International Scientific Journal of Engineering and Management
Volume: 02 Issue: 04 | April – 2023 DOI: 10.55041/ISJEM00418
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata
ISSN: 2583-6129
www.isjem.com

DE-DUPLICATION ON VOTING SYSTEM

Dr.G.ROSLIN NESAKUMARI

ERAVENI YASWANTH

BONU SEKHAR SANDEEP

BOPPENAPELLI VAMSHI

BOPPENAPELLI SAGAR

ABSTRACT

Data de duplication is one of important data compression techniques for eliminating duplicate copies of repeating data, and has been widely used in cloud storage to reduce the amount of storage space and save bandwidth. To protect the confidentiality of sensitive data while supporting de duplication, the convergent encryption technique has been proposed to encrypt the data before outsourcing. To better protect data security, this project makes the first attempt to formally address the problem of authorized data de duplication. Different from traditional de duplication systems, the differential privileges of users are further considered induplicate check besides the data itself. We also present several new de duplication constructions supporting authorized duplicate check in hybrid cloud architecture. Security analysis demonstrates that our scheme is secure in terms of the definitions specified in the proposed security model. As a proof of concept, the proposed work implements a prototype of our proposed authorized duplicate check scheme and conduct test bed experiments using our prototype. The proposed work shows that our proposed authorized duplicate check scheme incurs minimal overhead compared to normal operations.

ISSN: 2583-6129
www.isjem.com

CHAPTER 1

INTRODUCTION

1.1 OVERVIEW

Cloud computing provides seemingly unlimited "virtualized" resources to users as services across the whole Internet, while hiding platform and implementation details. Today's cloud service providers offer both highly available storage and massively parallel computing resources at relatively low costs. As cloud computing becomes prevalent, an increasing amount of data is being stored in the cloud and shared by users with specified *privileges*, which define the access rights of the stored data. One critical challenge of cloud storage services is the management of the ever-increasing volume of data. To make data management scalable in cloud computing, de duplication has been a well-known technique and has attracted more and more attention recently.

Data de duplication is a specialized data compression technique for eliminating duplicate copies of repeating data in storage. The technique is used to improve storage utilization and can also be applied to network data transfers to reduce the number of bytes that must be sent. Instead of keeping multiple data copies with the same content, deduplication eliminates redundant data by keeping only one physical copy and referring other redundant data to that copy. De duplication can take place at either the file level or the block level. For filelevel de duplication, it eliminates duplicate copies of the same file. De duplication can also take place at the block level, which eliminates duplicate blocks of data that occur in non-identical files. Although data deduplication brings a lot of benefits, security and privacy concerns arise as users' sensitive data are susceptible to both insider and outsider attacks. Traditional encryption, while providing data confidentiality, is incompatible with data deduplication.

Specifically, traditional encryption requires different users to encrypt their data with their own keys. Thus, identical data copies of different users will lead to different cipher texts, making de duplication impossible. Convergent encryption has been proposed to enforce data confidentiality while making de duplication feasible.

It encrypts/decrypts a data copy with a convergent key, which is obtained by computing the cryptographic hash value of the content of the data copy. After key generation and data encryption, users retain the keys and send the cipher text to the cloud. Since the encryption operation is Deterministic and is derived

ISSN: 2583-6129
www.isjem.com

from the data content, identical data copies will generate the same convergent key and hence the same cipher text. To prevent unauthorized access, a secure proof of ownership protocol is also needed to provide the proof that the user indeed owns the same file when a duplicate is found. After the proof, subsequent users with the same file will be provided a pointer from the server without needing to upload the same file. A user can download the encrypted file with the pointer from the server, which can only be decrypted by the corresponding data owners with their convergent keys.

Thus, convergent encryption allows the cloud to perform de duplication on the cipher texts and the proof of ownership prevents the unauthorized user to access the file. However, previous de duplication systems cannot support differential authorization duplicate check, which is important in many applications. In such an authorized de duplication system, each user is issued a set of privileges during system initialization. Each file uploaded to the cloud is also bounded by a set of privileges to specify which kind of users is allowed to perform the duplicate check and access the files. Before submitting his duplicate check request for some file, the user needs to take this file and his own privileges as inputs. The user is able to find a duplicate for this file if and only if there is a copy of this file and a matched privilege stored in cloud. For example, in a company, many different privileges will be assigned to employees.

In order to save cost and efficiently management, the data will be moved to the storage server provider (SSP) in the public cloud with specified privileges and the de duplication technique will be applied to store only one copy of the same file. Because of privacy consideration, some files will be encrypted and allowed the duplicate check by employees with specified privileges to realize the access control. Traditional de duplication systems based on convergent encryption, although providing confidentiality to some extent; do not support the duplicate check with differential privileges. In other words, no differential privileges have been considered in the de duplication based on convergent encryption technique. It seems to be contradicted if we want to realize both de duplication and differential authorization duplicate check at the same time.

1.2 ORGANIZATION OF THE THESIS

Chapter 1 deals with an introduction to the project where the existing system is been discussed. It also gives an overview of how de duplication technique has been implemented with the high security.

ISSN: 2583-6129
www.isjem.com

In Chapter 2, a detailed description of the literature survey of the papers which are referred during the course of the project was summarized.

Chapter 3 gives a brief explanation on the aim and scope of the project. Here proposed system has been compared with the existing system. The issues in the existing system and the advantages of the proposed system are also discussed.

Chapter 4 deals with the methods and algorithms used. The hardware and software requirements are provided along with the system design, architecture and flow of overall project.

Chapter 5 deals with the system implementation of the project.

In chapter 6, the Results and Discussion along with the screenshots of each module has been depicted.

Chapter 7 deals with the summary and conclusion of the project. It also includes the future scope of the project.

CHAPTER 2
LITERATURE SURVEY

Literature survey is the most important step in software development process. Before developing the tool it is necessary to determine the time factor, economy and company strength. Once these things are satisfied, then the next step is to determine which operating system and language can be used for developing the tool. Once the programmers start building the tool the programmers need lot of external support. This support can be obtained from senior programmers, from book or from websites. Before building the system the above consideration are taken into account for developing the proposed system.

The major part of the project development sector considers and fully survey all the required needs for developing the project. For every project Literature survey is the most important sector in software development process. Before developing the tools and the associated designing it is necessary to determine

ISSN: 2583-6129
www.isjem.com

and survey the time factor, resource requirement, man power, economy, and company strength. Once these things are satisfied and fully surveyed, then the next step is to determine about the software specifications in the respective system such as what type of operating system the project would require, and what are all the necessary software are needed to proceed with the next step such as developing the tools, and the associated operations

[1] In this paper, they proposed an architecture that provides secure deduplication storage resisting brute force attacks, and realize it in a system called dupLESS . It enables clients encrypted data with an existing service. The encryption for deduplicated storage can achieve performance and space saving close to that of using the storage service with plaintext data.

[12] There is a mechanism to reclaim space from incidental duplication to make it available for controlled file replication. This mechanism convergent encryption, which enable duplicate files to be coalesced into the space file, even if the files are encrypted with different users keys.

[15] It is a baseline approach in which each user holds an independent master key for encrypting the convergent keys and outsourcing them to the cloud. However, such a baseline key management scheme generates an enormous number of keys with the increasing number of users and requires users to dedicatedly protect the master key.

[17] In this project , they construct a private de duplication protocol based on the standard cryptographic assumptions is then presented and analyzed. They show that the private data de duplication protocol is probably secure assuming that the underlying hash function is collision-resilient, the discrete logarithm is hard and the erasure coding algorithm can erasure up to many fractions of the bits.

[21] In this paper, they design an encryption scheme that guarantees semantic security for unpopular data and provides weaker security and better storage and bandwidth benefits for popular data. This way, data de duplication can be effective for popular data, whilst semantically secure encryption protects unpopular content. We show that our scheme is secure under the Symmetric External Decisional Diffie-Hellman Assumption.

ISSN: 2583-6129
www.isjem.com

CHAPTER 3
AIM AND SCOPE OF THE PROJECT

3.1 SCOPE OF THE PROJECT

Data de duplication techniques are widely employed to backup data and minimize network and storage overhead by detecting and eliminating and redundancy among data.

3.2 OBJECTIVE

The main goal is to enable de duplication and distributed storage of the data across multiple storage servers.

3.3 PROBLEM DEFINITION

The existing system only performs the de duplication either on block level or file level. It does not provide very high security needed for the message to be transmitted. Due to this the third party or hacker may find the data that is being transmitted between the users.

3.4 EXISTING SYSTEM

Data de duplication systems, the private cloud are involved as a proxy to allow data owner/users to securely perform duplicate check with differential privileges. Such architecture is practical and has attracted much attention from researchers. The data owners only outsource their data storage by utilizing public cloud while the data operation is managed in private cloud.

Data de duplication is a specialized data compression technique for eliminating duplicate copies of repeating data in storage. The technique is used to improve storage utilization and can also be applied to network data transfers to reduce the number of bytes that must be sent. Instead of keeping multiple data copies with the same content, de duplication eliminates redundant data by keeping only one physical copy and referring other redundant data to that copy.

De duplication can take place at either the file level or the block level. For file level de duplication, it eliminates duplicate copies of the same file. De duplication can also take place at the block level, which eliminates duplicate blocks of data that occur in non-identical files. Identical data copies of different users will lead to different cipher texts, making de duplication impossible.

ISSN: 2583-6129
www.isjem.com

3.4.1 *Disadvantages*

- Traditional encryption, while providing data confidentiality, is incompatible with data de duplication.
- Identical data copies of different users will lead to different cipher texts, making de duplication impossible.

3.5 PROPOSED SYSTEM

In this proposed work, the system enhanced with security. Specifically, it present an advanced scheme to support stronger security by encrypting the file with differential privilege keys. In this way, the users without corresponding privileges cannot perform the duplicate check. Furthermore, such unauthorized users cannot decrypt the cipher text even collude with the S-CSP. Security analysis demonstrates that our system is secure in terms of the definitions specified in the proposed security model.

Convergent encryption has been proposed to enforce data confidentiality while making de duplication feasible. It encrypts/decrypts a data copy with a convergent key, which is obtained by computing the cryptographic hash value of the content of the data copy. After key generation and data encryption, users retain the keys and send the cipher text to the cloud. Since the encryption operation is deterministic and is derived from the data content, identical data copies will generate the same convergent key and hence the same cipher text. To prevent unauthorized access, a secure proof of ownership protocol is also needed to provide the proof that the user indeed owns the same file when a duplicate is found.

3.5.1 *Advantages*

- The user is only allowed to perform the duplicate check for files marked with the corresponding privileges.
- We present an advanced scheme to support stronger security by encrypting the file with differential privilege keys.
- Reduce the storage size of the tags for integrity check. To enhance the security of de duplication and protect the data confidentiality.

ISSN: 2583-6129
www.isjem.com

CHAPTER 4
METHODS AND ALGORTIHMS USED

4.1 HARDWARE REQUIREMENT

The hardware requirements may serve as the basis for a contract for the implementation of the system and should therefore be a complete and consistent specification of the whole system. They are used by software engineers as the starting point for the system design. It shows what the system does and it also shows how it should be implemented. It specifies the speed of the system that should be used in this project. The hardware requirement for this project is mentioned below.

System : Pentium IV 2.4 GHz

Hard Disk : 40 GB
Floppy Drive : 44 Mb
Monitor : 15 VGA Colour
Ram : 512 Mb

4.2 SOFTWARE REQUIREMENT

The software requirements document is the specification of the system. It should include both a definition and a specification of requirements. It is a set of what the system should do rather than how it should do it. The software requirements provide a basis for creating the software requirements specification. It is useful in estimating cost, planning team activities, performing tasks and tracking the team and tracking the team's progress throughout the development activity. The software requirement specifies the application software that is being used in the project. The software needed to develop this project is mentioned below.

Operating system : Windows XP/7
IDE : Eclipse
Coding Language : Java

International Scientific Journal of Engineering and Management
Volume: 02 Issue: 04 | April – 2023 DOI: 10.55041/ISJEM00418
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata
ISSN: 2583-6129
www.isjem.com

4.3 SYSTEM ARCHITECTURE

The system architecture establishes the basic structure of the system, defining the essential core design features and elements that provide the framework for the system. The systems architecture provides the architects view of the users' vision for what the system needs to be and do, and the paths along which it must be able to evolve and strives to maintain the integrity of that vision as it evolves during detailed design and implementation.

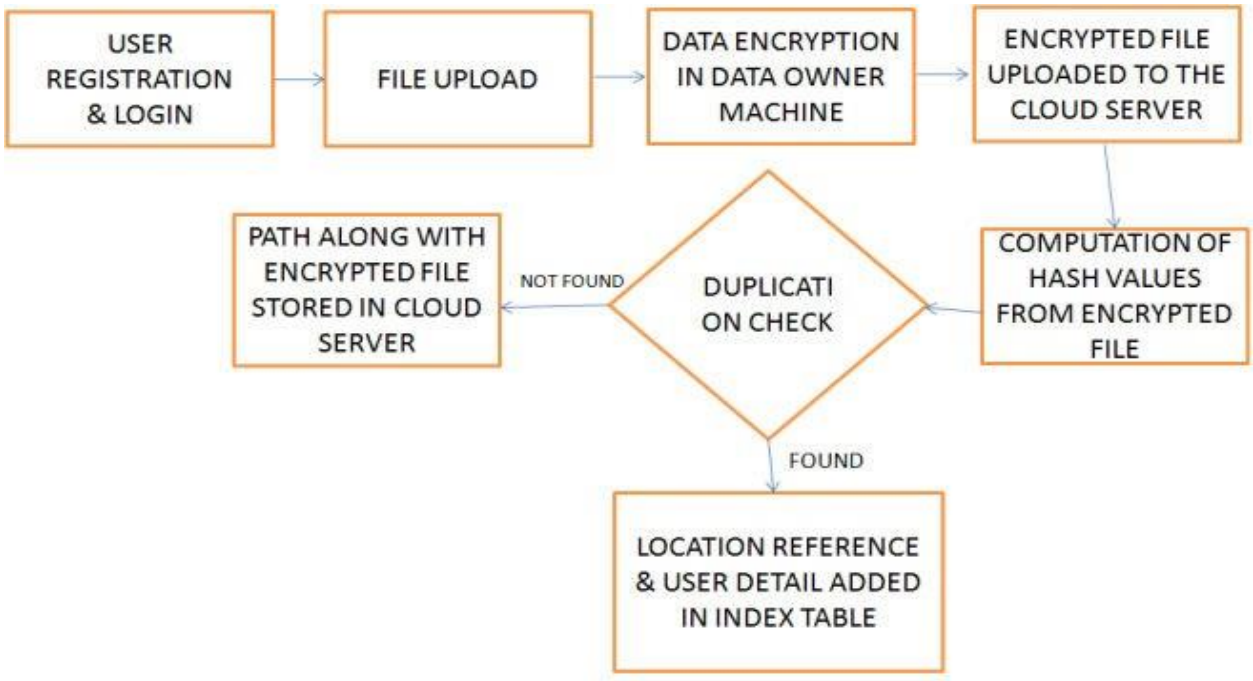


Fig. 4.1 Architecture of the system

SYSTEM IMPLEMENTATION

DATA FLOW DIAGRAM

The Data Flow diagram is a graphic tool used for expressing system requirements in a graphical form. The DFD also known as the "bubble chart" has the purpose of clarifying system requirements and identifying major transformations that to become program in system design.Thus DFD can be stated as the starting point of the design phase that functionally decomposes the requirements specifications down to the lowest level of detail.The DFD consist of series of bubbles joined by lines. The bubbles represent data transformations and the lines represent data flows in the system. A DFD describes what that data flow in rather than how they are processed. So it does not depend on hardware, software, data structure or file organization

International Scientific Journal of Engineering and Management
Volume: 02 Issue: 04 | April – 2023 DOI: 10.55041/ISJEM00418
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata
ISSN: 2583-6129
www.isjem.com

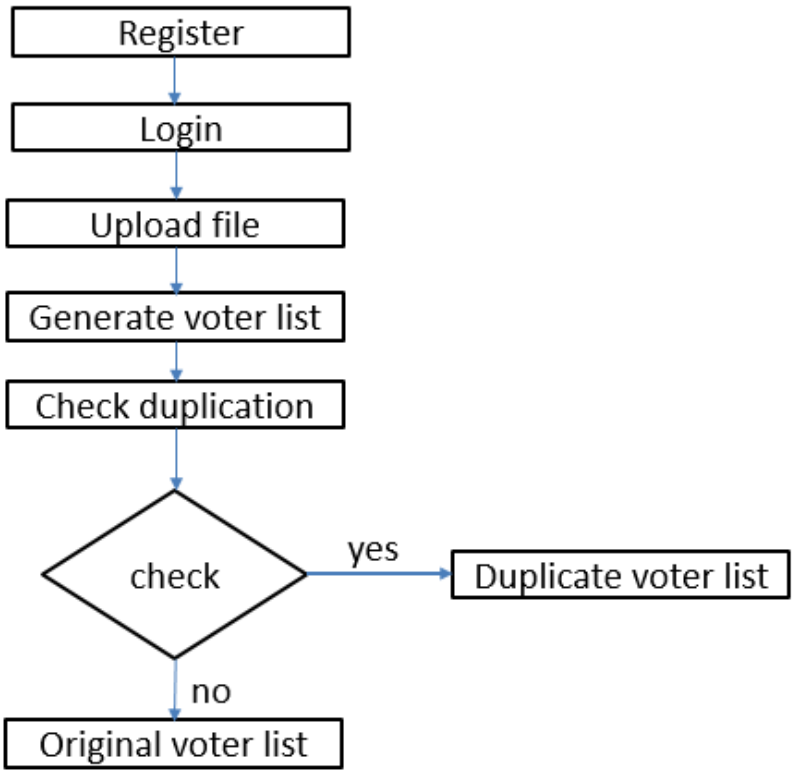


Fig. data flow diagram

CLASS DIAGRAM

International Scientific Journal of Engineering and Management
Volume: 02 Issue: 04 | April – 2023 DOI: 10.55041/ISJEM00418
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata
ISSN: 2583-6129
www.isjem.com

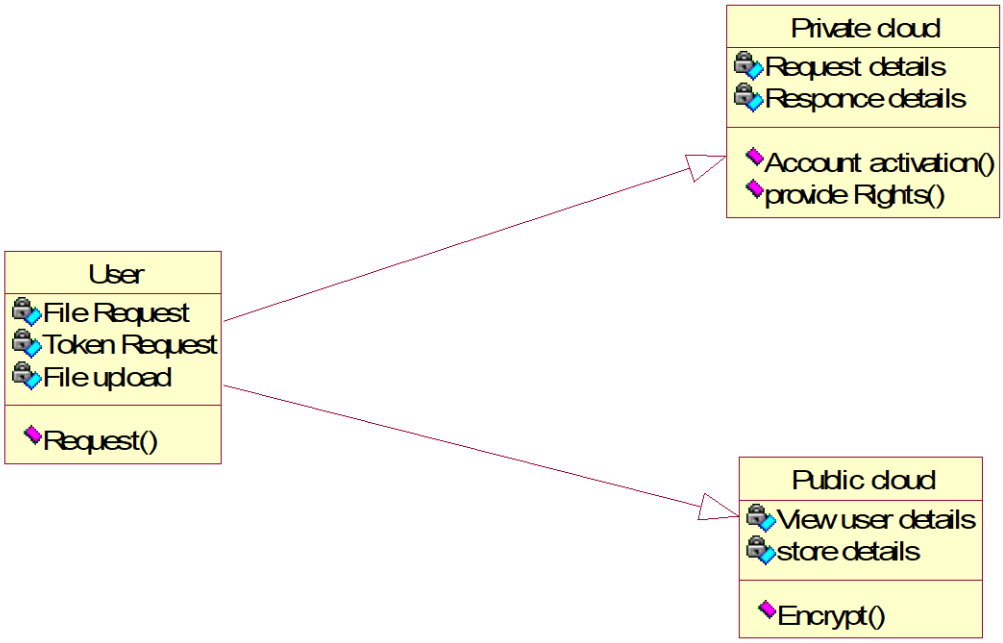


Fig. Class diagram

The class diagram is the main building block of object oriented modeling. It is used both for general conceptual modeling of the systematic of the application, and for detailed modeling translating the models into programming code. The message sending procedure involves user registering with the system and logs into the system. The authenticated secret message involves the message extraction and authentication of the message.

USE CASE DIAGRAM

A use case is a set of scenarios that describing an interaction between a user and a system. A use case diagram displays the relationship among actors and use cases. The two main components a user or another system that will interact with the system modeled. A use case is an external view of the system that represents some action the user might perform in order to complete a task.

International Scientific Journal of Engineering and Management
Volume: 02 Issue: 04 | April – 2023 DOI: 10.55041/ISJEM00418
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata
ISSN: 2583-6129
www.isjem.com

Fig. Use case diagram

ACTIVITY DIAGRAM

International Scientific Journal of Engineering and Management
Volume: 02 Issue: 04 | April – 2023 DOI: 10.55041/ISJEM00418
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata
ISSN: 2583-6129
www.isjem.com

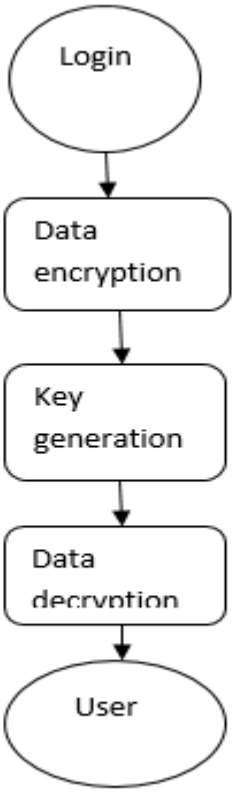


Fig. Activity diagram

ISSN: 2583-6129
www.isjem.com

SEQUENCE DIAGRAM

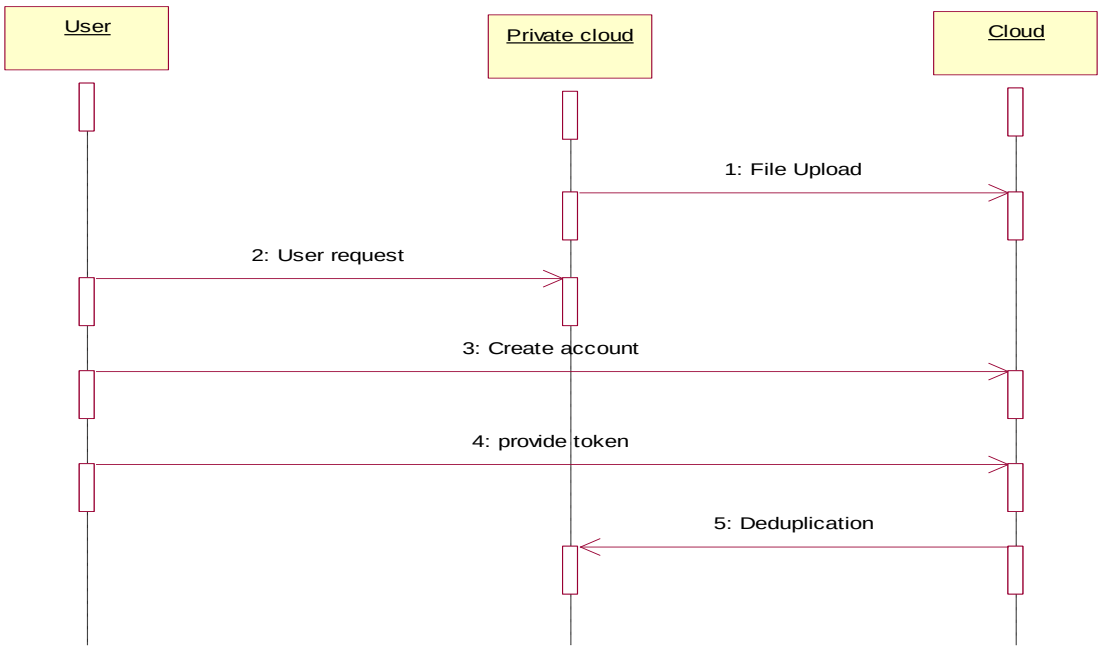


Fig. Sequence diagram

The sequence diagram is used to show the flow of the system with the time frame of each activity. The sender logs into the system and generates the file. Then it gets split using the chunk algorithm and encrypts the secret message using blowfish algorithm. The key is generated. The sender gets the private key when he registers with the system and applies inverse blowfish algorithm and split the image into different blocks. Using the private key the message is decrypted and original message is received by the user.

CONCLUSION

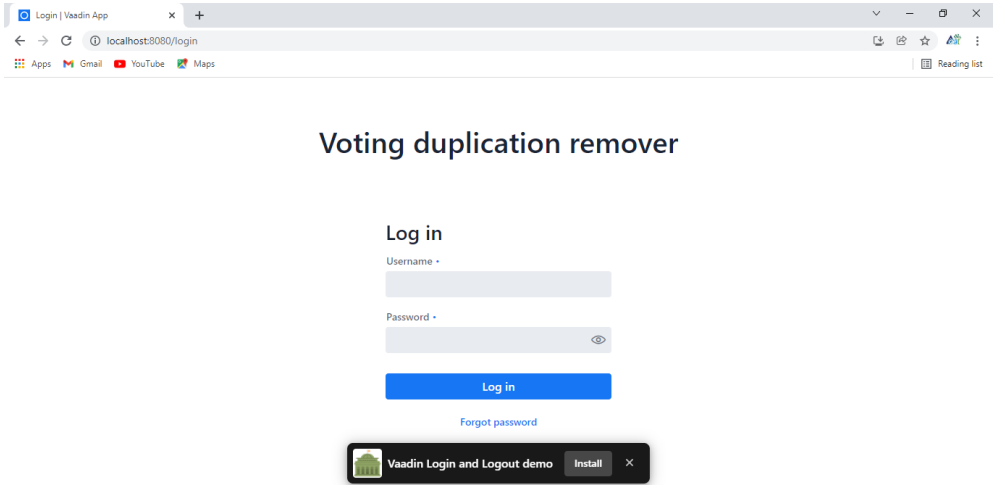
In this paper, the notion of authorized data de duplication was proposed to protect the data security by including differential privileges of users in the duplicate check. We also presented several new de duplication constructions supporting authorized duplicate check in hybrid cloud architecture, in which the duplicate-check tokens of files are generated by the private cloud server with private keys. Security analysis demonstrates that

ISSN: 2583-6129
www.isjem.com

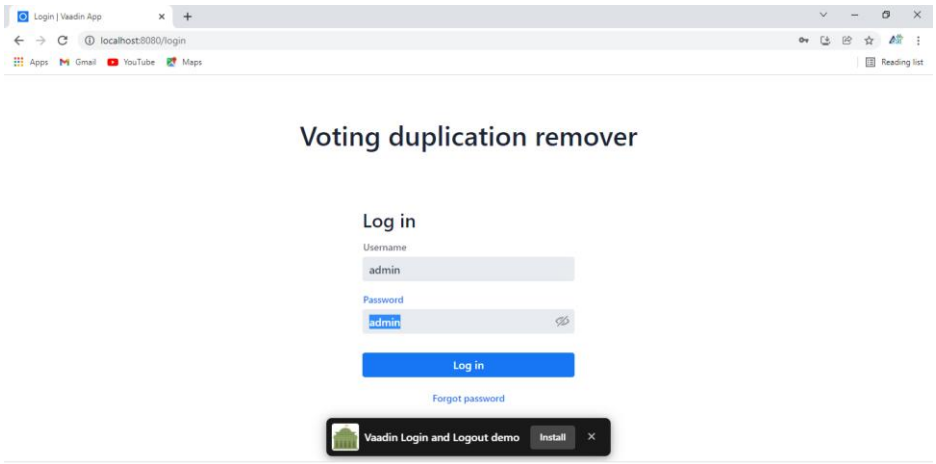
our schemes are secure in terms of insider and outsider attacks specified in the proposed security model. As a proof of concept, we implemented a prototype of our proposed authorized duplicate check scheme and conduct test bed experiments on our prototype. We showed that our authorized duplicate check scheme incurs minimal overhead compared to convergent encryption and network transfer.

IMPLEMENTATION

Home Page

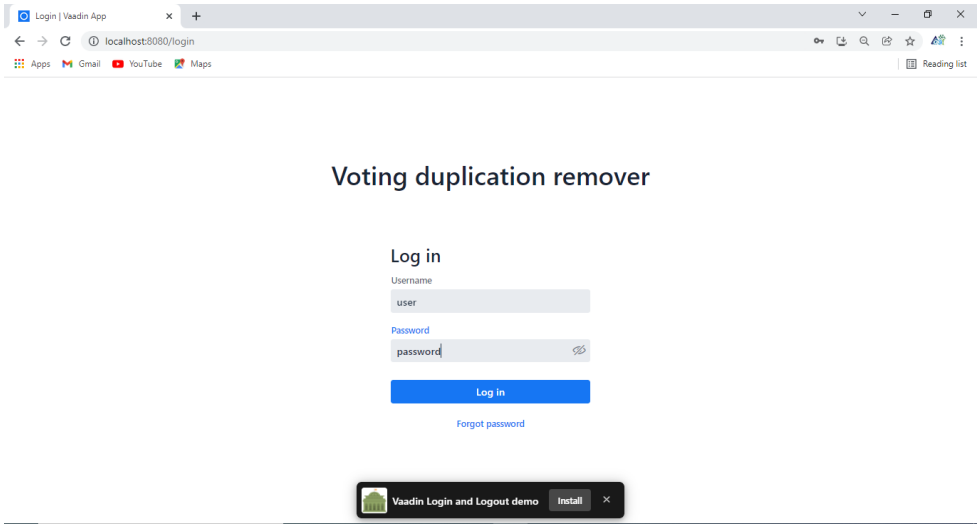


Admin Login Page

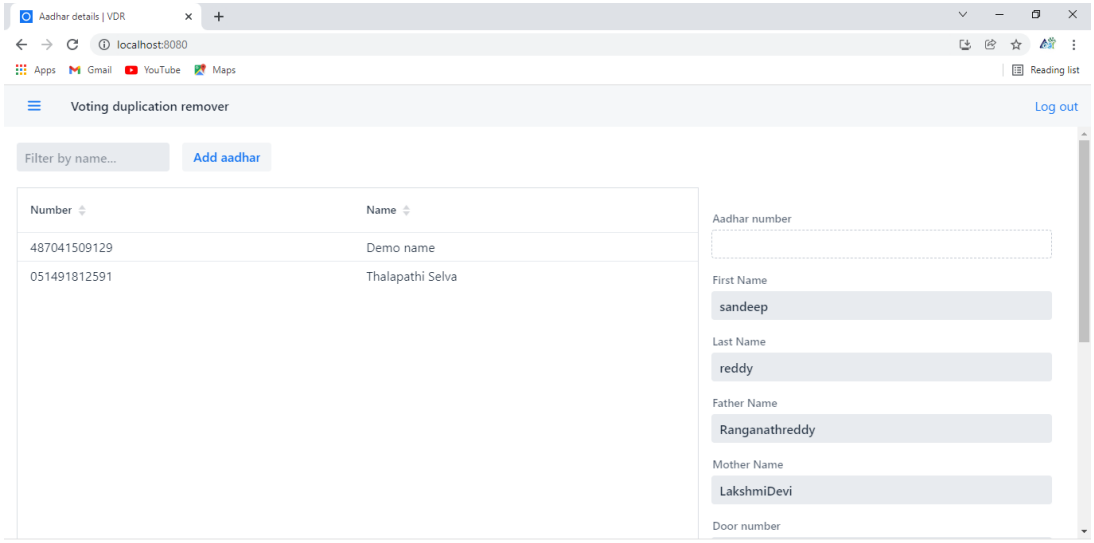


International Scientific Journal of Engineering and Management
Volume: 02 Issue: 04 | April – 2023 DOI: 10.55041/ISJEM00418
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata
ISSN: 2583-6129
www.isjem.com

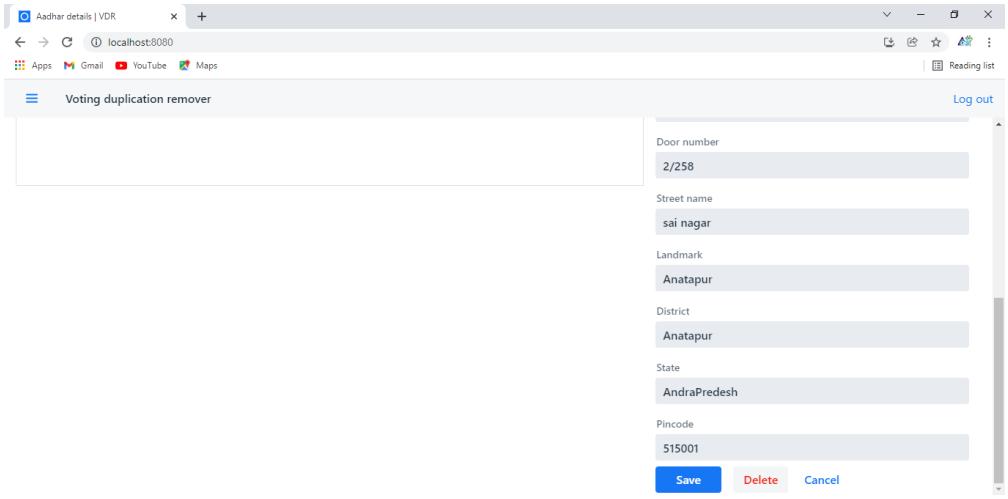
User Login Page



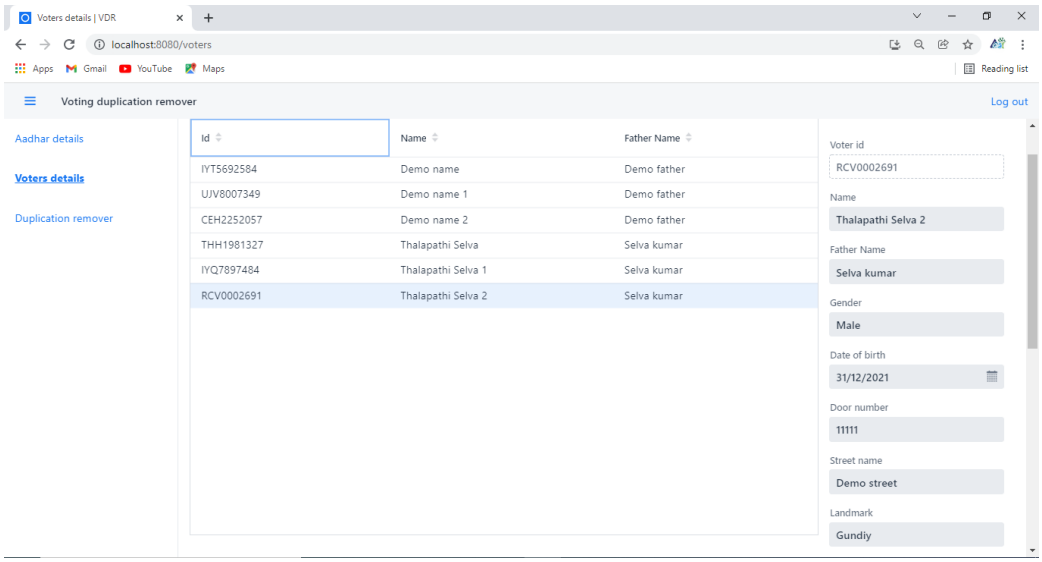
Add Aadhar Details



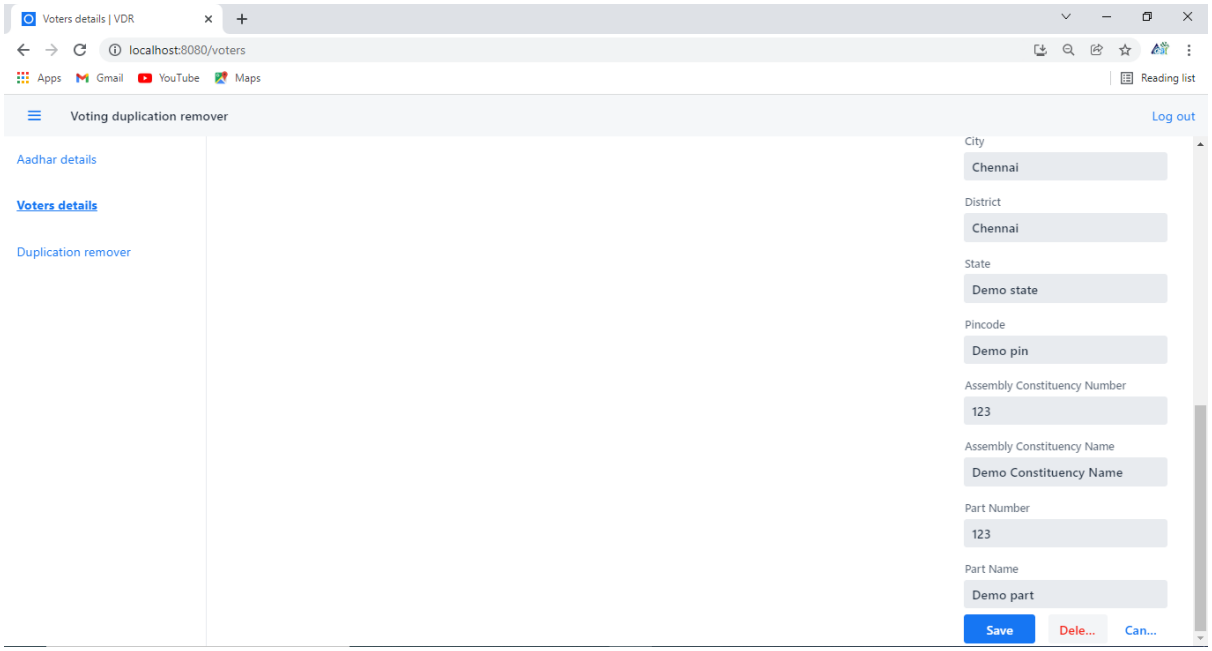
International Scientific Journal of Engineering and Management
Volume: 02 Issue: 04 | April – 2023 DOI: 10.55041/ISJEM00418
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata
ISSN: 2583-6129
www.isjem.com

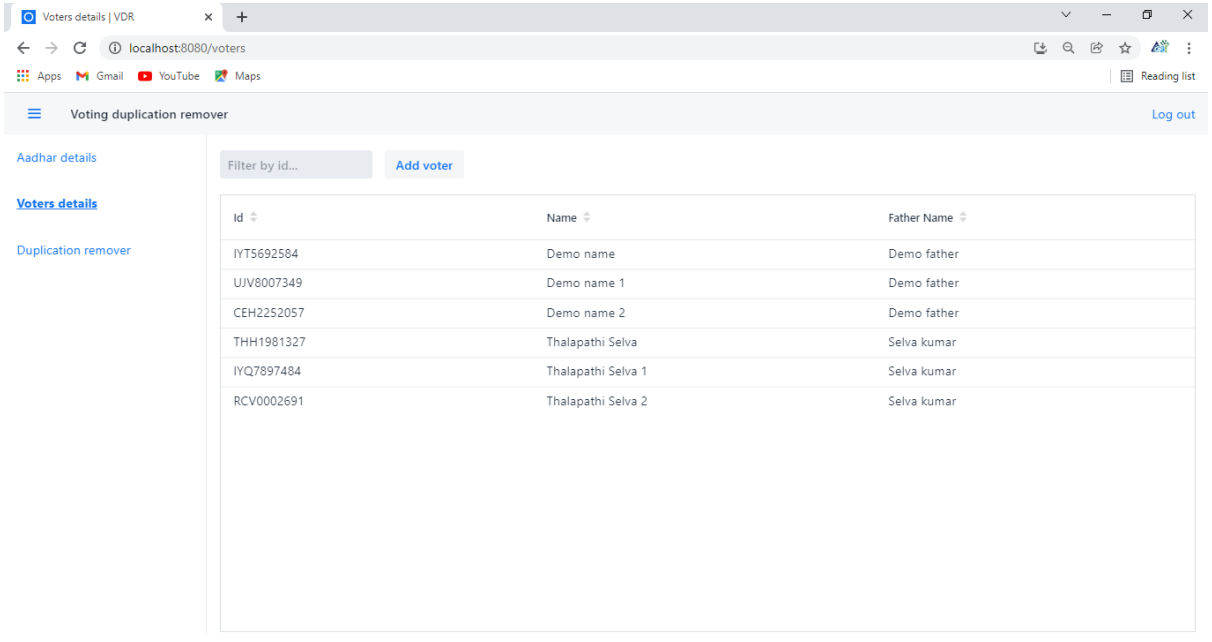


Add Voters Details



International Scientific Journal of Engineering and Management
Volume: 02 Issue: 04 | April – 2023 DOI: 10.55041/ISJEM00418
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata
ISSN: 2583-6129
www.isjem.com





ISSN: 2583-6129
www.isjem.com

REFERENCES

1. M. Bellare, S. Keelveedhi, and T. Ristenpart. Dupless: Serveraided encryption for deduplicated storage. In USENIX Security Symposium, 2013.
2. M. Bellare, S. Keelveedhi, and T. Ristenpart. Message-locked encryption and secure deduplication. In EUROCRYPT, pages 296–312, 2013.
3. M. Bellare, C. Namprempre, and G. Neven. Security proofs for identity-based identification and signature schemes. J. Cryptology, 22(1):1–61, 2009.
4. M. Bellare and A. Palacio. Gq and schnorr identification schemes: Proofs of security against impersonation under active and concurrent attacks. In CRYPTO, pages 162–177, 2002.
5. S. Bugiel, S. Nurnberger, A. Sadeghi, and T. Schneider. Twin clouds: An architecture for secure cloud computing. In Workshop on Cryptography and Security in Clouds (WCSC 2011), 2011.
6. P. Anderson and L. Zhang. Fast and secure laptop backups with encrypted de-duplication. In Proc. of USENIX LISA, 2010.
7. M. Bellare, S. Keelveedhi, and T. Ristenpart. Dupless: Serveraided encryption for deduplicated storage. In USENIX Security Symposium, 2013.
8. M. Bellare, S. Keelveedhi, and T. Ristenpart. Message-locked encryption and secure deduplication. In EUROCRYPT, pages 296– 312, 2013.
9. M. Bellare, C. Namprempre, and G. Neven. Security proofs for identity-based identification and signature schemes. J. Cryptology, 22(1):1–61, 2009.
10. M. Bellare and A. Palacio. Gq and schnorr identification schemes: Proofs of security against impersonation under active and concurrent attacks. In CRYPTO, pages 162–177, 2002.
11. S. Bugiel, S. Nurnberger, A. Sadeghi, and T. Schneider. Twin clouds: An architecture for secure cloud computing. In Workshop on Cryptography and Security in Clouds (WCSC 2011), 2011.
12. J. R. Douceur, A. Adya, W. J. Bolosky, D. Simon, and M. Theimer. Reclaiming space from duplicate files in a serverless distributed file system. In ICDCS, pages 617– 624, 2002.
13. D. Ferraiolo and R. Kuhn. Role-based access controls. In 15th NIST-NCSC National Computer Security Conf., 1992.

ISSN: 2583-6129
www.isjem.com

14. S. Halevi, D. Harnik, B. Pinkas, and A. ShulmanPeleg. Proofs of ownership in remote storage systems. In Y. Chen, G. Danezis, and V. Shmatikov, editors, ACM Conference on Computer and Communications Security, pages 491–500. ACM, 2011.
15. J. Li, X. Chen, M. Li, J. Li, P. Lee, andW. Lou. Secure deduplication with efficient and reliable convergent key management. In IEEE Transactions on Parallel and Distributed Systems, 2013.
16. C. Ng and P. Lee. Revdedup: A reverse deduplication storage system optimized for reads to latest backups. In Proc. of APSYS, Apr 2013.
17. W. K. Ng, Y. Wen, and H. Zhu. Private data deduplication protocols in cloud storage. In S. Ossowski and P. Lecca, editors, Proceedings of the 27th Annual ACM Symposium on Applied Computing, pages 441–446. ACM, 2012.
18. R. D. Pietro and A. Sorniotti. Boosting efficiency and security in proof of ownership for deduplication. In H. Y. Youm and Y. Won, editors, ACM Symposium on Information, Computer and Communications Security, pages 81–82. ACM, 2012.
19. S. Quinlan and S. Dorward. Venti: a new approach to archival storage. In Proc. USENIX FAST, Jan 2002. [18] 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. In 3rd International Workshop on Security in Cloud Computing, 2011.
20. R. S. Sandhu, E. J. Coyne, H. L. Feinstein, and C. E. Youman. Role-based access control models. IEEE Computer, 29:38–47, Feb 1996.
21. J. Stanek, A. Sorniotti, E. Androulaki, and L. Kencl. A secure data deduplication scheme for cloud storage. In Technical Report, 2013.