6th International Conference on Multidisciplinary Research (ICMR-2019)
Osmania University Campus, Hyderabad (India)
30th-31st May 2019 www.conferenceworld.in
Conference World
ISBN : 978-93-87793-89-7

Private Message Retrieving System with Erasure Coding Method On Distributed Servers

Sakshi R. Awadhiya[1](PG Scholar)**, Prof. R.V. Mante**[2](Assistant Professor)

[1]*Department of Computer Science and Engineering, Government College of Engineering, Amaravati*

[2]*Department of Computer Science and Engineering, Government College of Engineering, Amaravati*

Abstract*—*Due to increasing threat of leaking of confidential data and management of large number of these data new concept is bought into existence. Also there is risk for privacy of user's identity as well as private data. In this work, Private Information Retrieval(PIR) problem is being studied in the presence of third party for secure distributed storage systems. This concept concentrates on the challenges of privacy and secrecy for the information of users. Also Secure distributed storage systems is designed to protect both user privacy from the databases and data security from an eavesdropper. In addition, secret sharing scheme is also used for security purpose. Secure Erasure coding method is used for node failure as well as for reduced backup storage. Our methodology insists double encryption for each data and there is no necessity to revise keys after user's addition or removal from the group where the message (data) is being shared.

Keywords- Private Information Retrieval, secret sharing scheme, data security, Erasure coding method.

I.INTRODUCTION

A Secure Distributed Servers are distributed systems where data stored are secured due to the implementation done in between the storing of data in different servers. We present our results on the information-theoretic PIR problem in a scenario where each of the databases ensures data security from an eavesdropper. The purpose of distributed storage systems is to store data reliably over long periods of time using a distributed collection of storage nodes which may be individually unreliable. Regardless of their purpose, the main service provided here is the additional security of confidential data and efficient storage space. Taking Intelligence Bureau as an example, the application is designed for specific purpose of managing the data of cases as well as the officers working on them. Nowadays, the crime of hacking the confidential data of cases and erasing the data are increasing as well as large amount of these data are hard to manage if some system broke suddenly, hence above described techniques are developed such as to secure and manage the data effectively by storing a huge amount of private and possibly sensitive information as a backup using erasure coding method which takes less space as compared to previous techniques used.

ISBN : 978-93-87793-89-7

Due to the increasing rate of hacking system, cases data and users privacy are at risk. Indeed, the confidential data and officers privacy as well as case related peoples privacy could not be maintained as well by previous techniques. Recent events have shown that, in addition to breaching of users privacy (internally or externally), third party introduce new privacy risks. To address the previous privacy issues, double security concept in addition to other techniques have considered. As for storage purpose applications involving storage in large data-centers and peer-to-peer storage systems are considered. In such a setting, our work is based on two classes of considerations for security: one is user privacy (concealing the index of the desired message) from each of the databases and the other is data security from an eavesdropper who can access to one of the databases or link between the database and the user to obtain information about messages. The contributions of this work are two-fold. We first propose a new secure distributed storage system and its corresponding PIR scheme that simultaneously ensure user privacy from the non-colluding databases and data security from an eavesdropper. We utilize a secret sharing scheme for distributed databases to preserve data security, and our PIR scheme relies on an existing PIR scheme for an information-theoretic PIR problem proposed to keep user privacy in the secure distributed databases. In contrast to existing secure distributed storage systems, the databases in our scheme possess the redundant secret shares which are exploited as a side information to increase the rate of the PIR. The first scenario considered here is that the databases do not have knowledge about which secret shares are stored in other databases. The second one is that each database is aware of the stored secret shares in other databases, which is a less favorable condition in that it can invade user privacy during a PIR procedure. Secondly, it is shown that the rates of our proposed PIR schemes are withina constant multiplicative factor from the derived upper-bound on the information-theoretic capacity of the PIR problem under secrecy constraints.

In the design of secure distributed databases and PIR process, we prevent an eavesdropper from obtaining information about the messages by preventing the individual database from obtaining any information about the messages from the stored data. In our proposed system, we will discuss about security of data on distributed systems. In distributed system, we need to deal with multiple servers at a time. To maintain security over multiple servers is a very difficult task. In distributed system, we have to keep a track of network adversaries. Here a new security technique is explained for secure message sharing on distributed servers.

ISBN : 978-93-87793-89-7

According to this technique, the message will be divided into n shares. The shares will be distributed over multiple servers. Each share is stored on every server/db, multiple no of times repetitively. Due to which the availability of the message is increased in case of any attack made on particular database. To achieve data availability in any case of node failure, it is necessary to store backup of all the shares. But on the other hand, the server space is remained consumed by the same shares of messages on multiple servers. Therefore to reduced server space required to store backup of data we proposed secure erasure coding method, in which the shares will be stored in polynomial format on backup servers. Also, to increase the security of the shares we proposed a user defined encryption technique to change the bytes of the messages before storing them on various servers. We proposed two level encryption technique using which we can improve the security of the existing system.

In our proposed system, we proposed secure message sharing over distributed network with integrity checking and data availability. In our system, the message will be encrypted on client side before transfer to the server. Then the message will be divided into n shares on server, shuffle their indices, perform second level encryption and store on multiple servers randomly. At the time of decryption, user need to specify the key, the message will be downloaded from different servers and on client side the message will be rejoined in the sequence to get combined message. After that the re-combined message will be decrypted on client side by the system automatically and deliver it to user. The keys required for client side encryption, will be maintained on the basis of upload date, time and some user defined algorithms. In the design of secure distributed databases and PIR process, we prevent an eavesdropper from obtaining information about the messages by preventing the individual database from obtaining any information about the messages from the stored data. We provide a secure scheme against not only an external eavesdropper who can access to the stored data in the database but also a possible internal eavesdropper (eavesdropping database) which intends to leak confidential information about messages and user privacy to the adversaries. In addition, we show that the lower-bound on the capacity of PIR can be further improved by modifying our scheme especially when the number of databases is two. To retrieve shares privately, the user must generates queries and send it to databases. The queries are generated without information about the messages at the user, thus they are independent of the messages. Assuming that there exists an eavesdropper who has access to one of the databases and/or the link between the database and the user. For data security from the eavesdropper, databases store securely encoded data of the messages, and thus the eavesdropper obtains no information on the messages from the stored data and the link between databases and the user.

II.LITERATURE REVIEW

To guarantee the essential attributes of storage systems such as reliability, security, and so on, the use of regenerating codes facilitates storage systems to efficiently cope with node failures where the message shares are stored in the form of codes and transferred to multiple servers. The message shares can be recovered using the codes hence there is no need to store replica of each share[1]. But lack of security and integrity check provide us a new way to do more research on the topic. In Generic repair schemes, a linear secret sharing

ISBN : 978-93-87793-89-7

schemes can be securely repaired. The author proposed another scheme where codes are stored in the form of polynomial. In case of any failure the shares can be recovered by solving the polynomials[2]. To study the secure repair bandwidth under the general repair model when the secret sharing scheme being repaired is one of the open problem. To study secure repair where active adversarial nodes are present that may deviate from the prescribed repair protocol is one of the interesting problem. Also in this scheme complexity level is too high to go through this.

In [3], the multi-round private information retrieval over distributed network is being studied where it proves that the capacity of multi-round PIR is the same as the capacity of PIR of single round. The result includes T-privacy constraints. There is a drawback of storage overhead and no advantage in terms of capacity from multi-round over single-round schemes, nonlinear over linear schemes.

In [4], a recursive techniques is proposed to hide extra information in between the parts of Shamir's secret sharing schemes. This hidden information may be used for validation of shares at the time of secret reconstruction.

Simultaneous node failures can be revealed by DSS which is needed to be recovered with local connections. The design of coding schemes for DSS satisfy these properties. No major encryption technique is justified[5].

The data stored should be right even when some servers failed are considered in Secure storage and retrieval of information [SSRI] [6]. SSRI extend a property where an adversary can corrupt servers totally but some during given time interval. It is assumed that faults can occur at reconstruction time which is major shortcomings. The capacity of PIR is especially significant because of the central role played by PIR across a diverse array of problems that include locally decodable and batch codes, secure multiparty computation, instance hiding, secret sharing, and oblivious transfer[7].

- Codes for Distributed Storage

The regenerating codes model introduced in [8] considers optimizing two important resources: the storage capacity required by each node, and the repair-bandwidth. There exists a substitute between two resources, and lower bounds on their requirements were derived. Subsequent to their work, several explicit codes were constructed for the MSR and the MBR regimes of regenerating codes, many of which meet these bounds.

- Shamir's secret sharing :

A possible method to ensure information-theoretic security from passive eavesdroppers is to employ Shamir's secret sharing scheme[14], where the data is encoded and stored in a set of n nodes such that the entire data can be retrieved from any k nodes, while access to data in any (k-1) or fewer nodes provides zero information about the data. During repair of a failed node, this scheme requires a download of the entire data to a central location, following which the replaced node's data is re-encoded. Thus, the repair operations are inefficient in classical erasure codes, mandating significant network resources.

- Secure Network Coding :

The literature on secure network coding [5] primarily considers a multicast setting where a single source of data and every destination is interested in obtaining all the data sent by the source. Furthermore, with respect to

security from passive eavesdroppers in the multicast setting, only the scenarios where the eavesdropper can access subsets of links is well understood in the literature.

III.METHODOLOGY

1. Existing Methodology

In existing methodology, the evaluation is shown so as to provide security to the whole data shared between different servers as well as data is being shared between the servers and client. But the existing system is not providing enough security from third party also it's taking too much space by storing the whole data in all servers connected for backup purposes. Hence due to information retrieval in the presence of third party needed additional security to keep the information confidential as well as storing large data as a backup will be needing much space, here comes our proposed idea. Also future scope in existing system shows the need to keep the size of original data same as after the encryption of original data, It will be covered in current work.

2. Proposed Methodology

In our proposed system, we proposed secure message sharing over distributed network with integrity checking and data availability. In our system, the message will be encrypted on client side before transfer to the server. Then the message will be divided into n shares on server, shuffle their indices, perform second level encryption and store on multiple servers randomly. At the time of decryption, user need to specify the key, the message will be downloaded from different servers and on client side the message will be rejoined in the sequence to get combined message. After that the re-combined message will be decrypted on client side by the system automatically and deliver it to user. The keys required for client side encryption, will be maintained on the basis of upload date, time and some user defined algorithms.

ISBN : 978-93-87793-89-7

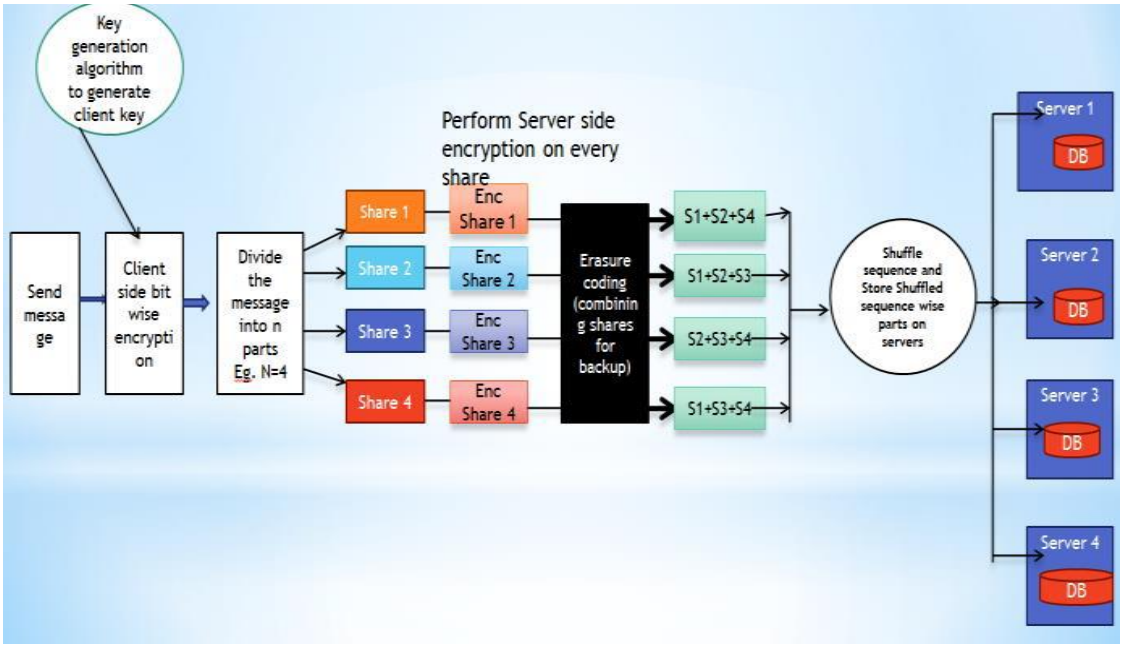

Fig a. Flowchart

Taking Intelligence Bureau into account there are various sections in which process is divided. Firstly administration panel page will be shown where various options will be given as shown in fig a. where admin will login. As admin is generally the head or controller of all the process happening, there are various work under him such as registering the officers and also registering the cases. He will also create groups where officers will be divided accordingly and can assign the cases to different groups. Group wise list is also shown in fig a. which comes in handy when new case will come. Officer group must have leader and members where leader can assemble its group members according to the work to be done on the case and also give direction to them according to the progress on investigation. In this page admin can track the server performance as well as fault tolerance where in case of any server failure, system will automatically recover the missing share using erasure codes stored on backup server.

Officer can also login using id and password which will be mailed to them as soon as admin will register them. As shown in fig b. officer can later change their password of their interest. Officer has contact with its group members and can send message about the assigned case to every member of their group members securely. The messages shared between group members will be encrypted firstly on client side by making key using the date and time of said message. The algorithm used for encryption is user defined and use time and date as input to secure the key as well as message. Server side encryption can also take place for additional security to message as well as all information backup in case of node failures.

IV.EXPERIMENTAL ANALYSIS

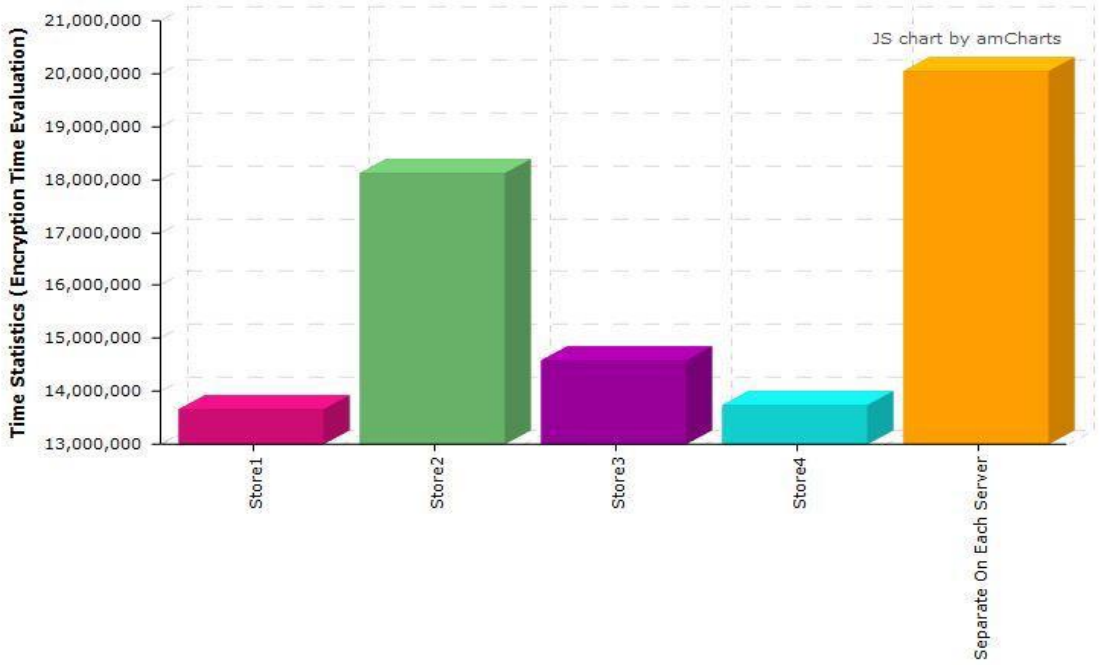

Fig 4.1. Encryption time wise Evaluation

Proposed System				Existing System
Store 1	Store 2	Store 3	Store 4	Encryption Time On Each Server
13649300	18131264	14583761	13746804	20052438

Fig 4.2. Encryption Time wise Data

As shown in Figure 4.1 and 4.2., there in our proposed system the time taken to encrypt the parts of message will be far less than the encryption time taken in existing system. Since in existing system there will be encryption taking place for whole message in each servers it will be taking 20052438 ns on each servers, whereas in our proposed system the messages will be divided according to the number of servers hence it takes time individually of parts of the message. Hence as shown in Figure 4.2, the tabular chart shows different values each store takes to calculate encryption time but each time it will always be less than existing time as average time taken will be 15027782 ns approx.

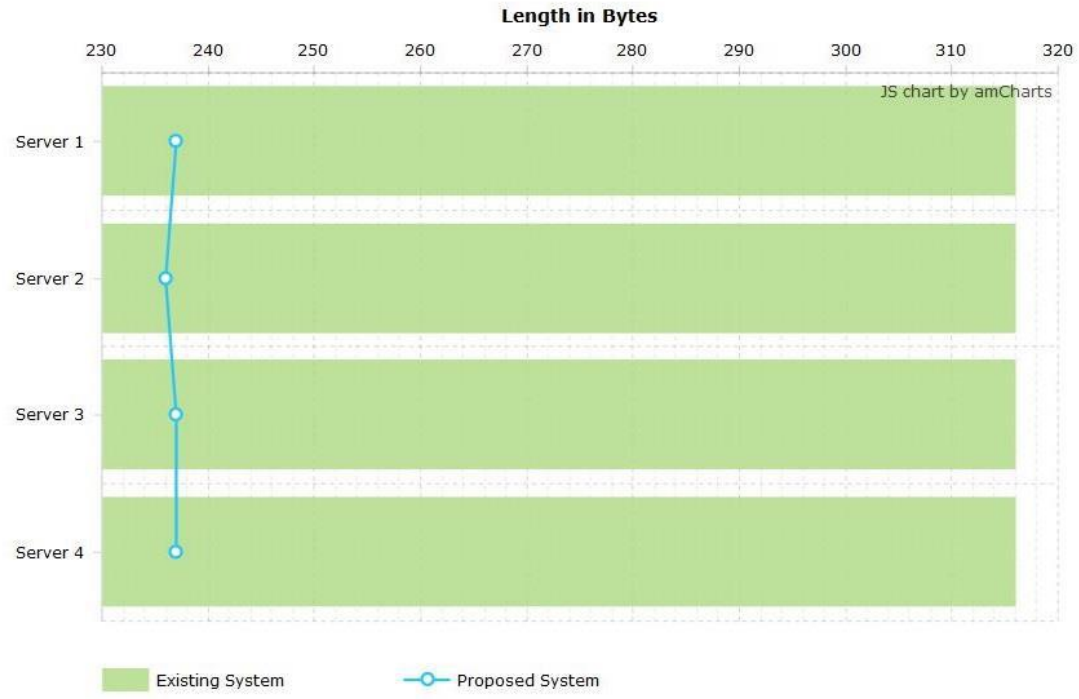

Fig 4.3. Average Storage Space Difference

Proposed System				Existing System
Store 1	Store 2	Store 3	Store 4	Average Length on All Servers
237	236	237	237	316

Fig 4.4. Table Data of Average Space Difference

Figure 4.3 and 4.4 shows the difference in average value of storage space occupied in existing system and in our proposed system. In the system since we are storing the message in encrypted format we will be showing the graph in bytes. Since in existing system whole data will be replicated in each server it takes to much space due toreplication but in our system we are using concept of erasure coding where the parts of encrypted message will be stored after double encryption in each servers as not whole message but in a format where any two server can provide whole message. Hence in fig 4.4, storage space occupied in existing system will be 316(bytes) while in our proposed system it occupies only 237(bytes) on average.

V.RESULT

A PIR problem is taken into account for distributed databases in the presence of an eavesdropper. Depending on whether the data indices in other databases are known at a database, we proposed two PIR schemes to ensure user privacy from each database and also security of data (in case there is an eavesdropper) at the same time. In

our proposed system, we proposed Secure erasure coding method. Therefore, we reduced the computational time of storing backup in case any server fails. We have also proposed user defined encryption algorithm which keeps the size of the document as it is after encryption. Therefore, our system is more efficient than the existing system.We overcome many of the existing drawbacks by managing space as well as security. Hence our system is more efficient.

REFERENCES

[1] A. Fazeli, A. Vardy, and E. Yaakobi, "Codes for distributed PIR with low storage overhead," in Proc. IEEE International Symposium on Information Theory (ISIT), Hong Kong, China, pp. 2852-2856, Jun. 2015.

[2] W. Huang and J. Bruck, "Generic secure repair for distributed storage," arxiv preprint arXiv:1706.00500, Jun. 2017.

[3] H. Sun and S. A. Jafar, "Multiround private information retrieval: capacity and storage overhead," arXiv preprint arXiv:1611.02257, Nov. 2016.

[4] A. Parakh and S. Kak, "Recursive Secret Sharing for Distributed Storage and Information Hiding," Information Sciences, vol. 181, no. 2, pp. 335-341, Dec2009.

[5] O. O. Koyluoglu, A. S. Rawat, and S. Vishwanath, "Secure cooperative regenerating codes for distributed storage systems," IEEE Transactions on Infomation Theory, vol. 60, no. 9, pp. 5228-5244, Sep. 2014.

[6] J. A. Garay, R. Gennaro, C. Jutla, and T. Rabin, "Secure distributed storage and retrieval," Theoretical Computer Science, vol. 243, no. 1-2, pp. 363-389, Jul. 2000.

[7] H. Sun and S. A. Jafar, "The capacity of private information retrieval with colluding databases," in Proc. IEEE Global Conference on Signal and Information Processing (GlobalSIP), Washington, DC, Dec. 2016.

[8] A. G. Dimakis, P. B. Godfrey, Y. Wu, M. Wainwright, and K. Ramchandran, "Network coding for distributed storage systems,"IEEE Transactions on Information Theory, vol. 56, no. 9, pp. 4539–4551, 2010.

[9] N. B. Shah, K. V. Rashmi, P. V. Kumar, and K. Ramchandran, "Distributed storage codes with repair-by-transfer and nonachievability of interior points on the storage-bandwidth tradeoff," IEEE Transactions on Information Theory, vol.58, no.3, pp.1837–1852, Mar. 2012.

[10] C. Tian, "Rate region of the (4, 3, 3) exact-repair regenerating codes," in IEEE International Symposium on Information Theory, Istanbul, Jul. 2013.

[11] B. Sasidharan, K. Senthoor, and P. V. Kumar, "An improved outer bound on the storage-repair-bandwidth tradeoff of exact-repair regenerating codes," arXiv preprint arXiv:1312.6079, 2013.

[12] Y. Han, R. Zheng, and W. Mow, "Exact regenerating codes for Byzantine fault tolerance in distributed storage," in Proc. IEEE International Conference on Computer Communications (INFOCOM), Florida, USA, March 2012.

[13] F. Oggier and A. Datta, "Byzantine fault tolerance of regenerating codes," in IEEE International Conference on Peer-to-Peer Computing, 2011, pp. 112–121.

[14] A. Shamir, "How to share a secret," Communications of the ACM, vol. 22, no. 11, pp. 612–613, 1979.

ISBN : 978-93-87793-89-7

[15] J. Feldman, T. Malkin, C. Stein, and R. Servedio, “On the capacity of secure network coding,” in Proc. 42nd Annual Allerton Conference on Communication, Control, and Computing, 2004.

[16] H. Yang, W. Shin, and J. Lee, "Private Information Retrieval for Secure Distributed Storage Systems," in IEEE Transactions on Information Forensics and Security, 2018.