International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
ISSN: 2457-0362

Achieving Secure, Verifiable, and Efficient Boolean Keyword Searchable Encryption for Cloud Data Warehouse

G. MADHU
madhu_godala@yahoo.co.in, MVSR Engineering College

Abstract: Cloud Data Warehouses (CDWs) offer scalable storage and accessibility for enterprise data management, yet ensuring data confidentiality during outsourced storage remains a critical challenge. Traditional encryption techniques hinder direct querying, limiting search efficiency on encrypted data. To address this, a Boolean Keyword Searchable Encryption (BKSE) scheme integrating Partial Homomorphic Encryption (PHE) is proposed, enabling secure and efficient query execution. A Binary Tree (BTREE) and Inverted Index structure facilitate rapid Boolean searches, supporting AND and OR operations for precise query matching. A Bit Mapping Function transforms user queries into encrypted search operations, ensuring secure execution. Blockchain-based Ethereum Smart Contracts provide tamper-proof authentication and immutable access control, while HMAC authentication secures result transmission, preventing unauthorized modifications. Query performance is optimized using a multi-VM packet routing technique, which partitions query responses for parallel processing, reducing latency and enhancing scalability. Experimental validation on a financial dataset demonstrates the approach's robustness in confidentiality, integrity, and computational efficiency, ensuring verifiable and efficient querying in cloud-based data warehouses while mitigating security risks associated with encrypted storage and retrieval.

Index Terms –*Cloud data warehouse, searchable encryption, Boolean expressions, homomorphic encryption.*

1. INTRODUCTION

In the digital age, enterprises rely on data warehouses (DWs) to store and analyze vast amounts of structured and semi-structured data, enabling critical business intelligence applications. A data warehouse serves as a centralized repository where aggregated results are derived from a multidimensional framework, ensuring high-performance analytical processing. With the advent of cloud computing, cloud data warehouses (CDWs) have gained prominence due to their scalability, resilience, and accessibility, allowing enterprises to leverage distributed storage and computing resources efficiently. However, data security remains a primary concern when outsourcing sensitive data to the cloud, necessitating robust encryption techniques before data is transmitted to cloud storage [1].

A data warehouse is built on a multidimensional model that organizes information into numerous dimensions, enabling efficient query execution through Online Analytical Processing (OLAP). One of the most prevalent models in OLAP applications is the multidimensional OLAP (MOLAP) model, which utilizes precomputed data cubes. Each cube represents a pre-aggregated perspective on dimension and fact data, facilitating rapid query responses [2]. However, traditional MOLAP systems face challenges when executing analytical queries over encrypted data, as direct computations on encrypted data are infeasible without decryption [3]. Consequently, researchers have explored searchable encryption (SE) techniques to enable efficient query processing over encrypted cloud-hosted data warehouses [4].

Searchable encryption (SE) is a cryptographic method that allows data consumers to retrieve relevant information without exposing the entire dataset. SE involves extracting keywords from a data cube, encrypting them, and storing them securely in the cloud. When a user submits a search query, the cloud server executes the search operation by matching the encrypted keywords with the stored index, returning only the relevant encrypted results. Authorized users can then decrypt the results using a secret key, ensuring confidentiality [5]. Several SE schemes, such as attribute-based searchable encryption (ABSE) [6], certificateless searchable encryption [7], and Boolean searchable encryption [8], have been proposed to enhance query efficiency and security in cloud-based data warehouses. These schemes ensure that keyword searches remain privacy-preserving while allowing dynamic updates and rankable search functionalities.

To further enhance secure query execution in cloud-hosted data warehouses, researchers have integrated advanced cryptographic techniques such as verifiable search [9], multi-keyword ranked search [10], and forward/backward privacy mechanisms [11]. These advancements aim to balance security, efficiency, and usability, ensuring that enterprises can securely perform analytical processing in cloud environments without compromising sensitive information. This paper explores various SE methodologies and their applications in encrypted data warehouse environments, highlighting their effectiveness in enabling secure and efficient query processing.

2. RELATED WORK

Cloud data warehouses have become essential for storing and managing vast amounts of data in a scalable and efficient manner. However, ensuring the confidentiality and security of data stored in cloud environments remains a significant challenge. Traditional encryption techniques provide security but limit the ability to perform efficient queries on encrypted data. To address this, searchable encryption techniques have been developed to allow secure data retrieval without exposing sensitive information.

Several approaches to secure searchable encryption have been proposed in recent years. Searchable symmetric encryption (SSE) enables keyword searches on encrypted data while maintaining confidentiality. A widely studied SSE scheme [9] focuses on efficiency and security by minimizing leakage during query execution. However, it suffers from the limitation of requiring the data owner to participate in search operations, which may not be ideal for cloud environments. To mitigate this, dynamic SSE schemes [10] have been introduced, allowing updates to encrypted databases without requiring complete re-encryption. These techniques enhance performance but introduce challenges related to forward and backward privacy.

Public key encryption with keyword search (PEKS) has been another key development in searchable encryption. This technique allows a user to encrypt data under a public key and later generate trapdoors to search for specific keywords [11]. While PEKS offers a more flexible security model compared to SSE, it often suffers from inefficiencies due to expensive pairing operations. In an attempt to improve efficiency, a searchable encryption model based on lattice-based cryptography was proposed [12], which enhances resistance against quantum attacks. However, its computational complexity remains a concern for practical deployments.

Homomorphic encryption (HE) has also been explored for secure computations on encrypted

cloud data. Fully homomorphic encryption (FHE) enables arbitrary computations on encrypted data without decryption, making it a promising solution for secure cloud-based analytics [13]. Despite its strong security guarantees, FHE is computationally expensive and impractical for large-scale cloud databases. To balance efficiency and security, partially homomorphic encryption (PHE) and leveled homomorphic encryption (LHE) have been proposed, but they still impose computational overhead that limits their adoption in cloud environments.

Attribute-based encryption (ABE) is another approach that enhances data security in cloud computing by enforcing fine-grained access control. In ABE schemes [14], access policies are embedded in ciphertexts, allowing only authorized users to decrypt specific data. This model provides improved security compared to traditional access control mechanisms but introduces challenges in scalability and key management. Ciphertext-policy ABE (CP-ABE) and key-policy ABE (KP-ABE) have been developed to address these limitations, though they still suffer from high computational costs and complexity in policy updating.

In addition to encryption-based approaches, secure multi-party computation (SMPC) and zero-knowledge proofs (ZKP) have been investigated for secure query execution in cloud environments. SMPC enables multiple parties to collaboratively compute a function without revealing their individual inputs [15]. This technique is beneficial for privacy-preserving data analytics but requires high computational resources, making it difficult to scale for large cloud datasets. ZKP, on the other hand, allows a prover to demonstrate knowledge of a secret without disclosing it [16]. This approach has been integrated with searchable encryption to enhance security and integrity verification of cloud data but introduces overhead that affects real-time query performance.

Blockchain technology has also been leveraged to improve the security of encrypted cloud data warehouses. The integration of blockchain with searchable encryption has been proposed to provide tamper-proof auditability and decentralized access control [17]. In these models, encrypted indexes are stored on a blockchain to prevent unauthorized modifications and provide transparency in access logs. While blockchain-based searchable encryption offers strong security guarantees, it faces challenges related to transaction costs and scalability in high-throughput environments.

Several works have focused on optimizing index structures for encrypted search. Traditional inverted indexes, widely used in plaintext search, must be modified for secure environments. Encrypted inverted indexes [18] allow efficient keyword searches while preventing leakage of term frequency and access patterns. However, they still suffer from leakage risks due to access pattern attacks. Oblivious RAM (ORAM) techniques have been proposed to mitigate these risks by obfuscating access patterns, but they introduce significant performance overhead, making them impractical for large-scale applications.

Recent research has also explored hybrid encryption models that combine multiple encryption techniques to balance security and efficiency. For instance, a hybrid searchable encryption scheme combining SSE and HE [19] has been designed to support more expressive queries while maintaining efficiency. Similarly, machine learning-based approaches have been proposed to optimize encrypted query execution. By leveraging predictive models, these systems can

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
IJARST
ISSN: 2457-0362

adapt encryption schemes dynamically to balance security and performance requirements [20].

In summary, various approaches have been developed to secure cloud data warehouses while enabling efficient search and computation over encrypted data. While SSE and PEKS provide practical solutions for keyword searches, they face challenges related to leakage and performance overhead. Homomorphic encryption and ABE offer stronger security guarantees but introduce computational inefficiencies. Blockchain and secure computation techniques enhance security but struggle with scalability. Future research must focus on optimizing these techniques to make searchable encryption more efficient and practical for real-world cloud applications.

3. MATERIALS AND METHODS

The system introduces a secure and efficient method for querying encrypted data in Cloud Data Warehouses (CDWs) using Boolean Keyword Searchable Encryption (BKSE) with Partial Homomorphic Encryption (PHE). A Binary Tree (BTREE) and Inverted Index structure the data, enabling Boolean operations such as AND and OR for accurate search results. A Bit Mapping Function converts queries into operations on encrypted data, ensuring secure execution. Blockchain-based Ethereum Smart Contracts implemented in Solidity provide immutable storage and authentication, preventing tampering. The HMAC authentication code secures data transmission, maintaining confidentiality and integrity. To enhance query performance, a multi-VM packet routing technique distributes query processing across virtual machines, reducing latency. Additional security is ensured by using HMAC for authenticating network packets, preventing unauthorized modifications. Splitting queries into smaller packets allows parallel processing, improving response times and scalability. Validated on a bank dataset, the system demonstrates robust encryption, authentication, and efficient query execution.

Fig.1 Proposed Architecture

This (Fig.1) diagram outlines a secure cloud data storage and retrieval system. An admin uploads encrypted data, indexed using B-trees and inverted indexes, ensuring secure storage and efficient search. Users submit queries, triggering Boolean searches on the encrypted data. The system employs homomorphic encryption, enabling searches without decryption, safeguarding privacy. Performance graphs compare search request counts against CPU speed and evaluate normal versus split search times.

i) Boolean Keyword Searchable Encryption (BKSE) with Partial Homomorphic Encryption (PHE)

The system integrates BKSE with PHE to enable secure and efficient encrypted data retrieval. BKSE allows users to perform Boolean operations such as AND and OR over encrypted data without decryption, ensuring confidentiality. The PHE technique permits limited mathematical operations on ciphertexts, enabling query execution while maintaining encryption. A Binary Tree (BTREE) and an Inverted Index structure the data, optimizing search efficiency. The Bit Mapping Function translates search queries into encrypted domain

operations, preventing unauthorized data access. This combination ensures that query results are accurate while maintaining data privacy, making the system highly secure for Cloud Data Warehouses (CDWs), where sensitive information must be protected from external threats.

ii) Blockchain-Based Authentication with Ethereum Smart Contracts

Blockchain technology enhances security by ensuring data integrity and authentication through Ethereum Smart Contracts. Implemented using Solidity, these smart contracts provide immutable storage for user authentication details and query logs, preventing data tampering. Each query request is recorded on the blockchain, creating a transparent and auditable transaction history. The system verifies user credentials before allowing access to encrypted data, reducing the risk of unauthorized access. By decentralizing authentication and logging mechanisms, the blockchain eliminates single points of failure, increasing security. This approach ensures that only authenticated users can execute queries, providing a robust access control mechanism that strengthens the confidentiality and integrity of data stored in cloud-based environments.

iii) Multi-VM Packet Routing and HMAC-Based Transmission Security

To improve query performance and scalability, the system employs a multi-VM packet routing mechanism that distributes query execution across multiple virtual machines, reducing response time and workload bottlenecks. Queries are split into smaller packets and processed in parallel, ensuring faster retrieval of encrypted data. Additionally, the system integrates HMAC (Hash-based Message Authentication Code) for securing data transmission, ensuring that query packets remain unaltered during transit. HMAC authentication prevents unauthorized modifications, preserving data integrity. This dual approach of optimized packet distribution and secured transmission enhances the overall efficiency of querying encrypted data in cloud environments, reducing latency while ensuring secure communication between users and cloud servers.

iv) Modules:

The system comprises Admin and User modules for secure data management.

ADMIN:

Admin Login: The admin logs into the system using credentials to manage the system, upload datasets, encrypt data, and oversee secure search operations.

Upload Dataset: Admin uploads datasets to the system, initiating the process of data encryption, storage, and generation of necessary indexing for secure retrieval operations.

Homomorphic Encryption & B+ Tree & Inverted Index Generation: The system performs homomorphic encryption, generates a B+ tree for data storage, and creates an inverted index for efficient search operations on encrypted data.

- ***Homomorphic Encryption:*** Data is encrypted using partial homomorphic encryption, enabling search operations on the encrypted data without needing decryption.
- ***B+ Tree:*** A B+ tree structure is used to store and organize data cubes, ensuring efficient access and retrieval during search operations.
- ***Inverted Index Generation:*** An inverted index is created to map user IDs or names to their respective records, facilitating efficient search operations on encrypted data.

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
ISSN: 2457-0362

Logout: The admin can log out of the system, securely ending the session and ensuring that no unauthorized access occurs after the session ends.

USER:

User Signup: Users can sign up by providing their details, which are stored in the blockchain, ensuring secure registration and future access to the system.

User Login: Users authenticate themselves through blockchain-based login, gaining secure access to the system to perform search and retrieval operations on encrypted data.

Process Query: Users input search queries, which are processed by the system. Boolean searches are performed on encrypted data, and results are returned securely to the user.

Number of Search Graph: This graph visualizes the number of search requests processed by the system, showing performance differences between low and high CPU speeds.

Extension Load Split Graph: A graph compares the execution time of the normal search method with the extended method, where requests are split across multiple virtual machines for faster processing.

Logout: Users can log out of the system, ending the session securely to prevent unauthorized access and ensuring data confidentiality.

4. RESULTS AND DISCUSSION

To run project double click on 'run.bat' file to get below screen

In above screen python server started and now open browser and enter URL as http://127.0.0.1:8000/index.html and press enter key to get below page

In above screen click on 'Admin Login' link to get below page

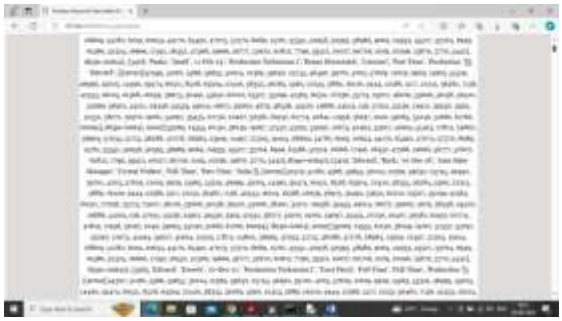
In above screen admin is login and after login will get below page

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
IJARST
ISSN: 2457-0362

In above screen click on 'Upload Dataset' link tp upload dataset and get below page

In above screen selecting and uploading 'bank.csv' dataset and then click on 'Open and Submit' button to load dataset and get below page

In above screen dataset loaded and now click on 'Homomorphic Encryption & B+ Tree & Inverted Index Generation' link to encrypt dataset and then generate BTREE + and get below output



In above screen can see encrypted data stored inside BTREE and now logout and sign up new user

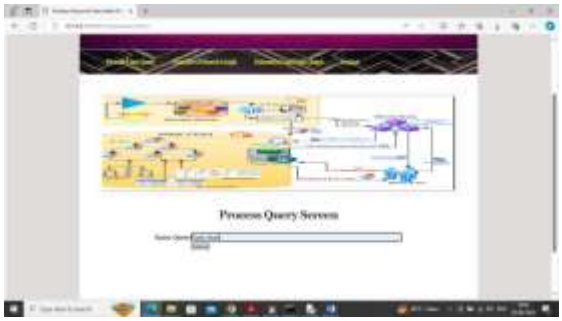
In above screen user is entering sign up details and then press button to get below output

In above screen user sign up details saved in Blockchain and in red colour text displaying all log obtained from Blockchain and this log contains details like Block No, Transaction No, hash code and many other details. Now click on 'User Login' link to get below page

In above screen user is login and after login will get below page

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
IJARST
ISSN: 2457-0362

In above screen click on 'Process Query' link to get below page



In above screen I entered some query to search detail of given person name and then press button to get below page

In above screen in blue colour text can see HMAC Extension1 authenticated code which will alert user about query result authentication and in tabular format can see the search result. Similarly you can search any details from dataset

In above screen giving another query and below is the output

In above screen can see result of another search and now click on 'Number Of Search Graph' link to get below graph

In above graph x-axis represents 'Number of Search' and y-axis represents processing speed time and orange line represents LOW CPU which has less sped and blue line represents HIGH CPU which has more speed and now click on 'Extension Load Split Graph' link to get below graph

In above graph x-axis represents 'Number of process request' and y-axis represents execution time and then orange line represents Extension execution time which is less faster because of splitting request into multiple parts so execution time will be faster. Blue line represents propose technique which will propose entire request in single CPU or VM so execution time will be more.

So by following above screens you can perform search operations on encrypted dataset.

5. CONCLUSION

The Boolean Keyword Searchable Encryption (BKSE) approach addresses critical challenges in querying encrypted data within Cloud Data Warehouses (CDWs). By integrating Partial Homomorphic Encryption (PHE) for secure storage and search operations, along with a Binary Tree (BTREE) and Inverted Index for efficient data organization, the system ensures precise and rapid query execution. The inclusion of Boolean expressions like AND and OR further enhances query accuracy, while the Bit Mapping Function facilitates seamless execution of user queries. Blockchain-based Ethereum Smart Contracts, built with Solidity, add a layer of trust and tamper-proof authentication, ensuring data integrity. The HMAC authentication code guarantees the secure transmission of search results, safeguarding against unauthorized alterations. Processing efficiency is significantly boosted by leveraging a multi-VM packet routing mechanism that optimally distributes query handling across virtual machines. Demonstrated on a bank dataset, the system effectively balances security, performance, and functionality, offering an advanced solution for secure and efficient querying in encrypted cloud environments. This development represents a significant advancement in ensuring confidentiality and integrity while maintaining practical usability for sensitive data operations.

The future scope of this system lies in enhancing scalability and performance for large-scale cloud data warehouses. Integrating advanced machine learning algorithms for dynamic query optimization and incorporating multi-cloud environments for distributed storage could further improve efficiency. Additionally, adopting privacy-preserving techniques such as Zero-Knowledge Proofs (ZKPs) could further enhance security. The system could also be extended to handle real-time data processing, allowing for faster updates and query execution, making it adaptable to a broader range of industries and use cases.

REFERENCES

[1] H. Yin, W. Zhang, H. Deng, Z. Qin, and K. Li, ''An attribute based searchable encryption scheme for cloud-assisted IIoT,'' IEEE Internet Things J., vol. 10, no. 12, pp. 11014–11023, Jun. 2023, doi: 10.1109/JIOT.2023.3242964.

[2] X. Liu, H. Dong, N. Kumari, and J. Kar, ''A pairing-free certificateless searchable public key encryption scheme for industrial Inter net of Things,'' IEEE Access, vol. 11, pp. 58754–58764, 2023, doi: 10.1109/ACCESS.2023.3285114.

[3] S. Guo, H. Geng, L. Su, S. He, and X. Zhang, ''A rankable Boolean searchable encryption scheme supporting dynamic updates in a cloud

environment,'' IEEE Access, vol. 11, pp. 63475–63486, 2023, doi: 10.1109/ACCESS.2023.3284904.

[4] B. Chen, T. Xiang, D. He, H. Li, and K. R. Choo, ''BPVSE: Publicly verifiable searchable encryption for cloud-assisted electronic health records,'' IEEE Trans. Inf. Forensics Security, vol. 18, pp. 3171–3184, 2023, doi: 10.1109/TIFS.2023.3275750.

[5] L. Chen, Y. Xue, Y. Mu, L. Zeng, F. Rezaeibagha, and R. H. Deng, ''CASE-SSE: Context-aware semantically extensible searchable symmetric encryption for encrypted cloud data,'' IEEE Trans. Services Comput., vol. 16, no. 2, pp. 1011–1022, Mar. 2023, doi: 10.1109/TSC.2022. 3162266.

[6] X. Li, Q. Tong, J. Zhao, Y. Miao, S. Ma, J. Weng, J. Ma, and K. R. Choo, ''VRFMS: Verifiable ranked fuzzy multi-keyword search over encrypted data,'' IEEE Trans. Services Comput., vol. 16, no. 1, pp. 698–710, Jan. 2023, doi: 10.1109/TSC.2021.3140092.

[7] Q. Zhang, S. Wang, D. Zhang, J. Sun, and Y. Zhang, ''Authorized data secure access scheme with specified time and relevance ranked keyword search for industrial cloud platforms,'' IEEE Syst. J., vol. 16, no. 2, pp. 2879–2890, Jun. 2022, doi: 10.1109/JSYST.2021. 3093623.

[8] X. Wang, J. Ma, X. Liu, Y. Miao, Y. Liu, and R. H. Deng, ''For ward/backward and content private DSSE for spatial keyword queries,'' IEEE Trans. Dependable Secure Comput., vol. 20, no. 4, pp. 3358–3370, Jul. 2023, doi: 10.1109/TDSC.2022.3205670.

[9] J. Fu, N. Wang, B. Cui, and B. K. Bhargava, ''A practical framework for secure document retrieval in encrypted cloud file systems,'' IEEE Trans. Parallel Distrib. Syst., vol. 33, no. 5, pp. 1246–1261, May 2022, doi: 10.1109/TPDS.2021.3107752.

[10] F. Li, J. Ma, Y. Miao, Z. Liu, K. R. Choo, X. Liu, and R. H. Deng, ''Towards efficient verifiable Boolean search over encrypted cloud data,'' IEEE Trans. Cloud Comput., vol. 11, no. 1, pp. 839–853, Jan. 2023, doi: 10.1109/TCC.2021.3118692.

[11] R. Zhou, X. Zhang, X. Wang, G. Yang, H.-N. Dai, and M. Liu, ''Device oriented keyword-searchable encryption scheme for cloud-assisted industrial IoT,'' IEEE Internet Things J., vol. 9, no. 18, pp. 17098–17109, Sep. 2022, doi: 10.1109/JIOT.2021.3124807.

[12] L. Xue, ''DSAS: A secure data sharing and authorized search able framework for e-Healthcare system,'' IEEE Access, vol. 10, pp. 30779–30791, 2022, doi: 10.1109/ACCESS.2022.3153120.

[13] Y. Yang, R. H. Deng, W. Guo, H. Cheng, X. Luo, X. Zheng, and C. Rong, ''Dual traceable distributed attribute-based searchable encryption and ownership transfer,'' IEEE Trans. Cloud Comput., vol. 11, no. 1, pp. 247–262, Jan. 2023, doi: 10.1109/TCC.2021.3090519.

[14] P. Zhang, Y. Chui, H. Liu, Z. Yang, D. Wu, and R. Wang, ''Efficient and privacy-preserving search over edge–cloud collaborative entity in IoT,'' IEEE Internet Things J., vol. 10, no. 4, pp. 3192–3205, Feb. 2023, doi: 10.1109/JIOT.2021.3132910.

[15] J. Liu, Y. Li, R. Sun, Q. Pei, N. Zhang, M. Dong, and V. C. M. Leung, ''EMK-ABSE: Efficient multikeyword attribute-based searchable encryption scheme through cloud-edge coordination,'' IEEE Internet Things J., vol. 9, no.

19, pp. 18650–18662, Oct. 2022, doi: 10.1109/JIOT.2022.3163340.

[16] Q. Liu, Y. Tian, J. Wu, T. Peng, and G. Wang, ‘‘Enabling verifiable and dynamic ranked search over outsourced data,’’ IEEE Trans. Services Com put., vol. 15, no. 1, pp. 69–82, Jan. 2022, doi: 10.1109/TSC.2019.2922177.

[17] G. Liu, G. Yang, S. Bai, H. Wang, and Y. Xiang, ‘‘FASE: A fast and accurate privacy-preserving multi-keyword top-k retrieval scheme over encrypted cloud data,’’ IEEE Trans. Services Comput., vol. 15, no. 4, pp. 1855–1867, Jul. 2022, doi: 10.1109/TSC.2020.3023393.

[18] M. Zeng, H. Qian, J. Chen, and K. Zhang, ‘‘Forward secure public key encryption with keyword search for outsourced cloud storage,’’ IEEE Trans. Cloud Comput., vol. 10, no. 1, pp. 426–438, Jan. 2022, doi: 10.1109/TCC.2019.2944367.

[19] Z.-Y. Liu, Y.-F. Tseng, R. Tso, Y.-C. Chen, and M. Mambo, ‘‘Identity certifying authority-aided identity-based searchable encryption framework in cloud systems,’’ IEEE Syst. J., vol. 16, no. 3, pp. 4629–4640, Sep. 2022, doi: 10.1109/JSYST.2021.3103909.

[20] P. Chaudhari and M. L. Das, ‘‘KeySea: Keyword-based search with receiver anonymity in attribute-based searchable encryption,’’ IEEE Trans. Services Comput., vol. 15, no. 2, pp. 1036–1044, Mar. 2022, doi: 10.1109/TSC.2020.2973570.