IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

ISO 3297:2007 Certified ⁜ Impact Factor 8.066 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 10, Issue 7, July 2023

DOI: 10.17148/IARJSET.2023.10795

Authorized Redundant Check Support in a Hybrid Cloud Environment

Sahana H R[1], Seema Nagraj[2]

Student, Department of Master of Computer Applications, Bangalore Institute of Technology, Bengaluru, Karnataka, India[1]

Assistant Professor, Department of Master of Computer Applications, Bangalore Institute of Technology, Bengaluru, Karnataka, India[2]

Abstract: According to research findings, data deduplication is a highly effective data compression method used in cloud storage to eliminate duplicate instances of recurring data, thereby conserving storage space and bandwidth. To ensure the confidentiality of sensitive data while enabling deduplication, a convergent encryption approach has been developed for encrypting data before outsourcing. Our study represents the first explicit effort to address the concept of authorized data deduplication, aiming to enhance data security. Unlike traditional deduplication systems that solely focus on the data itself, our analysis takes into account the varying privileges of users during duplicate checks. Furthermore, we propose several innovative deduplication architectures that facilitate authorized duplicate checks in a hybrid cloud setup. Based on our research and security analysis, this approach aligns with the most secure practices defined in the proposed security model.To validate our proposal, we implement a prototype of the authorized duplicate check mechanism and conduct testbed tests. The results demonstrate that our method introduces minimal overhead compared to standard operations.

Keywords: deduplication, convergent encryption, privileges, data compression, cloud storage.

I. INTRODUCTION

Cloud computing's popularity is skyrocketing. It facilitates the virtualization process and facilitates the distribution of client/server and other types. Its assistance is well above our wildest dreams. Web services and software as a service are only two examples of the many services that will aid in the advancement of distributed, grid, utility, and autonomous computing. The cloud's flexible architecture, cost-effective operational model, and user-friendly delivery platform enable widespread usage of the software. In order to ensure the proper functioning of cloud computing, adjustments to Service-Oriented Architecture (SOA) and virtualization are necessary. Amazon, the leading online bookseller, is a frequently chosen option for outsourcing. The rapid delivery of diverse services in any context using virtualization technology gives rise to privacy and security apprehensions. While there are no known security vulnerabilities in the cloud itself, several issues exist in its construction, such as an excessive number of virtual machines, insecure mobile data, and access control problems. Consequently, many users remain cautious about adopting cloud technology due to potential security concerns. Cloud computing enables the virtualization of hardware and software, concealing their underlying structure while making the data inaccessible to users. Modern cloud storage services offer large and affordable storage space, with authorized users sharing the ever-expanding data. This growth is managed through storage visibility and access controls, posing a challenge for cloud storage providers. To address this, deduplication has become a significant focus in the technology industry. This method scales cloud data management by eliminating duplicates, optimizing byte utilization, storage space, and network traffic. Deduplication reduces data duplication by storing a single duplicate and "pointing" to others, especially in distributed computing platforms. The debate between file-level and block-level deduplication has valid arguments, but ultimately, redundant data is deleted during the process. Data compression works at the block level, removing duplicate data blocks within otherwise unique files. Despite the advantages of data deduplication, privacy and security concerns arise as access to sensitive user data is simplified, both within and outside the organization. To address security, data deduplication and conventional encryption work together to safeguard information, with individuals encrypting their own data using private keys.

II. LITERATURE SURVEY

Distributed scaling of symbolic data
Authors:Rosie Verde and Tony Balzanella

Various tasks can be accomplished using SDA data, which possesses a distributional value. Quantitative variable distributions like histograms, densities, and quantile functions describe the "multi-valued data" element values. "Multi-valued data" is a good term to use. Multi-valued data is also called "data with multiple dimensions." SDA works well with large datasets. Principal component analysis makes variables with many numbers of values easier to understand. A

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

ISO 3297:2007 Certified ⁜ Impact Factor 8.066 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 10, Issue 7, July 2023

DOI: 10.17148/IARJSET.2023.10795

PCA is a PCA. The best method. With this study, you can find new places to visit. PCA can be done on data with values from a distribution that are shown by a quantile variable. When the squared Wasserstein distance is used on distributions, it gives rise to new link measures. These measures show how different parts of a spread are connected. When applied to generated data with distribution-valued variables, principle component analysis (PCA) gives new ways of looking at the location, variability, and shape of distributions on factorial planes. Distribution charts show how the information is spread out. showing modeling.

Effective dimensionality reduction is needed to prepare large amounts of data in a streaming context.
Authors:Yin Jun
Massive datasets or real-time data streams necessitate dimension reduction for data preparation. It helps classifiers do their jobs better. This could take a long time. Two prevalent approaches to reducing the number of dimensions in data are feature extraction and feature selection. Feature extraction methods generally outperform feature selection methods. In feature extraction, features are obtained rather than chosen. Dealing with large data streams or datasets can be challenging, and the use of computers becomes essential in such scenarios. Choosing which features to use is an example of a greedy approach that might not lead to the best results. "Greedy strategies" is a theme that keeps coming up in the answers. This famous book talks about many ways to extract and choose traits. So, we were able to understand these ways better. Here, these methods are put to the test. We look at orthogonal centroid algorithm (OC)-based methods for reducing the number of dimensions from two to one. The first way to get features out is called incremental OC, or IOC. Orthogonal centroid feature selection (OCFS), which gives the best results that meet the OC requirements, is stressed in the second method. Both were made similar by the same improvement process. On the Reuters Corpus Volume-1 data set and other big text data sets, the two programs did a better job than the most advanced methods. The Reuters Corpus Volume-1 and other large text collections were used to do these comparisons and studies.

Using data mining to find bad parts of a program
Authors: Jesus S. Aguilar Ruiz, Daniel Rodriguez, and Roberto Ruiz.
Identifying software problems can be challenging at times, but data mining offers a solution by pinpointing the most susceptible components likely to encounter issues. From the PROMISE library datasets, we selected features for analysis. In the following stage, a genetic method was employed to distinguish different groups accurately. This technique proves beneficial in handling imbalanced datasets, especially when there is a higher number of high-quality samples compared to low-quality ones, facilitating effective predictions for part breakdowns.

Based on the Semantic Web and classical document modeling, software maintenance took a big step forward.
Authors:Carrington and Kaplan, D. and S.
In this article, we will examine the ongoing efforts to enhance software maintenance by describing software in terms of paradigms, along with an overview of the conducted studies in this field. The writers look into Kuhn's theories in order to describe software systems and give metadata. Next, the writers say that software system information should be collected. Some examples of metadata are the definitions of metrics, the results of tests, the links between components, Additionally, we will discuss the specifications of both functional and non-functional requirements. The writers say that changes to metadata should be written down and tracked so that coders are aware of changing needs and quality measures that could affect maintenance. It is explained how to track changes to metadata. The writers then think about how information could use these traits to its advantage. Users can use Semantic Web technology to handle relationship software systems even if they don't speak English. Software tools are easier to keep up and running. In this case, technologies for handling software in an autonomous way could be useful.

III. MODELING AND ANALYSIS

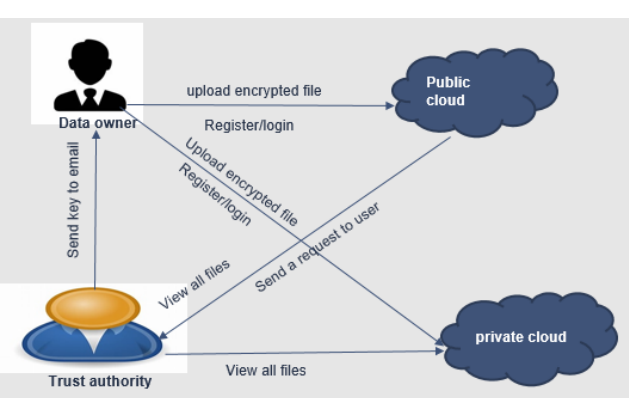

Figure 1: Flow chart of the Hybrid Cloud Environment.

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

ISO 3297:2007 Certified ⁜ Impact Factor 8.066 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 10, Issue 7, July 2023

DOI: 10.17148/IARJSET.2023.10795

The cloud computer stores data, looks at data, checks requests, and gives out data. Consider the number of files you possess and the frequency with which people require them. Here's where you can find it: - The number of files is kept in the cloud. With fuzzy logic's key generation for encryption and decryption, which includes a file private key and a trapdoor key, users can send data to the cloud without worrying about security. Fuzzy logic can be used to secure and decrypt this information. Check the facts. Inquire about: Locate secret papers. The user waits for an answer from the staff at the key generation center after sending the request. The user can download the file without encryption if they have the secret key. The TPA can take care of the accuracy of data, the authorization of Data Users, and the tracking of requests. Once the KGC has handled the User's Request to upload File and Data, the Private Key user will send an email to inform other users.

IV. RESULTS AND DISCUSSION

EXISTING SYSTEM

- In a compression system, permissions are first given to people that the system administrator has checked out. We call this "initializing the system."
- Even though traditional encryption gives security, it can't be used with data compression.
- Since different cypher messages would be made by many users' copies of the same data, deduplication won't work.

DISADVANTAGES OF EXISTING SYSTEM

- Accounting staff have to look at file names to see if there are duplicates, which takes more time and costs more money.

V. PROPOSED SYSTEM

- Academics have closely studied this architecture for its potential usefulness in the future, particularly in regard to checking for similarity.
- The private cloud ensures the capability to detect duplicates, while also managing distinct rights for data owners and users.
- Data owners exclusively utilize public cloud services from a third-party to store their data.

DISADVANTAGES OF PROPOSED SYSTEM

- Made things safer. The main benefit of multi-factor authentication is that it keeps your files and data safe.

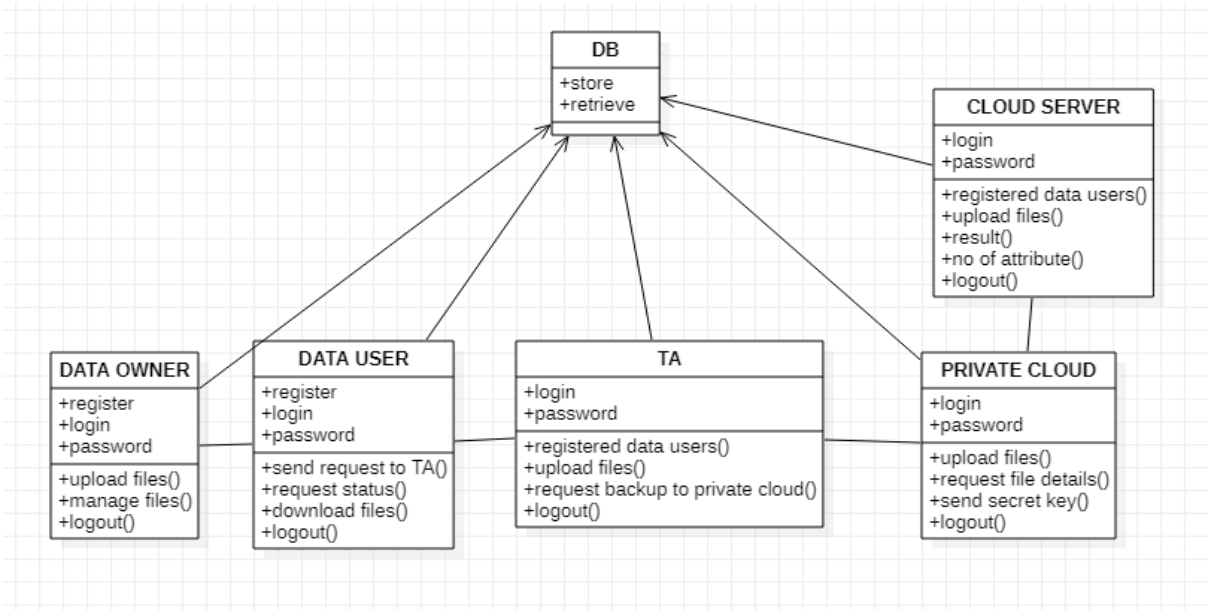

Figure 2: Proposed System Class Diagram

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

ISO 3297:2007 Certified ⁜ Impact Factor 8.066 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 10, Issue 7, July 2023

DOI: 10.17148/IARJSET.2023.10795

VI. CONCLUSION

The concept of permitted data deduplication was offered as a method for ensuring data security according to the implementation and checking for duplicates. This would be accomplished by taking into consideration the different permissions that are provided to different users. We also showed hybrid cloud deduplication systems that can validate duplicate checks. In these recently developed designs for deduplication, the private cloud server creates file duplicate check tokens utilizing private keys. We also showed hybrid cloud deduplication systems that can validate duplicate checks. These sorts of assaults might come from either within the organization or from outside the organization

REFERENCES

1. The OpenSSL Project can be found at http://www.openssl.org/.
2. L. Zhang and P. Anderson. The use of encrypted deduplication technology enables quick and secure laptop backups. Presented in The USENIX LISA Conference Proceedings for 2010.
3. M. Bellare, T. Ristenpart, and S. Keelveedhi. The "Dupless" technique, a server-assisted encryption approach for deduplicated storage, was presented at the 2013 USENIX Security Symposium.
4. M. Bellare, T. Ristenpart, and S. Keelveedhi. Employable methods of message-locked encryption and secure data duplication. Pages 296-312 of EUROCRYPT's 2013 version.
5. According to the reference in footnote number 5, M. Bellare, C. Namprempre, and G. Neven wrote an essay on security justifications for methods using signers' identities for identification. Published in the 2009 issue of the Journal of Cryptology, volume 22, number 1, pages 1-61.
6. M. Bellare and A. Palacio. Presenting proofs of security against impersonation under active and concurrent attacks for the Gq and Schnorr identification systems. Pages 162-177 of CRYPTO, 2002.
7. T. Schneider, A. Sadeghi, S. Bugiel, and S. Nurnberger. "Twin clouds" system enables computations in the cloud. Presented at the Workshop on Cryptography and Security in Clouds (WCSC 2011) in 2011.
8. J. R. Douceur, A. Adya, W. J. Bolosky, D. Simon, and M. Theimer. Reclaiming space in a serverless distributed file system previously occupied by duplicate files. ICDCS, 2002, pages 617-624.
9. Role-based access controls concept developed by David Ferraiolo and Robert Kuhn. Celebrated at the NIST-NCSC National Computer Security Conference in 1992, its 15th anniversary.
10. Learn more about GNU Libmicrohttpd at http://www.gnu.org/software/libmicrohttpd/.
11. A. Shulman-Peleg, B. Pinkas, D. Harnik, and S. Halevi. Ownership proof documentation stored in off-site storage facilities. Presented in the ACM Conference on Computer and Communications Security Proceedings, edited by Y. Chen, G. Danezis, and V. Shmatikov, on pages 491-500, ACM, 2011.
12. X. Chen, P. Lee, W. Lou, M. Li, J. Li, and J. Li. A reliable and effective data duplication technique using convergent key management for safeguarding sensitive data. Published in the IEEE Transactions on Parallel and Distributed Systems edition from 2013.
13. Libcurl is accessible at http://curl.haxx.se/libcurl/.
14. Peter Lee and Christopher Ng are associated with the storage system "Revdedup," which uses reverse data duplication and is optimized for reading from the most recent backups. The article was released in the APSYS Proceedings in April 2013.
15. Y. Wen, H. Zhu, and W. K. Ng. Private data deduplication methods for cloud storage. Presented in the proceedings of the 27th Annual ACM Symposium on Applied Computing on pages 441-446, edited by S. Ossowski and P. Lecca, ACM, 2012.
16. R. D. Pietro and A. Sorniotti. Increasing the safety and effectiveness of the proof of ownership process for deduplication. Presented at the ACM Symposium on Information, Computer, and Communications Security, pages 81 and 82, edited by H. Y. Youm and Y. Won, ACM, 2012.
17. S. Dorward and S. Quinlan. Venti, a revolutionary method for archival storage. Presented in the USENIX FAST Proceedings in January 2002.
18. J. C. S. Lui, A. Rahumed, H. C. H. Chen, and Y. Tang, with contributions from P. P. C. Lee. A safe backup program housed in the cloud that ensures previous data iterations won't be maintained. Presented at the third annual international workshop on security in cloud computing in 2011.
19. Models for access restriction depending on the posts being filled, published in IEEE Computer, February 1996, pp 38-47, authored by R. S. Sandhu, E. J. Coyne, H. L. Feinstein, and C. E. Youman.
20. The paper authored by J. Stanek, A. Sorniotti, L. Kencl, and E. Androulaki presents a reliable and safe data deduplication technique specifically designed for use with cloud storage. The Technical Report is dated 2013.

