International Journal of Combined Research & Development (IJCRD)
eISSN:2321-225X; pISSN:2321-2241 Volume: 13; Issue: 7; July- 2024

INFORMATION SECURITY ELIMINATION OF DUPLICATION AND RETRIEVAL USING THE PUBLIC KEY CRYPTOGRAPHY AND KEYWORD LOOKUP

Sowmya J
Assistant Professor
Department of Master of Computer Applications
The Oxford College of Engineering
sowmyaj@theoxford.edu

Shraddha Mahadev Hegde
PG Student
Department of Master of Computer Applications
The Oxford College of Engineering
hegdeshraddha440@gmail.com

Abstract:

Users still require data storage because of the preset data explosion, and moving data to the cloud is now the very common choice for both consumer and enterprise. Users may exchange and backup data more easily when they take advantage of cloud services, which lowers their storage costs. The server's storage needs decrease as a consequence. In a cloud server due to lack of trust, data published by numerous users is frequently saved multiple times. However protect client privacy; data storage on cloud servers must be secured. However, data entered into a blank form will be erased right away. In addition, we learn how indexing mechanisms for public-key searchable passwords and trapdoor- compatible joins are equivalent in ciphertext to secure deduplication. The data re-encryption key for the user's copies is placed in the archive of the written record. When duplicating files the cloud server uses the re-encryption key to create the converted files. To find the information of data, the user decrypts the converted ciphertext using his/her private key.

Keywords: Encryption, Public key Retrieval, Cloud Computing, Ciphertext Transformation, Cloud Security.

1. INTRODUCTION

This functionality may help users minimize their storage expenditures and increase their job productivity. Considering that cloud computing technology is becoming more mature. Among the many cloud service providers (CSPs) currently available in the market, the most well-known is Badu Cloud, i.e. Amazon Cloud, etc. A cloud service provider (CSP) is responsible for monitoring and managing data stored on a server in the cloud. Users can upload and store their personal data in the cloud. However, the following are regular occurrences of security vulnerabilities associated with cloud computing. Individuals and businesses alike possess a capacity to access their personal files, corporate contracts, user the records of transactions, ecological location data, and other sensitive information that is kept on the cloud server However, there are still cases where users' privacy is violated and sensitive information is leaked. More worrying is that some communications services make money by selling user data to generate company revenue. There should be a considerable amount of focus placed on the problem of safeguarding information in cloud storage. Both big data collection and cloud computing are undergoing fast development. It's because of the reason that people throughout the globe are producing an enormous amount of data, which has led to a significant rise in the needs for cloud servers. Data deduplication will be a good choice to find the solution for big data. In the case of raw data, testing for equality

can be done by direct comparison. Another point of view, user data concerns the user's personal privacy, and the act of transferring or maintaining it in unencrypted form on cloud servers might result in the risk of user privacy being compromised. Encrypting data is an effective method for protecting the privacy of users. regarding about the real situation; various users utilize different keys to encrypt data. Additionally, when there are unpredictable variables associated in the encryption process, the cipher text that is created from an identical file is different.

This is why, in situations when there are few users, it is urgently necessary to design a safe de-duplication solution for encrypted data that uses several keys. These days, safe data de-duplication solutions are frequently designed using convergent encryption [1]. Content hash keying is exposed to a number of risks, however, including chosen-plaintext attacks, phony attacks, and data leaks [2, 3], [4]. Because the hash value of a user's data file is utilized to produce the encryption key that is used in content hash keying, numerous files belonging to the same user will generate several separate keys, which will result in an issue with key management [5]. Both the method of encrypting data and the method of de-duplicating data have an effect on one another. When multiple users use the same key to resolve data, the resulting ciphertext will be identical. Direct comparison of cipher texts with one another is the strategy that's utilized to produce secure data de- duplication; In spite of , this process will results in a difficulty with key management. The key management issue may be successfully decreased if various individuals use distinct keys to encrypt data; In spite of, it is difficult to establish equality test if this is the case. In light of this, the primary study topics of the current work are to investigate how different users may encrypt knowledge with the identical key without interacting with each other, hence creating identical ciphertext shortly after encoding data that is identical, and how users can retrieve their data.

2. EXISTING SYSTEM

In the field of secure cloud storage, research and development are ongoing processes. On the other hand, there were several ideas and academic articles that discussed the theoretical features of such System and likewise possibility of putting it into practice. One thing that should be taken into consideration is that some practical implementations may not have been extensively accepted or commercially accessible at that particular time period.

DISADVANTAGES

The search efficiency, security overhead, limited search flexibility, scalability, and complex key management are all causes that should be considered.

3. PROPOSED SYSTEM

To enable an online storage system that is both safe and efficient, it is important to enable de-duplication and give users the Capability to search for encrypted data using keywords.

Cloud storage presents a number of issues, including the requirement for effective de-duplication and data recovery procedures, as well as concerns around data security and privacy.

ADVANTAGES

- Data Confidentiality
- Keyword Search over Encrypted Data
- Data Deduplication Efficiency and Security against Insider Security breaches are critical aspects worth

considering.

4. ANALYSIS

Agile Methodology

The agile strategy is a form of project administration that separates every step across a number of distinct stages in order to more effectively handle the project. This approach was developed by the Agile Software Community. In addition to the ongoing growth that occurs at each stage, it is necessary to maintain regular contact as well as cooperation with the multiple parties who are engaged. Teams start a process that includes planning, carrying out, and reviewing the work they've already performed as soon as they start working on a project. This process begins almost immediately. Constant collaboration is necessary with the various individuals that make up the collective as well as with those who have financial stakes in the project.

In the context of project management, the term "agile project management" refers to a methodology that advocates for the use of iterative processes and emphasizes the need of continuous cooperation. The base of agile project management is the notion that a project may experience ongoing improvement throughout its existence, with adjustments being implemented in a timely manner responsive way. This is the fundamental idea of agile project management. The agile approach is built on the foundation of this core concept. Among the several methods that are used to manage projects, agile project management has quickly become highly popular ways. The responsiveness of agile project management to change, the high degree of client interaction, and the flexibility of this process is mostly responsible for this. Exactly what are the components that make up the Scrum methodology?

When something is heuristic, it indicates that it is constructed on the basis of on-going education and the capacity to adjust to changing circumstances. The framework of scrum is heuristic, which implies that it is constructed on the principle of continuous learning. This accepts the reality how the team will acquire additional abilities as they gain additional expertise and that they are unaware of everything at the beginning of a project given reasons such as that they cannot comprehend everything. It also admits that your team will acquire new talents as they gain more experience. Scrum is designed to make it simpler for teams to make natural adjustments to their process in accordance with changing circumstances and increasing expectations from customers. It will be

possible for your team to continually learn and grow if the process incorporates reprioritization, whether the release cycles are maintained as short as possible. In a scrum team, that is three components that remain constant throughout the course of period of time and continue as the major focus of our concentration and dedication. These components are the "scrum team" components.

Algorithm

Attribute-based encryption:

Log encryption is a possible use for ABE. It is feasible to encrypt a log just with qualities that match the attributes of the receivers, as opposed to encrypting each individual component of the log utilizing the keys for all recipients.

Copy of the text:

Because it includes a content of plaintext as its foundation that is unintelligible by a person or machine without the appropriate cipher to decode it, cipher text can be called as encrypted or data that is encoded this is because of that it comprises a form thereof. Conversion of encrypted text into

plaintext that will read is known as decryption, which is just the opposite of encryption

5. MODULE DESCRIPTION

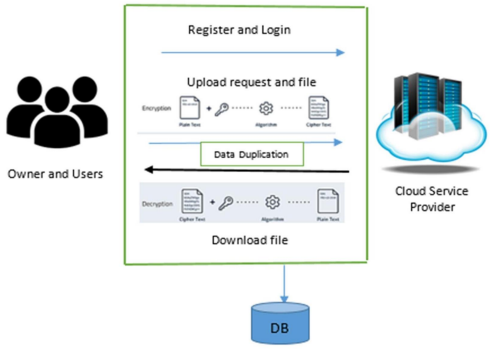

Fig.5.1 Architecture Diagram

DATA OWNER

The process involves registering, logging in to the account, and uploading the file using the encryption format with authorization from the cloud owner. Although the de-duplication process has been finished, the same file cannot be accepted for the same server since it's copy. This cloud platform is mostly used for storage purposes. I am capable to know user demand and submit a demand for re-encryption. I am also able to view the request and re-encrypt the key. I am able to log out.

DATA USER

Once the cloud user has authorized the account, the user may proceed to see their profile, search for a file, and submit a request to download it. Additionally, the user can view the key and download the content. Finally, the user can log out of their account.

TEST AUTHORITY

The account may be logged in by the appropriate authority with the relevant credentials. Users and owners are capable to see and authorize them. View the user's request to download the key and send it and Sign out.

CLOUD SERVER PROVIDER

All we must need to do is log in to the account, authorize the owner, authorize the user, see all the files which have been uploaded, send the key via the user mail, create a graph, and log out.

6. IMPLEMNTATION

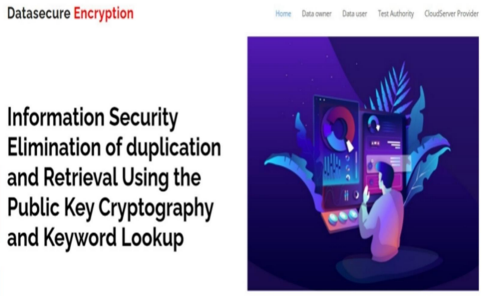

Fig.6.1 Home Page Login

Fig.6.2.Owner login page

7. RESULTS

Elimination of duplication and retrieval of information using public key cryptography and keyword lookup involves a sophisticated interplay of cryptographic techniques, data

management strategies. Public key cryptography (PKC), known for its usage of asymmetric key pairs (public and private keys), is fundamental in ensuring data security and integrity. In this context, PKC can help in the elimination of duplication by leveraging digital signatures and cryptographic hash functions. When data is ingested into a system, it is first hashed using a cryptographic hash function to generate a unique hash value. This hash value acts as a digital fingerprint for the data. The data is then signed with a digital signature created using the sender's private key, ensuring its authenticity and integrity. By storing and comparing these hash values and signatures, the system can easily identify and eliminate duplicate data entries

1	Test Case	Upload Product
2	Precondition	Load the Product View the Product details
3.	Description	View the Product details from the database
4.	Test Steps	Upload the Product View the Product
5.	Expected Output	Files are successfully uploaded in the database.
6.	Actual Output	Files are successfully uploaded in the database and it will be viewed from the database
7.	Status	Success

Table.7.1.Testcase

8. CONCLUSION

Cloud storage should provide confidential information to optimize space utilization. This work introduces a secure deduplication method that uses keyword search public key cryptography (PEKS) to evaluate data duplication in the encrypted state. By comparing the content with password-searchable trapdoors, the system can identify and remove duplicate files without compromising data security. Proxy re-encryption is used to facilitate data recovery and ensure the deduplication process is secure and efficient. The numbers parallel equations to limits the measurement with large data. Experimental results show that this method works well in cloud storage environment, preserving data security and integrity while improving performance. The cost of data deduplication may be impacted by the present systems' poor ability to measure data duplication, particularly in cases where data changes just little. Subsequent investigations endeavor to mitigate this constraint by refining the capacity to handle minute variations in information, consequently raising the total expense of deduplication and augmenting the quality of cloud storage employed.

9. FUTURE ENHACEMENT

One potential area of emphasis for future research may be the optimization of the execution of PEKS-based systems, with the aim of minimising computational overhead and enhancing search efficiency. To improve the pace of the search and decrease the duration it takes, methods can be explored by simultaneous processing, indexing, and caching.

10.REFERENCE

[1] "Reclaiming space from duplicate files in a serverless distributed file system," written by J. R. Douceur, A. Adya, W. J. Bolosky, P. Simon, and M. Theimer, was published in the proceedings of the 22ndInternational Conference on Distributed Computing Systems, which took place in Vienna,

Austria, in 2002, on pages 617–624.
[2] In the proceedings of the IEEE International Conference on Systems, Man, and Cybernetics (SMC), which took place in Banff, Alberta, Canada, in October 2017, pages 2176–2181, A. Agarwala, P. Singh, and P. K. Atrey presented their work titled "DICE: A dual integrity convergent encryption protocol for client side secure data deduplication."
[3] P. Anderson and L. Zhang, "Fast and secure laptop backups with encrypted deduplication," published in the proceedings of the 24th Large Installation System Administration Conference, held in San Jose, California, United States of America, in 2010, pages 1–12.
[4] "Side channels in cloud services: Deduplication in cloud storage," by D. Harnik, B. Pinkas, and A. Shulman-Peleg, was published in the November/December 2010 issue of IEEE Security Privacy, volume 8, number 6, pages 40–47.

[5] In June of 2014, the IEEE Transactions on Parallel Systems published an article titled "Secure deduplication with efficient and reliable convergent key management."