Improving Efficient Reliability Based Secure Deduplication Technique on Encrypted Big Data

K Uma[1], E Jayabalan[2]

Research Scholar, Department of Computer Science, Government Arts College, Salem, India[1]

Assistant Professor, Department of Computer Science, Government Arts College, Salem, India[2]

ABSTRACT: Deduplication is most important technique in cloud computing, it's offered to eliminating duplication from storage file in cloud. Deduplication technique also providing some advantages, which is new security, facing attacker and privacy challenges. Now a day's space and network charges are not a problem. The big problem in cloud computing is space balancing and hanging. Also we are facing in cloud security. Now a day's clients are increasing their data volume. Most of the organization facing attacker problems like Masquerade attacks, Host based attacks and Network based attacks are there in cloud. In cloud computing most of the company's uploaded huge number of data. While uploading enormous data, there occurring some hanging problem. In any type of organization accessing their data through internet, for that cloud service provider providing more space and secure authentication id through mail access. But most of the hacker still try to hacking others cloud account using new technique and number of users still hacking other account with authentication for that in this article proposing new algorithm. The proposed algorithms are to improving their authentication security. And also cloud user only can access their file without permission no one cannot access their data. In this article mainly we focused on the file sharing without duplication and with security.

KEYWORDS: Deduplication, Secret sharing, Reliability, cloud Based Intrusion Detection, Activation, Distributed Storage file.

I. INTRODUCTION

Cloud computing is more friendly user for organization. It is a internet based computing that provides shared files and computer resource. Also it's proving limitless "virtualized" resources to cloud users as services across the whole internet, while masking platform and implementation details. Today's cloud service providers make an effort both highly available storage and massively parallel computing resources at comparatively low costs. The third-party companies stored the data in cloud storage environment in the space offered.

Virtualization machine and physical machines are using for stored data of own purpose. Here, physical machines are using on particular space depend upon our amount of memory. Also the virtualization machine are using to store data depend upon the amount of bandwidth. Virtualization provided unlimited resource with higher secure service. In [1], article PDP methods are using for Message authentication code it's not only for bandwidth service it's also MAC service also providing.

The remote handling technology of cloud computing is comes under big storage are for accessing their data at anywhere. The cloud computing is computationally performed at many real time problems. The every aspect of cloud services is to provide amazing flexibility to the users. The enormous data are allocated in the memory contributing with many individual users as well as organization. The replications of data are possible to the cloud in different type of cloud users. The common content of files are storing means the memory space, energy, process time are perform at slowly. Overcome the situation in cloud de-duplication is necessary one for storage servers.

The compressing technique of de-duplication is basically derived that avoid repeated data. The two type of de-duplication is available in cloud. One is file level and another one is block level. De-duplication process impossible for invalid users. It is an authorized pattern for accessing high volume of data in cloud. The security concerns multiple technology improved for better solution.

Secure and confidentiality is major roles on cloud computing. In some time it can give less performance in cloud because attacking hackers. Therefore authorized de-duplication only give effective result. To maintain data in high priority without hacking hackers on cloud is must need for de-duplication. Reduction and saving energy is securely maintained for authorized de-duplication in cloud.

For reducing attacker we proposed the CIDS algorithm which means network based attacker, Masquerade attackers and host based attackers. IDS handle easy targets for cloud user. CIDS architecture is scalable and elastic without central coordinator [10].

II. RELATED WORKS

X. Sun, et al [7], Proposed Minimum execution time (MET) and Minimum Completion Time (MCT) Algorithm. MCT algorithm providing gathers statistics to determine the node with the MCT. MET Algorithm, on the other hand, with the MET allots the unprocessed tasks to the node. OLB algorithm always keep the node is busy. It does not mind each node is the current workload. Also it is easy and direct. In 2009 taking place due to load imbalance, also Gmail crashed around the world several times. Then, some problem occurred easily at the same time while Gmail crashing. For example, the long time or control the traffic when the next work is insufficient while the demander has to wait.

Tin-Yu Wu, et al [8], Proposed Index Name Server (INS) and Domain Name Server (DNS). This paper working for the duplication and also maintaining data as soon as possible balance the data through the sharing. While sharing the data are distributed to others. Single node is giving remote backup for efficient comparison. INS process not the file compression, chunk matching, data Deduplication real-time feedback control IP information and busy level index monitoring, But the file storage, optimized node selection, and server load balancing. MD5 are using for secured through finger print. If the finger print is match then only can access ay type of organization data.

Rongmao Chen, et al[6], proposed Block Level Message-Locked Encryption (BL-MLE) for security of large number of file. While Message sharing to lock the text through internet for encrypted method. This is fully secured of the Message sharing. This scheme is proven to secure in the random oracle model and it is possible to design efficient BL_MLE schemes. Therefore it's is possible to design more over effective.

Zheng yan , et al [3], proposed the ownership challenge and Proxy re-encryption. It's handle the encryption method for big data, also it is more flexible method for sharing data and then can upload file and update with Deduplication even the data holder are offline. Encrypted data can be obtained because the data holders are authorized. Also it can obtain the symmetric keys used for the data decryption.

Shengmei Luo, et al [4], proposed algorithm is EMC's stateful routing algorithm for BOAFFT methods it's given to achieve scalable throughput and capacity using capacity multiple data severs for data Deduplication. BOAFFT method is more efficient data routing algorithm for random disk reads and writes of large amount of data. It constructs a hot fingerprint cache in each data server based access.

Jin Li, et al [2], proposed new distributed technique for Deduplication system with higher reliability in which the data chunks are across in multiple cloud serves. The security improvement is more efficient to stored the data in cloud with proof of ownership. The proposed system is demonstrated that the incurred realistic environments are very limited. They are compared with PDP, PoR for recover his outsource data through interactive proof of server. The proof of data possession (PDP) introduced the concept is Ateniese et al[1].

Jin Li, Yan Kit Li, et al [5], proposed hybrid approach for supporting authorized duplicate check in cloud architecture. Security analysis is secure in terms of our prototype for testbed experiments. Its improving their duplication approach through hybrid method in cloud service provider using convergent Encryption and proof of ownership in 2014.

Pasqalepuzio et al [9], proposed cloud dedup, a secured and systematic key storage service which assures block-level de-duplication, file level de-duplication and data trustworthy at the same time.

III. METHODOLGY

Our scheme provides to manage the encrypted file for Security of organization data. Because organizations uploading huge number of files for security purpose. Cloud Service provider providing efficient and secure space for data. Multi users can access at the same time. But still attackers improved their access method with authentication. For that in this article proposed new technique to improve efficient file sharing without duplication and with efficient. Its

providing efficient technique to improve more than existing system without file. From existing system they proven by 92% without duplication file. Here, in this article trying to increasing to 93%. Through IBM, its more over efficient and flexible for user. IBM provides 5GB space for trial space. Beyond that cost of service is high. IBM is one of the biggest storage service providers. Its providing effective service. To the work developed by the Blue mix service provider. It's one of the tools for cloud service provider. Deduplication technique is more over eliminating duplicate file form the cloud user. Here we design to balance the data while sharing. Few IBM services are very costly. Then the cloud sim and cloud analyst also the same few GB they may give free.

The cloud users each one has secret key for file sharing through that we can save the file. While uploading file we can access the data is duplicate or not duplicate also know as it is which type of duplicate which means file level deduplicate and block level deduplicate. Cloud service provider can find the file is which type of duplicate file. If it is duplicate it's automatically move to deactivation then if it is non-duplicate file means activate the file automatically. If u want to activate again means can activate and can share the file through internet. It is on- demand service provider.

IBM Service provider is giving comparative and resource management to cloud user. Also it reduces Masquerade attackers while new technique that is CIDS Algorithm. And also a using trible key algorithm. Without authorization no one can access.

BLOCK DIAGRAM
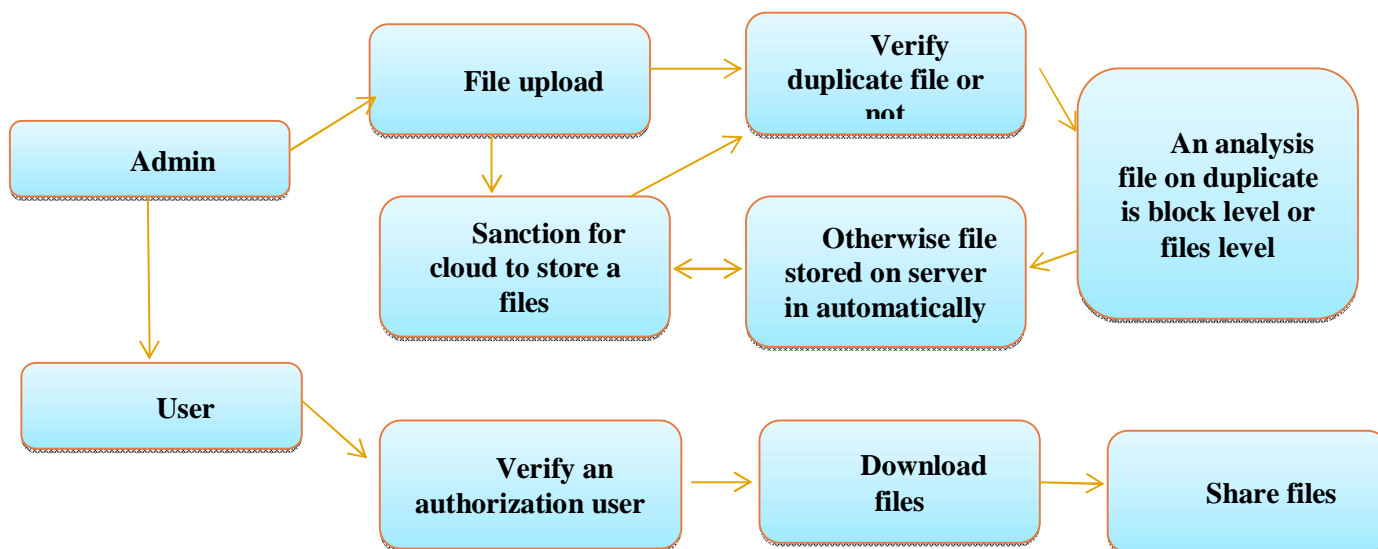
IV. EXPERIMENTAL RESULTS

In generally, Deduplication techniques are providing to calculating data volume and managing data sharing with efficient.

Data Set: In any type of files are upload in cloud there is limit depend upon the amount of data space. For that PDF file cannot upload while trial use.

Data Set	Accuracy in physical machine	Accuracy in virtual machine
PDF	high	low
TEXT	high	high
Word pad	high	high

Table.1. Deduplication accuracy Compare between PM to VM

Number of Stored Files: We can upload 10 to 1000 numbers of files. Store and retrieve the number of file and data. Token checking is done with the help of hash code.
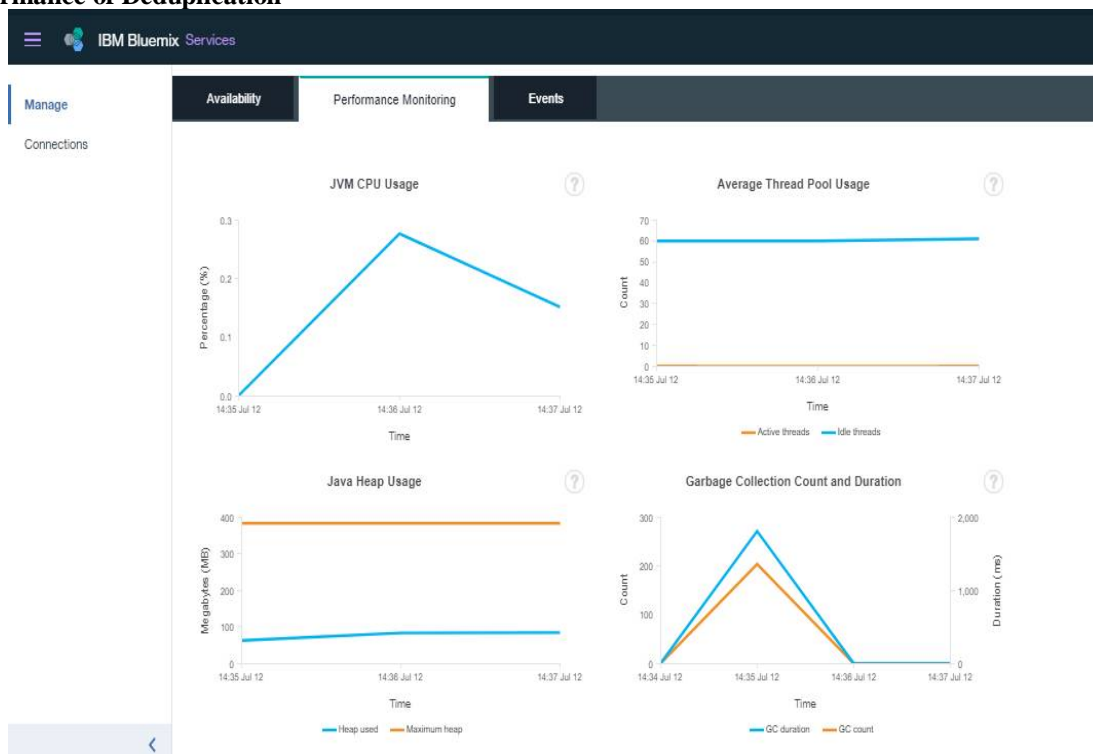
Performance of Deduplication

Fig.1. Deduplication performance monitoring report
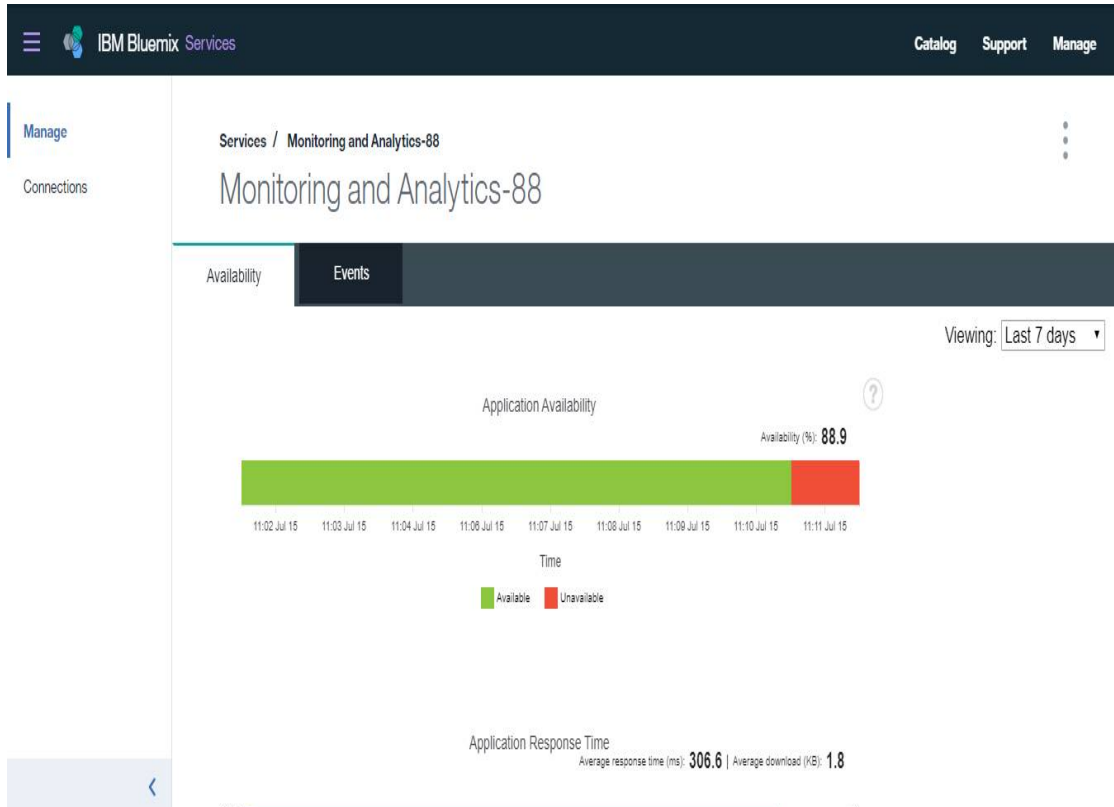
Fig.2. Deduplication availability report
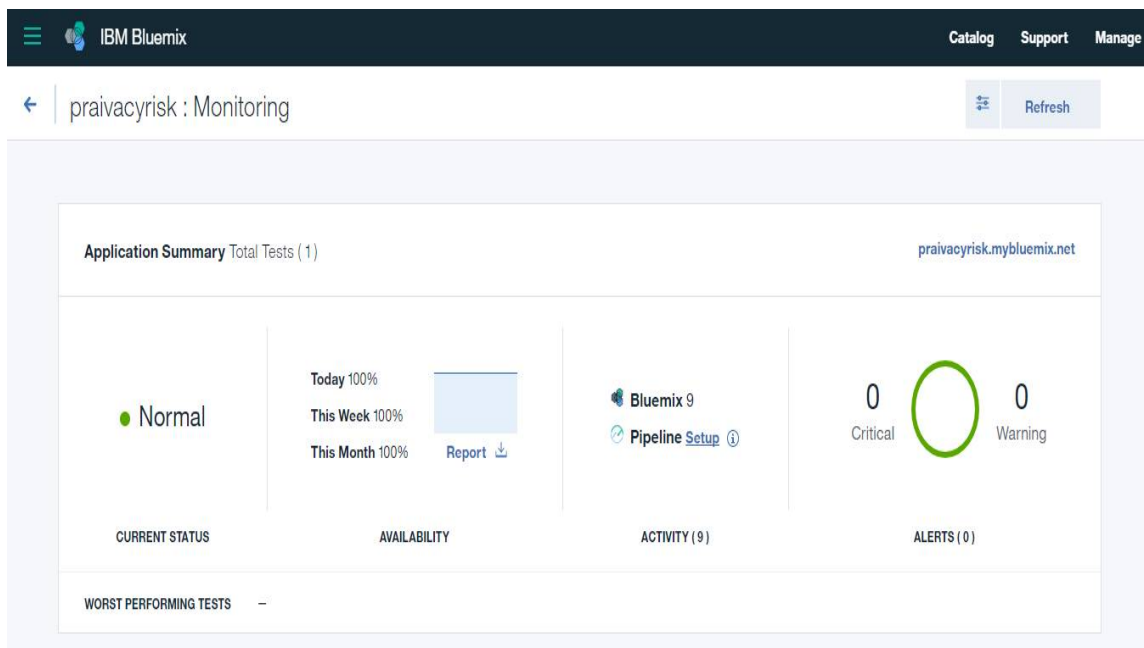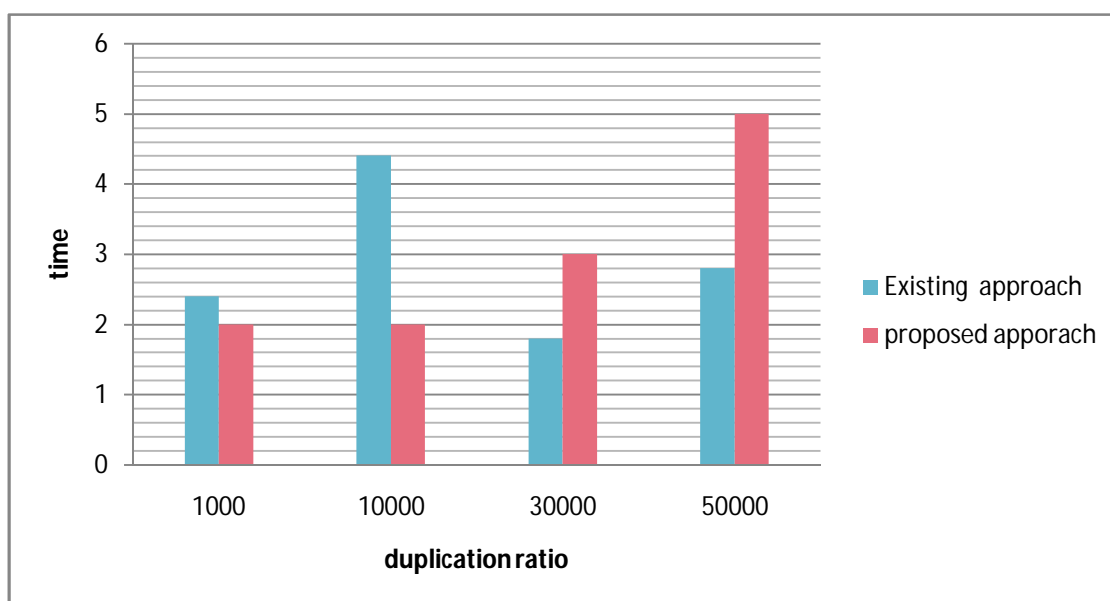

Fig.3. Application response time

Fig.4. privacy risk monitoring

Deduplication Ratio The Deduplication ratio to evaluate the effective work, we prepared two unique data, each of which consists of 10 200MB files. Here, uploaded two data sets are clear and effective file. We first upload the first data set as an early upload. We pick a portion of 10 files for the second upload, according to the given Deduplication ratio, from the early data set as duplicate files and remaining files from the second set as unique files. The average time of uploading the second data set is presented in Figure.

V. CONCLUSION

Deduplication is very important to manage a data without duplication, Its is more important techniques for save bandwidth and secure purpose. In cloud computing is using for sharing file and storing data through internet with security. Here, any type of organization uploading massive amount of data for that every one improving new technique of our work. In this article we proposed new technique for reducing duplication and attacker. The big issue of cloud computing is attacker for that we proposed algorithm is trible key method and CIDS for improving higher efficiency, flash memory and reduce attacker. Here we have proven if duplicate file with others can need and can share with activation to anyone.

In the future work, to improve the efficiency with secure through mail access with id number. To achieve the secure via mail or message lock code. And also can do the message scanning via phone and also can share with phone scanning.

REFERENCES

[1]G. Ateniese, R. Burns, R. Curtmola, J. Herring, L. Kissner, Z. Peterson, and D. Song, "Provable data possession at untrusted stores," in Proceedings of the 14th ACM conference on Computer and communications security, ser. CCS '07. New York, NY, USA: ACM, 2007, pp. 598–609. [Online]. Available: http:// doi.acm .org/10.1145/1315245.1315318

[2]Jin Li, Xiaofeng Chen, Xinyi Huang, Shaohua Tang and Yang Xiang, and Mohammad Mehedi Hassan and AbdulhameedAlelaiwi ―Secure Distributed Deduplication Systems with Improved Reliability‖, 10.1109/TC.2015.2401017, year 2015.

[3]ZhengYan,Wenziu Ding, Xixun Yu, Haiqi Zhu and Robert H.Deng, ―Deduplication on encrypted big data in cloud‖ 2016 IEEE Transactions on Vol.2, No.2.

[4]ShengmeiLuo, Guangyan Zhang, Chengwen Wu, Samee U. Khan, and Keqin Li, Boafft: Distributed Deduplication for Big Data Storage in the Cloud, IEEE Transactions On Cloud Computing, vol. 61, no. 11, January 2015

[5]Jin Li, Yan Kit Li, Xiaofeng Chen, Patrick P. C. Lee, Wenjing Lou, ―A Hybrid Cloud Approach for Secure Authorized Deduplication‖:DOI 10.1109/TPDS.2014.2318320, IEEE Vol:PP No:99 year 2014.

[6]Rongmao Chen*, Yi Mu*, Guomin Yang, and FuchunGuo, BL-MLE: Block-Level Message-Locked Encryption for Secure Large File Deduplication‖: DOI 10.1109/TIFS.2015.2470221, 2015.

[7]X. Sun, K. Li, and Y. Liu, "An efficient replica location method in hierarchical P2P networks," in Proc. 8th IEEE/ACIS Int. Conf. Comput. Inform. Sci. , Jun. 2009, pp. 769–774.

[8]Tin-Yu Wu, Jeng-Shyang Pan and Chia-Fan Lin, ―Improving Accessing Efficiency of Cloud Storage Using De-Duplication and Feedback Schemes‖, IEEE SYSTEMS JOURNAL, VOL. 8, NO. 1, MARCH 2014.

[9]P. Puzio, R. Molva, M. Onen, and S. Loureiro, ―ClouDedup: Secure deduplication with encrypted data for cloud storage,‖ in Proc. IEEE Int. Cof. Cloud Comput. Technol. Sci., 2013, pp. 363–370, doi:10.1109/CloudCom.2013.54.

[10]Hisham A.Kholidy, Fabrizio Baiardi, -- CIDS: A framework for Intrusion Detection in cloud system, Ninth Internation conference on information Technology 2012.