Design of a Secure and Authorized System for Data Deduplication

Srinivas L N B, Rajasekar P

Abstract- De-duplication is a technique of compression of data and it eliminates duplicate copies of repeating data. To protect the data security better, the problem of authorized data de-duplication is addressed. The idea is presented which is used for differentiating data according to their popularity. The problem of authorized de-duplication is addressed in this paper. To remove duplicate data from database is the main aim. It also helps for saving the storage space and bandwidth. Data de-duplication makes the system more secure and it eliminates the file of same name .This paper gives benefit to both the user and storage provider by the de-duplication technique. Key system is used in this project to support the security feature. Because of the rapid rise in the number of users, number of files created is increasing day by day. This paper shows how the security is obtained in database during data de-duplication process.

Index Terms— Data Compression, duplication, protection, security, deduplication

1 Introduction

The increase in the volume of data has raised major problem in the data protection. We have used a concept in this paper that has been used for storing huge amount of data and resources sharing. Data de-duplication is a concept that has been useful in reducing the cost. Storage utilization can be improved by the technique of data de-duplication [1]. It. De-duplication technique removes the data which is repeated and keeps only one copy. De-duplication method can be executed in many ways and there are basically three types of data de-duplication i.e. file level, block level and byte level. The database storage can handle huge amount of data [2]. The good quality of service is provided and cost is also reduced. De-duplication is a very good technique but on the encryption level it is very difficult to achieve. The encryption done using convergent keys plays an important role in the data de-duplication. The OTP feature is added to this paper to make the system more secure.

To put it in a simple manner, file level data deduplication is one among the two popular approaches [3]. It actually eliminates duplicate copies of similar file. This methodology is also referred as single instance storage (SIS). The other effective approach is block level approach. The block level data deduplication method eliminates duplicated blocks of data that occur in non-identical files [4]. When these two approaches are compared with one other, the second approach (i.e.) block-level deduplication frees up more space than compared to file level deduplication.

The remainder of this paper is organized as follows. Section2 deals in detail about the relevant work carried out by researchers in this domain. In Section 3, the proposed system is presented. The architecture of the proposed system is explained with the block diagram and with the help of an Use Case diagram, the system design is explained in detail. Conclusion is provided in the fourth section of the paper.

2 Related Work

A Secure Data Deduplication using convergent encryption key system. There are two main categories of data de-duplication namely first level and block level de-duplication [3], [5]. The block size is either fixed or variable in the block level de-duplication. Now-a-days people store huge amount of data on laptops or personal computers. Convergent encryption is the mechanism used by us. A technology which helps to reduce the cost is data de-duplication. To reduce the risk of data usage, cross level de-duplication is enabled text data.

Key management and convergent system are the techniques used in this system [6], [7], [8]. There are two problems in this system. Firstly, the enormous number of keys is generated when the number of user gets increased.

A Secure Deduplication with Efficient and Reliable Convergent Key Management. To remove duplicate copies of data, data de-duplication is being used and it is widely applied to database storage to reduce not only storage space but also to upload bandwidth [9]. To accomplish secure data de-duplication in cloud storage is a promising challenge. For secure data de-duplication, convergent encryption has been extensively required. Key management and convergent

- *Srinivas L N B is currently working as Assistant Professor in the Department of Information Technology, School of Computing, SRM Institute of Science and Technology, Tamilnadu, India. E-mail: srinivas.l@ktr.srmuniv.ac.in*
- *Rajasekar P is currently working as Assistant Professor in the Department of Information Technology, School of Computing, SRM Institute of Science and Technology, Tamilnadu, India. E-mail: rajasekar.p@ktr.srmuniv.ac.in*

encryption are the techniques used. It is user friendly and more secure [10].

The limitations of the present techniques are discussed here. Data deduplication does not work with traditional encryption techniques. While using data deduplication technique it should not reduce fault tolerance mechanism [8], [11]. This system only provides performance benefits when there are less update operations than read operations. Increasing complexity in deduplication process is main limitation of the model. The systems still have limitations including linear or quadratic communication and computational cost. It does not provide support of public verifiability and malicious database server. In order to function more efficiently, more capacity will be needed by different deduplication processes [12]. The best example for this instance is post processing deduplication, which requires more space to retain new data, before it can be deduplicated to limit the amount of actual space.

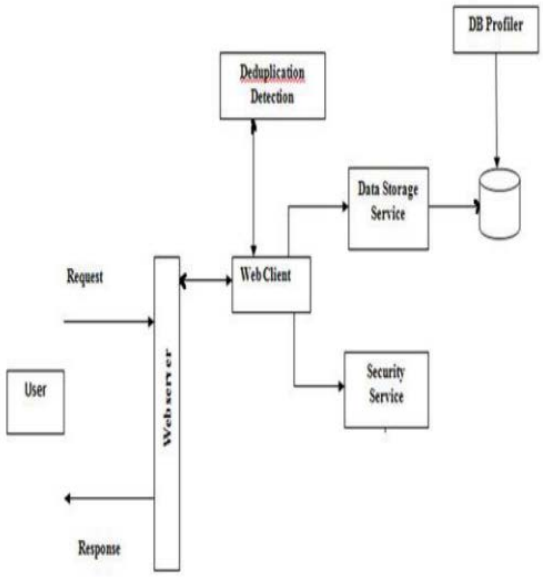

Fig. 1. Architecture of Data deduplication process

3 System Architecture and Design

Database storage is getting popular in the recent years. In file level de-duplication, duplicate copy of same file is eliminated and in block level, duplicate block of data is eliminated. The main issue with data de-duplication is security [13]. Database provides a better way of storing the data with nominal cost. De-duplication has many advantages but it has some security issues as well [9], [14], [15]. The proof of data by the data owner is provided for achieving data de-duplication in our proposed system. It is used at the time when the file gets uploaded. The user can upload their file in the database storage. The file gets stored in the database in the format of encryption. The security is maintained when the user when the authorized user downloads the specific file and hence the files get decrypted.

The architecture of the proposed system is presented in Fig.1. Web server gets the request for uploading the file from the user when user wants to upload the file to the database storage. For that purpose only the users have got approval from the web server can only upload their files. The file gets split into the block when it gets uploaded i.e. 4 kb is the default size. Block occurs depending upon the size of file.de-duplication detection occurs after that. Security service and database storage service are the two services which web client has. Security service and uploaded files is contained in the data storage server and provides security to those files.

In Fig.2., the data deduplication process is presented as an Use Case Diagram. Use cases and users interaction is being captured by the use case diagram. Functional requirement of the system is described by it. Overview of all is provided or the usage requirement parts for the organization or the system in the form of business model or an essential model. The scope of a development project is communicated. One or more use case diagram is comprised by the use case model. It supports documentation such as actor definition and use case specification.

The time taken to upload and download file of variable size is measured in order to improve system performance. When the theoretical design is turned into a working system is the stage of the implementation of the project. It is very critical stage. The experiment can be performed on the simulation of the model proposed by us. In this proposed work, the size of the file varies. Three different operations are tested viz. upload, test and delete operations. The system is more secure by the key feature and OTP feature. The user needs to request the key from the admin in order to view the file. The admin needs to approve the key request of the user.

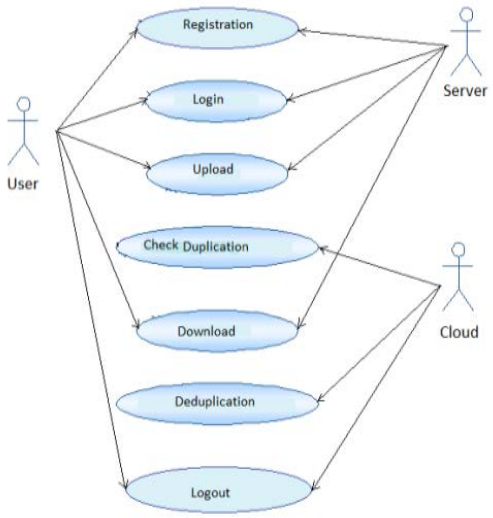

Fig. 2. Use case diagram of Data deduplication process

The system faced scalability problems takes a longer time when the number of files gets increased, when a single de-duplicator is used. The performance of the system cannot be maintained by the single user. The result shows that it helps in the reduction of the processing time, when the number of de-duplicators is increased. The experiment is being performed to delete the files. The processing time can be decreased, by addition of more de-duplicators, but the result of upload and update cases is slightly different from delete files.

4 Conclusion

Nowadays majority of users are using database storage to store their data. Increase in the amount of data in database is one of the major concerns. For reducing the space and utilizing it efficiently, data de-duplication is used. An encrypted data duplication mechanism is proposed in this work. The proposed system includes proof of data owner and helps for the implementation of better security issues in database. As the demand is increasing and data storage in the database, data de-duplication is one of the techniques which are useful for the improvement of efficiency of storage. The design is expected to give high performance and the scalability problem can also be handled effectively. The

proposed mechanisms help to improve the privacy preservation.

References

[1] Jin Li, Yan Kit Li, Xiaofeng Chen, Patrick P. C. Lee, Wenjing Lou, "A Hybrid Cloud Approach for Secure Authorized Deduplication", IEEE Transactions on Parrallel and Distributed Systems, Vo.26, Issue 5, 2015.

[2] M. Bellare, S. Keelveedhi, and T. Ristenpart, "Dupless: Serveraided encryption for deduplicated storage", USENIX Security Symposium, 2013.

[3] J. Stanek, A. Sorniotti, E. Androulaki, and L. Kencl, "A secure data deduplication scheme for database storage", Technical Report, 2013.

[4] S. Quinlan and S. Dorward. Venti, "A new approach to archival storage", Proc. USENIX FAST, Jan 2002.

[5] Edward J. Coynek, Hal L. Feinsteink and Charles E. Youmank Ravi S. Sandhu, "Role-Based Access Control Models", IEEE Computer, vol. 29, pp. 38-47, Feb 1996.

[6] A. Adya, W. J. Bolosky, D. Simon, and M. Theimer. J. R. Douceur, "Reclaiming space from duplicate files in a serverless distributed", ICDCS, pp. 617-624, 2002.

[7] P. Anderson and L. Zhang, "Fast and secure laptop backups withencrypted de-duplication", Proc. of USENIX LISA, 2010.

[8] M. Bellare, C. Namprempre, and G. Neven, "Security proofs foridentity-based identification and signature schemes", J. Cryptology,22(1):1–61, 2009.

[9] S. Bugiel, S. Nurnberger, A. Sadeghi, and T. Schneider Twinclouds, "An architecture for secure cloud computing", Workshop on Cryptography and Security in Clouds (WCSC 2011), 2011.

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

[11] 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 Systems, 2013.

[12] C. Ng and P. Lee. Revdedup, "A reverse deduplication storage system optimized for reads to latest backups", Proc. of APSYS, Apr 2013.

[13] W. K. Ng, Y. Wen, and H. Zhu, "Private data deduplication protocols in cloud storage", S. Ossowski and P.

Lecca, editors, Proceedings of the 27th Annual ACM Symposium on Applied Computing, ACM, 2012.

[14] R. D. Pietro and A. Sorniotti, "Boosting efficiency and security in proof of ownership for deduplication", H. Y. Youm and Y. Won, editors, ACM Symposium on Information, Computer and Communications Security, ACM, 2012.

[15] A. Rahumed, H. C. H. Chen, Y. Tang, P. P. C. Lee, and J. C. S.Lui, "A secure cloud backup system with assured deletion and version control", 3rd International Workshop on Security in Cloud, 2012.