ISSN (Online): 2456-0448
International Journal Of Innovative Research In Management, Engineering And Technology
Vol. 9, Issue 3, April 2024

Secure Data Vault: Harnessing Elliptic Curve Cryptography for Ranking Index Based Keyword Search

[1] A. Kalaiyarasi, M.E., [2] P.V. Krithika, [3] R. Nirosha, [4] S. Sheela

[1]Assistant Professor, Department Of Information Technology, Muthayammal Engineering College (Autonomous), Rasipuram - 637 408, Tamil Nadu, India
[2] [3] [4]Student Department Of Information Technology, Muthayammal Engineering College (Autonomous), Rasipuram - 637 408, Tamil Nadu, India

Abstract: ***The widespread adoption of cloud storage services has emphasized the need to strengthen data privacy and security in cloud environments. While encryption methods provide a basic level of confidentiality, they traditionally hinder the efficient search and retrieval of specific data, limiting the usability of cloud storage systems. In response to this challenge, we propose an innovative approach to keyword search on encrypted cloud data, utilizing Elliptic Curve Cryptography (ECC) encryption. Our method harnesses ECC encryption to ensure robust data confidentiality while enabling efficient keyword-based search operations without the need for decryption. By leveraging ECC encryption and novel index structures, our approach facilitates multi-keyword ranked searches, with dynamic index updates to maintain search relevance and efficiency. Access to encrypted data remains restricted to authorized users with the necessary decryption keys, thus safeguarding sensitive information against unauthorized access. Moreover, to enhance data integrity, our framework incorporates External Validators (EVs) as trusted entities, reducing dependence on the data storage provider and reinforcing data integrity mechanisms. Through a careful integration of cryptographic techniques and index optimizations, our solution achieves a balanced trade-off between data privacy and search functionality, providing a robust mechanism for securely storing and retrieving data from cloud environments. Experimental evaluations validate the effectiveness of our proposed approach in real-world cloud storage scenarios, highlighting its potential to significantly enhance cloud security. This research offers a valuable contribution to the field by presenting a practical and efficient solution for strengthening data privacy while facilitating seamless data search and retrieval in cloud infrastructures.***

Index Terms — Cloud data storage, Data encryption, Elliptic curve cryptography, Keyword construction, Index generation, Keyword ranking, Index updating, Integrity checking, External Validators.

I. INTRODUCTION

Resource management is better handled by cloud computing since it relieves the user of the burden of finding resources for storage. A user can ask the cloud provider for more storage if needed, and when they're done using it, they can either release the storage by simply ceasing to use it or shift the data to a longer-term, less expensive storage option. Because they are no longer concerned with storage and costs associated with both new and old resources, the user is further able to use more dynamic resources efficiently. Cloud computing service models are all inside in the cloud sing and laptops, desktops, phones and tablets are acts like clients to get services from the cloud. Servers provide services to clients according to their request or pay base. Cloud computing provides a shared pool of configurable IT resources on demand, in which needs minimal effort of management to get better services. Services are based on various agreements SLA (Service Level Agreement) between service providers and consumers.

As cloud computing spreads, more and more private data, including emails, government documents, and personal health records, are being stored there. The owners of the data can be freed from the responsibility of data storage and maintenance by putting it in the cloud, allowing them to take advantage of the on-demand, high-quality data storage service. However, since the cloud server may no longer be completely trusted because the data owners and the cloud server are not in the same trusted domain, the outsourced data may be at risk. As a result, sensitive data should often be encrypted before being outsourced in order to protect data privacy and prevent unauthorized access. Given that there may be many outsourced data files, data encryption makes it very difficult to utilize data effectively. Additionally, in cloud computing, data owners can make their outsourced data available to a wide user base. During a particular session, the individual users might only want to retrieve particular data files that interest them. One of the most popular methods is to use keyword-based search to selectively retrieve files as opposed to trying to get back all of the encrypted files, which is absolutely unfeasible in cloud computing environments. Unfortunately, data encryption limits the usage of keyword searches by users, rendering outdated plaintext

search techniques useless for cloud computing. In addition, data encryption necessitates the preservation of keyword privacy because keywords frequently hold crucial details about the data files. Although keyword encryption can safeguard keyword privacy, it also makes conventional plaintext search approaches worthless in this situation.

A form of Internet-based computing known as "cloud computing" makes data and shared computing resources available on demand to computers and other devices. This paradigm enables widespread, on-demand access to a pool of configurable computing resources (such as servers, networks, storage, applications, and services), which can be swiftly deployed and released with little administration labour. Users and businesses can store and process their data in privately owned or third-party data centres that may be located far from the user, ranging in distance from across the city to across the world, thanks to cloud computing and storage options. Similar to a utility (like the electricity grid) over an electrical network, cloud computing relies on resource sharing to achieve coherence and scale economies.

In the current digital landscape, data security and privacy have emerged as critical concerns for individuals and organizations alike. With the proliferation of sensitive information stored electronically, protecting data from unauthorized access and breaches has become a top priority. Encryption stands as one of the most robust mechanisms for safeguarding data, and the Advanced Encryption Standard (AES) is a widely recognized and trusted encryption algorithm used for this purpose. However, ensuring secure storage and controlled sharing of encrypted data, while managing access to this data, presents an ongoing challenge. The proposed project with the advent of cloud computing, data owners are motivated to outsource their complex data management systems from local sites to the commercial public cloud for great flexibility and economic savings. However, sensitive data must be encrypted before being outsourced in order to ensure data privacy, making plaintext keyword search obsolete for typical data usage. Therefore, it is crucial to enable an encrypted cloud data search service. It is required to permit numerous keywords in the search request and return documents in the order of their relevance to these keywords due to the high number of data users and documents in the cloud. Related works on searchable encryption focus on single keyword search or Boolean search, and rarely sort the search results. This project aims to address these challenges by proposing and implementing an innovative solution—a comprehensive AES-based encrypted data and keyword storage system with identity based data sharing.

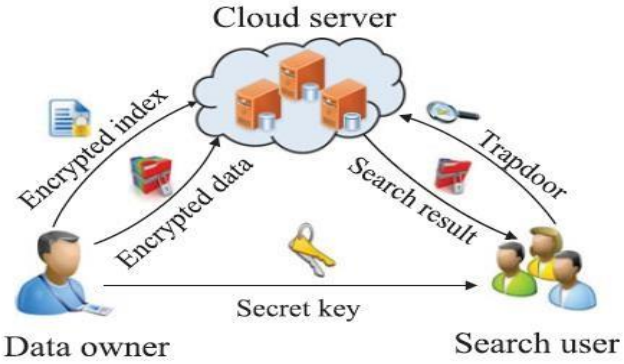

Fig.1 Cloud Data Sharing Model

II. RELATED WORK

Ma, et.al,…[1] presented an attribute-based blind signature scheme based on elliptic curve cryptography (ECC), and the security of new scheme is proved under the intractability of elliptic curve discrete logarithm problem (ECDLP). Our scheme is a key policy attribute-based signature (KP-ABS). The new scheme uses linear secret sharing scheme (LSSS) matrix technology that does not require recursive operation to achieve more flexible and fine-grained access control. In addition, the scheme is based on Elliptic Curve Cryptography (ECC) using scalar multiplication on an elliptic curve instead of a bilinear pairing operation. In addition, current attribute-based blind signature schemes are based on the access tree structure. The access tree structure can represent flexible access control policies. However, because the access structure is represented as a tree, recursion is required to perform operations. When the recursion depth reaches a certain level, the running time space of the program is affected to a certain extent. The linear secret sharing scheme (LSSS) access structure solves this problem well. LSSS uses the linear recombination property of the linear secret-sharing scheme to reconstruct secrets without recursive operation, which is more efficient, and the expressivity of LSSS and the access tree structure is equivalent. Kaur, et.al,…[2]

developed a security protocol for smart devices is difficult because of its resource constraints such as low power, bandwidth and speed. Due to this problem, traditional cryptosystem such as Diffie–Hellman and Rivest, Shamir and Adleman (RSA) cannot be used to provide lightweight authentication scheme in case of smart home network. Elliptic Curve Cryptography (ECC) can be used as an alternative to these cryptosystem since it has less arithmetic requirements and have smaller key size as compared to them. In the smart home environment, the data is transferred between the entities over an insecure channel. Authentication is a mechanism, which provides a secure communication between them. In the past, many user authentication protocols for smart home environment have been proposed, but those protocols are not secured as attacks such as insider, impersonation, etc. have been found in them. Therefore, to have a secure transmission of data over this environment, a secure communication between smart devices and user via GWN is required. In this paper, an authentication protocol has been proposed which is not vulnerable to all those attacks. Security analysis of the scheme is done and it is shown that this scheme is not secure from insecure session key agreement problem, insider attack, replay attack, offline password guessing attack and GWN bypass attack. An improved protocol is proposed by overcoming the shortcomings of the Shuai et al.'s scheme. Formal security analysis of the proposed protocol is shown using random oracle model. ProVerif tool is used to verify mutual authentication and key agreement in the proposed protocol.

Liu, et.al,…[3] presented a new searchable encryption scheme that addresses the above problems simultaneously, which makes it practical to be adopted in distributed systems. It not only enables multi-keyword search over encrypted data under a multi-writer/multi-reader setting but also guarantees the data and search pattern privacy. To prevent KGA, our scheme adopts a multi-server architecture, which accelerates search response, shares the workload, and lowers the key leakage risk by allowing only authorized servers to jointly test whether a search token matches a stored ciphertext. Present a new public-key searchable encryption scheme that can address the above security, privacy and functionality issues. Our scheme is suitable for a distributed environment which comprises multiple data writers and readers and can deploy multiple designated servers to assist the public cloud storage server to perform privacy-preserving keyword search over encrypted data. Our solution is called Searchable Encryption based on Efficient Privacy-preserving Outsourced calculation framework with multiple keys (SE-EPOM). Specifically, different from existing works, our multi-user access refers to accommodating both multiple writers (and data owners) and multiple readers (or data users) simultaneously, which is important for adoption in a distributed system.

Asharov, et.al,…[4] presented a blockchain-based multi-cloud storage data auditing scheme to protect data integrity and accurately arbitrate service disputes. Here not only introduce the blockchain to record the interactions among users, service providers, and organizers in data auditing process as evidence, but also employ the smart contract to detect service dispute, so as to enforce the untrusted organizer to honestly identify malicious service providers. This approach use the homomorphic verifiable tags to design network storage service verification mechanism in a multi-cloud system, which supports users to verify the integrity of outsourced data in batches, thus reducing the overhead during the service verification phase. We also introduce blockchain technology to enable trusted public audits. The user outsources his data to multiple CSPs.

Then user and CSPs jointly generate the integrity metadata for data auditing. During the auditing phase, the user randomly generates a challenge nonce and requires CSPs to respectively generate an integrity proof as a response based on the specified data blocks. The organizer, a special cloud service provider that is responsible for managing multiple CSPs, will aggregate all CSPs responses into a single value for the resource-constrained user to verify data integrity in batch. Interaction records between these entities are all recorded on the blockchain, and if there is a service dispute, the smart contract on the blockchain can accurately and automatically detect violations based on the records. If there is data corruption, the smart contract will ask the untrusted organizer to find the malicious service providers. The tamper-resistance feature of the blockchain forces the organizer to search honestly and fairly.

Liang, et.al,…[5] proposed a verifiable attribute based multi-keyword search scheme by using an improved k-NN technology in this paper. Firstly, compared with single keyword search, we adopt multiple keywords search to get the more accurate search results. Secondly, we improved k-NN search to make the CS return the k files which are the highest relevant to the multiple queried keywords. Additionally, we remove the splitting configuration in the traditional k-NN technology to reduce the communication costs and eliminate the key management problem. And we add the preference factor to the k-NN technology, which makes a data user to get more accurate results. Before uploaded, outsourced files are encrypted to protect their privacy and confidentiality. And by presetting an access policy, a data owner can control who can access his files to provide fine-grained searchable control. Hidden the access policy can protect the privacy of user attributes. The keyword indexes and search queries are extended and encrypted to further protect the privacy of data and users. Unlike other existing KSE schemes without considering the verification and decryption of search results, our VPAMS scheme allows data users

not only can check the correctness of the returned files, but also can decrypt the search files, which greatly improve the practicality of KSE.

III. BACKGROUND OF THE WORK

In existing system propose an anonymous EMR access control framework with multiple authorities, which provide user anonymity against the untrusted authorities. Second, we achieve traceable attribute-based Boolean keyword search, which enables the authorized user who satisfies the policy to conduct Boolean keyword search over the encrypted EMRs. Here propose a fine-grained and anonymous EMR sharing framework with multiple authorities, to guarantee the EMR confidentiality with CP-ABE, and protects user privacy against the untrusted authorities by issuing secret keys anonymously. Based on the oblivious transfer protocol, the user can get the secret key corresponding to her/his attributes, but the authority cannot get any useful information about the attribute values chosen by the user. Provide traceable attribute-based Boolean keyword search, which allows authorized users to search the encrypted EMRs with their expressive keywords. Hence, a certain EMR can be searched only if the user satisfies the access policy and the search conditions are also matched. Moreover, it can trace the traitor by revealing user identity from the trapdoor, which is required for every valid search. Proposed TABKS scheme is secure against chosen plaintext attack (CPA) and chosen keyword attack (CKA), and show that TABKS protects the AAs' security and user attribute privacy.

CP-ABE (Ciphertext Policy – Attribute Based Encryption)

Ciphertext Policy – Attribute Based Encryption (CPABE) is a type of attribute-based encryption that allows data owners to encrypt their data.

- It can only be decrypted by users who possess specific attributes that match a policy associated with the data.

IV. INDEX BASED DATA STORAGE WITH INTEGRITY CHECKING IN CLOUD

The rapid growth of cloud storage services has provided users with convenient and scalable data storage solutions. However, the outsourcing of data to third-party cloud providers raises significant concerns regarding data privacy and security. To address these concerns, researchers have developed advanced encryption techniques, such as keyword based searchable encryption, to enable secure data retrieval and search functionalities while preserving data confidentiality. Proposed system combines the benefits of searchable encryption and indexing techniques to enable efficient searching over encrypted data stored in the cloud. It allows users to store their data in an encrypted form on the cloud while still being able to perform keyword-based searches on the outsourced data. With the cloud service being more and more popular in modern society, ECC technology has become a promising orientation. It allows users to use flexible access control to access files stored in the cloud server with encrypted form. Given a cipher text and a transformation key, CSP transforms a cipher text into a simple cipher text.

The user only needs to spend less computational overhead to recover the plaintext from simple cipher text. By using this approach, users can securely store their data in the cloud while maintaining the ability to search for specific information without compromising data confidentiality. Also propose a brand new idea for achieving multi-keyword ranked search. Based on the keyword ranking, the index of the keyword was changed dynamically. Besides the search result, the cloud server should not deduce any keyword information of the file set from secure indexes and trapdoors. Keyword privacy requires indexes and queries are properly represented and securely encrypted. The proposed scheme aims to strike a balance between security and efficiency, ensuring that the data retrieval process is both fast and reliable. During the auditing phase, the user randomly generates a challenge nonce and requires CSPs to respectively generate an integrity proof as a response based on the specified data blocks. The TPA (Third Party Auditor), a special cloud service provider that is responsible for managing multiple CSPs, will aggregate all CSPs responses into a single value for the resource-constrained user to verify data integrity in batch.

ISSN (Online): 2456-0448
International Journal Of Innovative Research In Management, Engineering And Technology
Vol. 9, Issue 3, April 2024

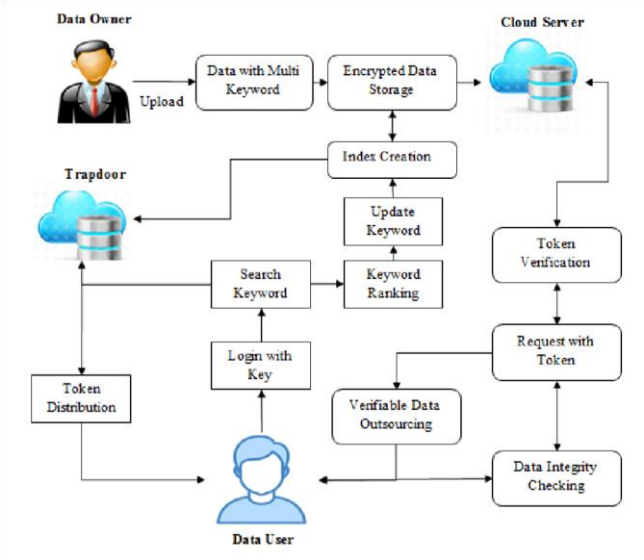

Fig 1: Secure Data Vault Architecture

Data Storage Framework

Cloud computing is the long dreamed vision of computing as a utility, where cloud customers can remotely store their data into the cloud so as to enjoy the on-demand high quality applications and services from a shared pool of configurable computing resources. Its great flexibility and economic savings are motivating both individuals and enterprises to outsource their local complex data management system into the cloud. For the first time, here explore the problem of keyword search over encrypted cloud data, and establish a set of strict privacy requirements for such a secure cloud data utilization system. This module contains three types of users such as cloud owner, cloud server and users. This module helps the owner to register their details and also include login details. An authorize user in to the system. To adds security to the user data. The login credentials are secured by encryption and they are decrypted back by the server to avoid eavesdropping. Server can store the files in cloud storage. And user can search files using keywords. This proposed framework also utilizes blockchain technology for securing the uploaded data and multi keyword storage.

Data Encryption

This module helps the owner to upload his file with encryption using ECC algorithm. This ensures the files to be protected from unauthorized users. Data owner after logging into system to added the data from crawling web and the data are stored in the structure so it can be accessed easily. The data will be large so that it should be stored in the proper structure. Elliptic Curve Cryptography (ECC) is a public key cryptography. In public key cryptography each user or the device taking part in the communication generally have a pair of keys, a public key and a private key, and a set of operations associated with the keys to do the cryptographic operations. Only the particular user knows the private key whereas the public key is distributed to all users taking part in the communication. Some public key algorithm may require a set of predefined constants to be known by all the devices taking part in the communication.

Index Creation

Design and implement a data structure to store the encrypted keywords and their corresponding index details. This structure could be a database table, a key-value store, or any other suitable data storage mechanism. When working with encrypted keywords and creating an index, special considerations need to be taken into account to ensure the security and privacy of the data.

Index is created as a list of mappings which correspond to each keyword. The list for a particular keyword contains details such as:

1. File ids of the files which has the particular keyword
2. Term frequency for each file which denotes the number of times the keyword has occurred in the file. This measures the importance of the keyword in that file.
3. Length of each file

4. Relevance score for each file
5. Number of files that has the particular keyword Data structures such as tables can be used to store this data.

Term frequency, length of the file, number of files for the keyword are used to calculate the relevance score for each file by scoring mechanisms which is discussed later in the Ranking modules.

Data Access Request

The search and data access request process involves securely searching for and retrieving encrypted data. To initiate a search, the user formulates a query describing the desired data. The query is encrypted using an appropriate algorithm and sent as a data access request over a secure channel. The receiving system verifies the user's authentication and securely processes the request while keeping the search query encrypted. The system performs an index lookup using the encrypted query, identifying the relevant encrypted data entries. Match the encrypted search query with the encrypted keywords in the index to retrieve the corresponding index details. The system then retrieves the encrypted data or record identifiers that match the query.

Token Distribution

To provide access permissions to the query user for retrieving encrypted data, a token distribution system can be implemented. When granting access permissions to the query user for retrieving encrypted data, a token distribution system is employed. The system generates unique tokens that act as access credentials for authorized users. These tokens are securely distributed to users who have been granted access to specific data or resources. To request access to encrypted data, the query user presents their token along with the data access request. The system verifies the authenticity and validity of the token to ensure that the user has the necessary permissions. The token serves as proof of authorization and allows the query user to retrieve the encrypted data. By using token distribution, access permissions can be controlled and monitored effectively. This approach enhances the security and privacy of encrypted data, ensuring that only authorized users with valid tokens can retrieve and decrypt the data they are permitted to access.

Verifiable Data Access

To facilitate a verifiable data accessing process, token verification and data access using a shared decryption key can be implemented. A verifiable data accessing process can be established through token verification and data access using a shared decryption key. Users requesting access for encrypted data using their access token. The system verifies the authenticity and validity of the token to ensure that it has not been tampered with and is issued by a trusted authority. Once the token is successfully verified, the system shares a decryption key securely with the user. This shared decryption key allows the user to decrypt the requested data while ensuring its confidentiality. By implementing token verification and data access using a shared decryption key, the system ensures that only authorized users with valid tokens can access and decrypt the data.

DATA AUDITING

Integrity checking using an auditor refers to a process in which a trusted third party, known as the auditor, verifies and ensures the integrity of data, files, or systems to detect unauthorized changes or corruption. The auditor, a trusted entity responsible for integrity checking, periodically assesses the integrity of the stored or transmitted data. The integrity checking process is continuous, ensuring that the data or systems are regularly audited for any signs of unauthorized changes or corruption.

METHODOLOGY

Elliptic Curve Cryptography:

- Elliptic Curve Cryptography (ECC) is a widely used public-key cryptography algorithm that is known for its strong security and efficiency.
- It is particularly well-suited for resource constrained environments, such as mobile devices and embedded systems.
- ECC provides a high level of security with shorter key lengths compared to traditional public-key cryptography systems like RSA.
- This makes ECC particularly well-suited for resource-constrained devices where computational efficiency and small key sizes are important.
- ECC encryption ensures that the outsourced data remains confidential and secure.
- IBSE enables efficient keyword-based searches on the encrypted data stored in the cloud.
- This proposed search method provides fine-grained access control over the outsourced data.

- It allows users to search for specific information without the need to decrypt the entire dataset.

V. CONCLUSION

Searchable encryption offers a powerful solution for preserving the privacy and confidentiality of sensitive data while enabling efficient search operations. By employing ECC, a widely adopted asymmetric encryption algorithm, the searchable encryption scheme ensures strong cryptographic protection for the data. ECC provides a high level of security with smaller key size and has been extensively studied and tested for its resistance against various cryptographic attacks. The index construction approach complements ECC encryption by allowing for efficient search operations on the encrypted data. Through the construction of indexes or data structures that capture the necessary information about the encrypted data, the scheme enables keyword-based searches or other types of queries without requiring the decryption of the entire dataset. The keyword search is adopted to capture the similarity between the keywords and the secure inner product computation is used to calculate the similarity score so as to enable result ranking. This approach significantly improves the search efficiency while maintaining the privacy and security of the encrypted information.

REFERENCES

[1] Huang, Qinlong, Guanyu Yan, and Yixian Yang. "Privacy-preserving traceable attribute-based keyword search in multi-authority medical cloud." IEEE Transactions on Cloud Computing (2023).
[2] Ma, Rui, and Linyue Du. "Attribute-based blind signature scheme based on elliptic curve cryptography." IEEE Access 10 (2022): 34221-34227. [3] Kaur, Damandeep, and Devender Kumar. "Cryptanalysis and improvement of a two-factor user authentication scheme for smart home." Journal of Information Security and Applications 58 (2021): 102787.
[4] Liu, Xueqiao, Guomin Yang, Willy Susilo, Joseph Tonien, Ximeng Liu, and Jian Shen. "Privacy-preserving multi-keyword searchable encryption for distributed systems." IEEE Transactions on Parallel and Distributed Systems 32, no. 3 (2020): 561-574.
[5] Zhang, Cheng, Yang Xu, Yupeng Hu, Jiajing Wu, Ju Ren, and Yaoxue Zhang. "A blockchain-based multi-cloud storage data auditing scheme to locate faults." IEEE Transactions on Cloud Computing 10, no. 4 (2021): 22522263.
[6] Liang, Yanrong, Yanping Li, Qiang Cao, and Fang Ren. "VPAMS: Verifiable and practical attribute-based multikeyword search over encrypted cloud data." Journal of Systems Architecture 108 (2020): 101741.
[7] Miao, Yinbin, Robert H. Deng, Kim-Kwang Raymond Choo, Ximeng Liu, and Hongwei Li. "Threshold multikeyword search for cloud-based group data sharing." IEEE Transactions on Cloud Computing 10, no. 3 (2020): 21462162.
[8] Dai, Xuelong, Hua Dai, Chunming Rong, Geng Yang, Fu Xiao, and Bin Xiao. "Enhanced semantic-aware multikeyword ranked search scheme over encrypted cloud data." IEEE Transactions on Cloud Computing 10, no. 4 (2020): 2595-2612.
[9] Wang, Haoyang, Kai Fan, Hui Li, and Yintang Yang. "A dynamic and verifiable multi-keyword ranked search scheme in the P2P networking environment." Peer-to-Peer Networking and Applications 13 (2020): 2342-2355. [10] Tariq, Husna, and Parul Agarwal. "Secure keyword search using dual encryption in cloud computing." International Journal of Information Technology 12 (2020): 1063-1072.