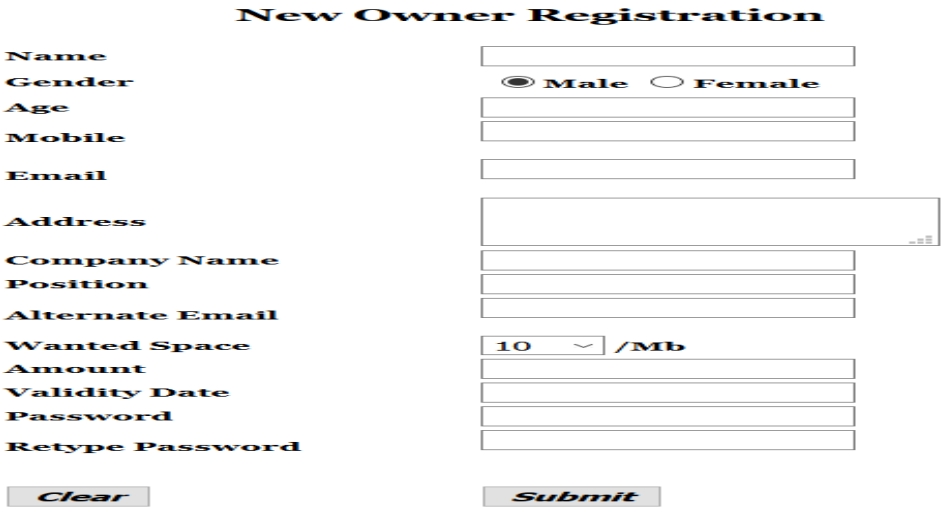


FIGURE 2: REGISTRATION FORM

FIGURE 3 : SERVER LOGIN PAGE

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 7, Issue 2, February 2019

FIGURE 4: OWNER LOGIN PAGE

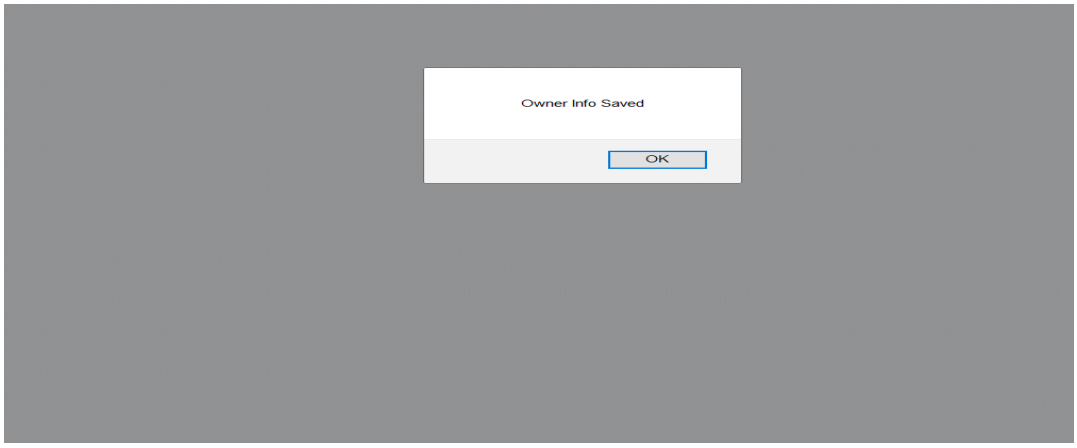


FIGURE 5:OWNER INFO SAVED

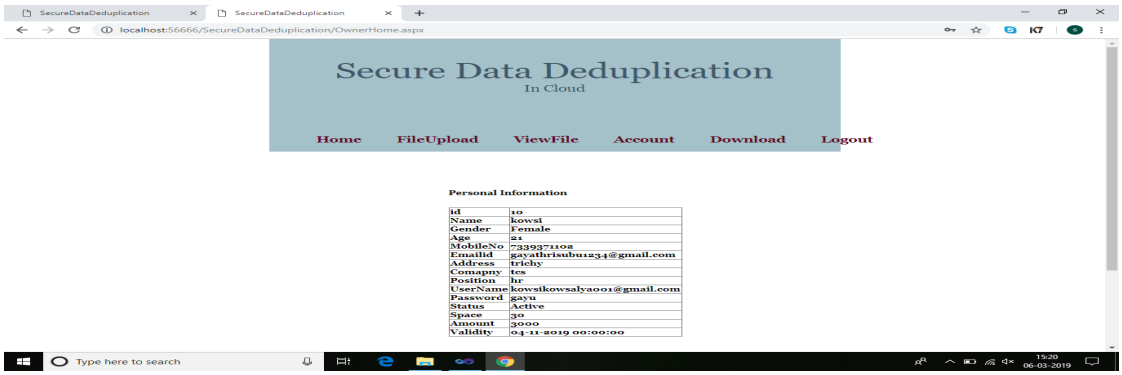


FIGURE 6: OWNER INFO

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 7, Issue 2, February 2019

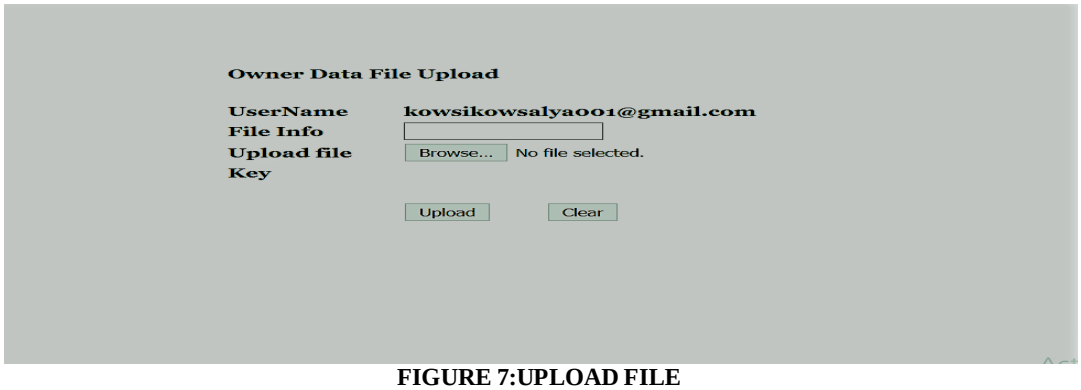


FIGURE 7:UPLOAD FILE

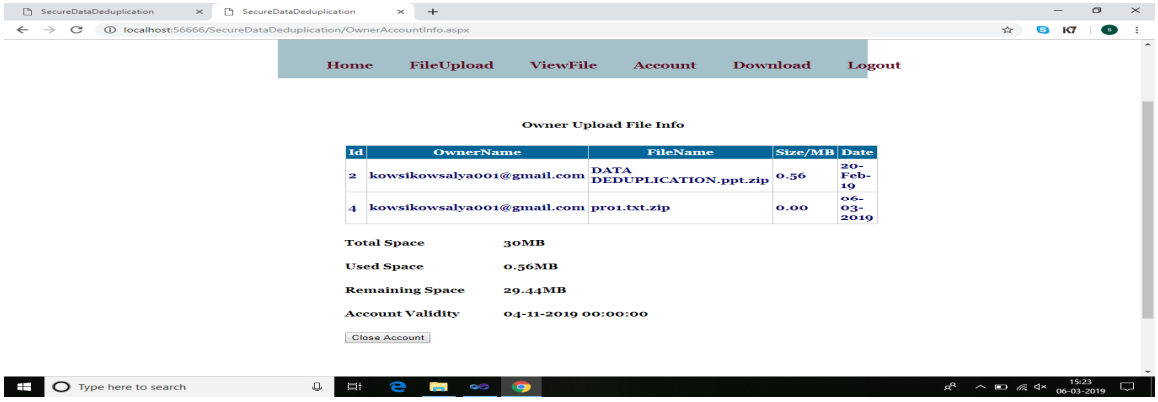


FIGURE 8: VIEW UPLOADED FILE

FIGURE 9: FILE DOWNLOAD

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 7, Issue 2, February 2019

VII. CONCLUSION

This system proposed the distributed compression systems to improve the reliability of data while achieving the confidentiality of the users and also shared authority outsourced data with an encryption mechanism. Then implemented the compression systems using the secret sharing scheme and demonstrated that it incurs small encoding/decoding overhead. In this work, have identified a new privacy challenge during data accessing in the cloud computing to achieve privacy-preserving access authority sharing for similarity files. Authentication is established to guarantee data confidentiality and data integrity. User privacy is enhanced by access requests to privately inform the cloud server about the users access desires. The backup recovery scheme is to improve the recovered scheme to avoid the blockages and also refund the amount to unused spaces in cloud system.

FUTURE WORK

In future, the framework can be extended to implement various encryption algorithms to improve the security and also be implemented in real time audio compression storage systems for various real time applications.

REFERENCES

[1] D. T. Meyer, and W. J. Bolosky, “A study of practical deduplication,” Proc. USENIX Conference on File and Storage Technologies 2011
[2] W. K. Ng, W. Wen, and H. Zhu, “Private data deduplication protocols in cloud storage,” Proc. ACM SAC’12, 2012.
[3] N. Baracaldo, E. Androulaki, J. Glider, A. Sorniotti, “Reconciling end-to-end confidentiality and data reduction in cloud storage,” Proc. ACM Workshop on Cloud Computing Security, pp. 21–32, 2014.
[4] P. Anderson, L. Zhang, “Fast and secure laptop backups with encrypted de-duplication,” Proc. USENIX LISA, 2010.
[5] J. Li, X. Chen, M. Li, J. Li, P. Lee, and W. Lou, “Secure deduplication with efficient and reliable convergent key management,” IEEE Transactions on Parallel and Distributed Sytems, Vol. 25, No. 6, 2014.
[6] Stanek, Jan, Alessandro Sorniotti, Elli Androulaki, and Lukas Kencl. "A secure data deduplication scheme for cloud storage." In Financial Cryptography and Data Security, pp. 99-118. Springer Berlin Heidelberg,2014
[7] J. Li, Y. K. Li, X. Chen, P. Lee, and W. Lou, “A hybrid cloud approach for secure authorized deduplication,” IEEE Transactions on Parallel and Distributed Systems, Vol. 26, No. 5, pp. 1206–1216, 2015.
[8] M. Bellare, S. Keelveedhi, T. Ristenpart, “DupLESS: Server aided encryption for deduplicated storage,” Proc. USENIX Security Symposium, 2013.
[9] M. Bellare, S. Keelveedhi, “Interactive message-locked encryption and secure deduplication,” Proc. PKC 2015, pp. 516–538, 2015.
[10] L. Mingqiang, C. Qin, P.P.C. Lee, and J. Li, “Convergent Dispersal: Toward Storage-Efficient Security in a Cloud-of- Clouds,” Proc. USENIX Conference on Hot Topics in Storage
and File Systems, 2014.
[11] J. Li, X. Chen, X. Huang, S. Tang, Y. Xiang, M. Hassan, and A. Alelaiwi, “Secure Distributed Deduplication Systems with Improved Reliability,” IEEE Transactions on Computer, Vol. 64, No. 2, pp. 3569–3579, 2015.
[12] Adya, Atul, William J. Bolosky, Miguel Castro, Gerald Cermak, Ronnie Chaiken, John R. Douceur, Jon Howell, Jacob R. Lorch, Marvin Theimer, and Roger P. Wattenhofer. "FARSITE: Federated, available, and reliable storage for an incompletely trusted environment." ACM SIGOPS Operating Systems Review 36, no. SI (2002):1-14
[13]Arunkumar, G., and Neelanarayanan Venkataraman. "A novel approach to address interoperability concern in cloud computing." Procedia Computer Science 50 (2015): 554-559.