ISSN: 0975-833X

Available online at http://www.journalcra.com

International Journal of Current Research
Vol. 10, Issue, 12, pp.76370-76373, December, 2018
DOI: https://doi.org/10.24941/ijcr.33581.12.2018

INTERNATIONAL JOURNAL OF CURRENT RESEARCH

RESEARCH ARTICLE

PRIVATE MESSAGE SHARING SYSTEM ON DISTRIBUTED SERVERS WITH ERASURE CODING METHOD

***Sakshi R. Awadhiya and R.V. Mante**

Department of Computer Science and Engineering, Government College of Engineering Amravati-444 604, Maharashtra, India

ARTICLE INFO

Article History:
Received 20th September, 2018
Received in revised form
30th October, 2018
Accepted 29th November, 2018
Published online 31st December, 2018

Key Words:
Private Information Retrieval,
Secret Sharing Scheme,
Data Security, Erasure Coding Method.

ABSTRACT

In this work, Private Information Retrieval (PIR) problem is being studied in the presence of third party for secure distributed storage systems. 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.

***Citation: Sakshi R. Awadhiya and R.V. Mante,* 2018.** "Private Message Sharing System On Distributed Servers With Erasure Coding Method", *International Journal of Current Research*, 10, (12), 76370-76373.

INTRODUCTION

A Secure Distributed Servers are distributed systems where stored data are secured during the implementation done in between the storing of data in different servers. Information-theoretic PIR problem is used where each of the databases ensures data security from an eavesdropper. Here our work is based on two classes for security: i) 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. A new secure distributed storage system and its corresponding PIR scheme is studied that simultaneously protect user privacy from the non-colluding databases and data security from an eavesdropper. A secret sharing scheme is made for distributed databases to preserve data security, and our PIR scheme relies on an existing PIR scheme to keep user privacy in the secure distributed databases. In contrast to existing secure distributed storage systems, the redundant secret shares which are exploited as a side information to increase the rate of the PIR are possessed by databases. We consider two different scenarios according to the availability of the information about which secret shares are stored in other databases. The first scenario is that there is no knowledge to database about what 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 can invade user privacy during a PIR procedure. Rate of our proposed PIR schemes should be in between a multiplicative factor (which is always constant) of the upper bound that is already derived on information theoretic capacity of PIR problem. 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 where multiple servers will run at a time. A new security technique is explained for secure message sharing on distributed servers. According to this technique, the message will be patitioned into n shares. The shares will be distributed over multiple servers. 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 became important to store backup of all the shares. But in other case, 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. Two level encryption technique can be used which can improve the security of the existing system. Here we proposed secure message sharing over distributed network with integrity checking and data availability. In the design of secure distributed databases and PIR process, we prevent an

***Corresponding author:* Sakshi R. Awadhiya,**
Department of Computer Science and Engineering, Government College of Engineering Amravati-444 604, Maharashtra, India.

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. We will also focus on the lower-bound on the capacity of PIR that can be further improved by modifying our scheme especially when the number of databases is less. 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. We consider two possible scenarios whether or not each database knows which secret shares are stored in other databases by coordination between databases:

- **Databases without coordination (Scenario 1):** The databases do not coordinate each other, thus they do not know the set of the indices of the secret shares stored in other databases, i.e., DB 1 has no information about its shares. This scenario can be realized when the main server operates the storage system does not share this information or it outsources the external databases to use their storage capabilities.
- **Databases with coordination (Scenario 2):** The databases are aware of the set of the indices of the secret shares stored in other databases by coordination.

Literature survey: 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 (Fazeli, 2015). In Generic repair schemes, a linear secret sharing 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 (Huang, 2017). 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. In (Sun, 2016), 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 (Parakh,. 2009), 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 (Koyluoglu, 2014). The data stored should be right even when some servers failed are considered in Secure storage and retrieval of information (SSRI) (Garay, 2006). 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 (Sun, 2016).

Codes for Distributed Storage: The regenerating codes model introduced in (Dimakis, 2010) 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. Furthermore, it was shown in (Shah, 2012; Tian, 2013; Sasidharan, 2013), that the bounds are loose at essentially all points in the interior of the tradeoff curve. The results of this paper are based on the product-matrix codes, and exploit certain unique features of the underlying product-matrix framework. The requirement of security where repair dynamics is present was first considered in "Securing dynamic distributed storage systems against eavesdropping and adversarial attacks" where the authors marks the storage lower bounds and bandwidth requirements under such a setting. In (Han, 2012), the authors deal with fault tolerance in byzatine by employing product-matrix codes. They use a cyclic redundancy check (CRC) to check the integrity of data during repair and reconstruction, and a feedback scheme to iteratively correct them. However, CRC based schemes are not applicable in the present setting of information-theoretic security since the CRC may also be corrupted by the adversary. In (Oggier, 2011), the authors derive bounds to determine the secure capacity in a "cooperative-repair" setting. The bounds show that such an attempt to cooperatively repair may adversely affect the system where malicious adversaries are present.

Shamir's secret sharing: A possible method to ensure information-theoretic security from passive eavesdroppers is to employ Shamir's secret sharing scheme (Shamir, 1979), 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 (Feldman, 2004) 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. The problem of secure distributed storage considered in this paper requires handling the case when nodes are compromised. The problem of node-compromise is typically treated as a case of link-compromise by assuming that the eavesdropper gains access to all links that are incident upon the compromised nodes. Schemes providing secrecy with only reconstruction requirement still makes the problem non-trivial with the requirement of node-repair addressing.

Private Information Retrieval: Chor, Goldreich, Kushilevitz, and Sudan first provided the idea of private information retrieval(PIR). The database is viewed as a binary string in the PIR model and it is shown that user can retrieve a single bit without revealing any identity about its index. But having two replicas can reduce the communication cost through which the communication complexity of k-server PIR has been further reduced in a series of ground breaking papers (Chor *et al.*, 1998). In another recent work, Chan *et al.* proved the substitution between storage overhead and communication complexity, for setups in which the size of each file is relatively large. Augot *et al.* studied the concept where the database are partitioned into several shares to avoid repetition and reduce the storage overhead. However, they didn't encode the parts of the database as we study in this work. In order to overcome the complications occurred by above methods, all known k-server information-theoretic PIR protocols can be efficiently emulated significantly removing the storage overhead. In (Gasarch, 2004), information-theoretic PIR is considered where privacy is strongest. While PIR is not as fundamental, it is both connecting to fields of interest and using interesting techniques. The biggest frustration about PIR's is the lack of good lower bounds. The topic of private information retrieval (PIR) has been fairly well explored in the literature on theoretical computer science and cryptography. However, all the previous works assume a replication-based setting. In another paper it is considered PIR when data is stored using erasure codes. Hence, we hope that our single-database computational PIR will be useful for the design of other cryptographic protocols. Many problems remain open. How do we reduce the intractability assumptions? Impagliazzo and Rudich show that implementing oblivious transfer based on general one-way functions (i.e. without trapdoor) is hard to do using black-box reductions. However, how to achieve communication-efficient cPIR is not known by assuming trapdoor one-way permutations can be implemented based on any trapdoor one-way permutations . How can this be done in cPIR setting? Basing cPIR solutions on other algebraic problems could also be of interest; for example, based on our paper, Man shows how to replace the quadratic residuosity problem in our protocol by the shortest lattice vector problem. Additionally, even assuming that our security parameter is polylogarithmic in n, we only achieve 2O(plog nlog log n) communication complexity. Existing work on private information retrieval (PIR) problems largely focused on uncoded data storage (where every storage node stores all the data records). Their main focus was to design a retrieval scheme which has the lowest total upload and download (for transmission of queries retrieved and data retrieval) costs. A fundamental question in PIR problems in coded storage is the characterisation of the substitution between storage costs and retrieval costs.

Regenerating codes: Crucial idea about regenerating codes is of sub-packetization. Each packet consists of some sub-packets, and when packet storage by a node fails then sub-packets from other nodes can be sent for recovery. Experience with data centers however suggests that extra storage nodes which are accessed is related to a considerable overhead. Hence there is not necessarily the right single measure of the recovery time by pure bandwidth consumption.

Locally decodable codes: These codes were introduced where a r-query Locally Decodable Code (LDC) encodes messages in a way that even after 10% of codeword coordinates are erased one can easily recover any message symbol by accessing only r codeword symbols. Thus LDCs are in fact very similar to (r, d)-codes with an important distinction that LDCs allow for local recovery even after number of symbols is erased, while (r, d)- codes provide locality only after a single erasure. Not surprisingly locally decodable codes require substantially larger codeword lengths then (r, d)-codes. The scheme encodes the database as a function over a group and the user will only be able to recover the bits of the database from the function. The reason is that, when one translates 2-server PIR schemes into LDC, the resulting alphabet of the code can be quite large. Thus, each one of the two queries used by the decoder is a string of s bits.

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.

Conclusion

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. The rates of our PIR schemes are also concluded to be within a constant multiplicative gap from the upper-bound on the capacity of the considered PIR problem. We overcome many of the existing drawbacks by managing space as well as security. Hence our system is more efficient.

REFERENCES

"Private information retrieval," J. ACM, 45, Earlier version in FOCS 95.

Chor, B., Kushilevitz, E., Goldreich, O. and Sudan, M. 1998.

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

Fazeli, A., Vardy, A. and Yaakobi, E. 2015. “Codes for distributed PIR with low storage overhead,” in Proc. *IEEE*

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

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

Gasarch, W.I. 2004. A survey on private information retrieval (column: Computational complexity), Bulletin of the EATCS, vol. 82, pp. 72–107.

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

Hua Sun and Syed A. Jafar. 2016. “Blind Interference Alignment for Private Information Retrieval”.

Huang W. and Bruck, J. 2017. “Generic secure repair for distributed storage,” arxiv preprint arXiv: 1706.00500, Jun.

International Symposium on Information Theory (ISIT), Hong Kong, China, pp. 2852-2856, Jun..

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

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

Parakh A. and Kak, S. 2009. “Recursive Secret Sharing for Distributed Storage and Information Hiding,” Information Sciences, vol. 181, no. 2, pp. 335-341, Dec.

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

Shah, N.B. K. V. Rashmi, P. V. Kumar and K. Ramchandran, 2012. “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.

Shamir, A. 1979. “How to share a secret,” Communications of the ACM, vol. 22, no. 11, pp. 612–613.

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

Sun H.and Jafar, S. A. 2016. “Multiround private information retrieval: capacity and storage overhead,” arXiv preprint arXiv:1611.02257, Nov.

Tian, C. 2013. “Rate region of the (4,3,3) exact-repair regenerating codes,” *in IEEE International Symposium on Information Theory*, Istanbul, Jul..

****