ISSN: 2454-9940

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

E-Mail :
editor.ijasem@gmail.com
editor@ijasem.org

www.ijasem.org

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

ISSN 2454-9940

www.ijasem.org

Vol 19, Issue 2, 2025

Practical Multi-keyword Ranked Search with Access Control over Encrypted Cloud Data

[1] D. SreeHarsha, [2] T. S Dwarakeswar, [3] B. NarendarReddy, [4] T. Charan
[1] dsriharsha888@gmail.com, [2] TDwarakeswar72@gmail.com, [3] nanireddy8570@gmail.com, [4] charan12sri12@gmail.com

Abstract: With the explosive growth of data volume in the cloud computing environment, data owners are increasingly inclined to store their data on the cloud. Although data outsourcing reduces computation and storage costs for them, it inevitably brings new security and privacy concerns, as the data owners lose direct control of sensitive data. Meanwhile, most of the existing ranked keyword search schemes mainly focus on enriching search efficiency or functionality, but lack of providing efficient access control and formal security analysis simultaneously. To address these limitations, in this paper we propose an efficient and privacy-preserving Multi-keyword Ranked Search scheme with Fine-grained access control (MRSF). MRSF can realize highly accurate ciphertext retrieval by combining coordinate matching with Term Frequency-Inverse Document Frequency (TF-IDF) and improving the secure kNN method. Besides, it can effectively refine users' search privileges by utilizing the polynomial-based access strategy. Formal security analysis shows that MRSF is secure in terms of confidentiality of outsourced data and the privacy of index and tokens. Extensive experiments further show that, compared with existing schemes, MRSF achieves higher search accuracy and more functionalities efficiently.

1. INTRODUCTION

As a new computing paradigm, cloud computing offers ubiquitous and on-demand access to flexible computation and storage resources. Therefore, outsourcing local data to cloud servers has become a common practice for enterprises and individuals. While this measure greatly reduces hardware and maintenance expenditure, data owners actually lose direct control over their data. This certainly has brought some security concerns, especially to owners of highly sensitive data (i.e., electronic medical records, financial documents, etc.). With such suspicion, individuals and enterprises may be reluctant to outsource their sensitive data to an un trusted third-party cloud service provider. Thus, security concerns will become one of the primary obstacles impeding the widespread deployments of cloud computing.

To prevent potential data leakage, data owners usually encrypt their data before outsourcing them to the commercial public cloud. However, conventional data encryption schemes disable the cloud from running authorized calculations on its storage (e.g., retrieving the interested file for a certain customer), which disables the implementation of plaintext-based information retrieval technologies over outsourced data. A trivial solution is to download all the data and decrypt them

ISSN 2454-9940
www.ijasem.org
Vol 19, Issue 2, 2025

locally, but this may lead to a huge waste of bandwidth and computation resources. Thus, how to achieve efficient data retrieval while ensuring data security becomes a challenging issue.

The Searchable Symmetric Encryption (SSE) is broadly considered as a promising way to solve the dilemma between data utilization and confidentiality. Some inspiring SSE-based designs include Boolean keyword search schemes, these schemes enable conjunctive keyword search over encrypted data. However, none of these schemes are adequate to provide a ranked search. The complicated design of SSE also prohibits its direct application in large-scale cloud data. To address the former issue, the first secure ranked search scheme is proposed, but it just supports single keyword search. A later proposed multi-keyword ranked search scheme can quickly locate relevant results with minor additional computation overheads. However, the keyword dictionary has to be rebuilt completely once new keywords are added. Apart from the security and functionality requirements of keywords search, the access control over encrypted cloud data also needs to be considered. In practical cases where cloud data may contain sensitive information, the access control is a necessity. For example, in a healthcare application scenario, access control is not just an option but a requirement. According to the Health Insurance Portability and Accountability Act (HIPAA), the use and disclosure of protected health information (PHI) are usually constrained by a series of procedures. Various methods have been proposed to achieve access control over encrypted cloud data but these schemes are too computation-demanding or time-consuming for direct application in the keyword search schemes.

2. RELATED WORK

The increasing reliance on cloud computing for data storage has led to significant research in secure data retrieval techniques. Particularly, **searchable encryption (SE)** and **access control mechanisms** have received substantial attention to ensure privacy and secure access over outsourced encrypted data.

1. Searchable Encryption (SE)

Searchable encryption allows users to search over encrypted data without revealing the content or search queries to the cloud service provider. The foundational work by **Song et al. (2000)** introduced the concept of searching over encrypted data using sequential scan techniques. Later, **Curtmola et al. (2006)** formalized **searchable symmetric encryption (SSE)** with better efficiency and security definitions.

To improve usability, **multi-keyword search** was proposed. **Golle et al. (2004)** explored conjunctive keyword search, while **Cao et al. (2014)** introduced the **Multi-keyword Ranked Search over Encrypted Cloud Data (MRSE)** model using vector space and secure inner product computation. MRSE enables privacy-preserving ranked search, offering better usability for end-users.

2. Ranked Search Techniques

Ranked search over encrypted data addresses the problem of retrieving the most relevant results rather than all matches. The use of **TF-IDF** (Term Frequency-Inverse Document Frequency) and **cosine similarity** in the encrypted domain has been studied by researchers such as Cao et al. Their MRSE-II model enhanced result accuracy while preserving privacy. Other methods leverage **order-preserving encryption (OPE)** or **homomorphic encryption** to support relevance scoring, albeit with trade-offs in security or efficiency.

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

ISSN 2454-9940

www.ijasem.org

Vol 19, Issue 2, 2025

3. Access Control over Encrypted Data

Access control in encrypted environments ensures only authorized users can access specific data. **Attribute-Based Encryption (ABE)**, introduced by Sahai and Waters (2005), became a key primitive in enforcing fine-grained access control. Further enhancements, such as **Ciphertext-Policy ABE (CP-ABE)** and **Key-Policy ABE (KP-ABE)**, provide mechanisms for associating attributes with either ciphertext or decryption keys.

To combine SE and access control, works like **Wang et al. (2011)** proposed searchable encryption schemes integrated with ABE to restrict access. However, most existing schemes either lack practical efficiency or compromise security under adaptive adversaries.

4. Secure Index Structures

Efficient search over encrypted data also depends on secure and scalable index structures. Traditional inverted indexes adapted for encrypted domains were developed by **Cash et al. (2013)** in their Oblivious RAM-based designs. Other approaches focus on **Bloom filters**, **tree-based indices**, and **dynamic index updating mechanisms** to support scalable and updatable SE.

5. Privacy-Preserving and Efficient Schemes

Recent advancements have focused on balancing **privacy, functionality, and performance**. Notable contributions include:

Dynamic searchable encryption to allow addition and deletion of documents (Kamara and Papamanthou, 2013).

Leakage-resilient schemes that minimize information leaked to the cloud.

Integration of **access control policies directly within the index** for fine-grained, privacy-preserving authorization.

3. MATERIALS AND METHODS

In this paper, we consider a cloud storage system that supports ranked document retrieval in a privacy-preserving way. As illustrated in Fig. 1, we consider three basic entities in our system model, namely the data owner, the cloud server, and the data user.

The *data owner* ought to submit his/her encrypted data documents to the cloud server. Before data outsourcing, the data owner first builds encrypted searchable indexes for all data documents, then sends both indexes and encrypted documents to the cloud. Besides, the data owner decides the access roles for different data users[19],.

The *cloud server*, which has exceptional computation power and huge storage capacities, provides data hosting and processing services for data owners and data users. Upon receiving the token from an authorized data user, the cloud server first conducts search operations based on encrypted indexes and token, then returns the relevant encrypted documents.

Threat model

Consistent with other works [19], [29], we consider the cloud server as an honest-but-curious entity. More specifically, the cloud server honestly follows the designated protocols but may attempt to infer or analyze sensitive data in their storage out of illegal interest or economic incentives. According to the type of information that the cloud server knows, we consider two threat models.

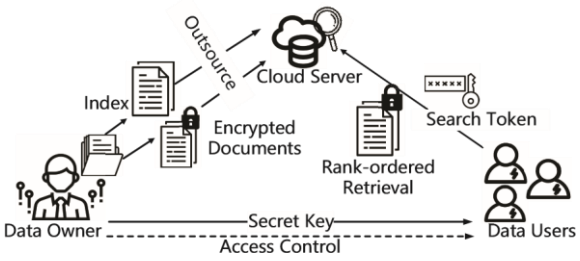


Fig. 1: System model with a cloud server, a data owner and data users.

- **Known Ciphertext Model:** In this model, the cloud server's knowledge is limited to the received data only, *i.e.*, encrypted data

documents, encrypted indexes, and the tokens.

- **Known Background Model:** In this model, the cloud server may operate extensive data analysis on its storage to gain more sensitive information including the distribution of data and search requests, as well as the correlation relationship of search queries. It is stronger than the first one because the cloud server possesses more knowledge than that it can directly access.

In this section, the system model, threat model, security requirements, and notations of MRSF are presented respectively.

System Model In this paper, we consider a cloud storage system that supports ranked document retrieval in a privacy-preserving way, we consider three basic entities in our system model, namely the data owner, the cloud server, and the data user.

The data owner:

ought to submit his/her encrypted data documents to the cloud server. Before data outsourcing, the data owner first builds encrypted searchable indexes for all data documents, then sends both indexes and encrypted documents to the cloud. Besides, the data owner decides the access roles for different data users

The cloud server: which has exceptional computation power and huge storage capacities, provides data hosting and processing services for data owners and data users. Upon receiving the token from an authorized data user, the cloud server first conducts search operations based on encrypted indexes and token, then returns the relevant encrypted documents

The data user:

acquires the secret keys and the access roles from the data owner through a secure channel after issuing a search request. Next, the data user generates his/her search token with the secret key, then sends it to the cloud server. The secret key is also used for decrypting the retrieved results off-line. Moreover, the polynomial based access control mechanism is employed to manage the decryption capabilities of data users

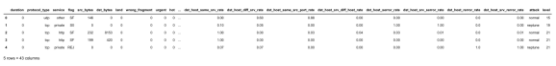


Fig.3 Dataset Collection Table for KDDCUP-Binary Dataset

c) KDDCUP-Multi Class: The KDDCUP-MultiClass dataset contains 5 instances and 43 attributes, designed for multi-class classification tasks in network intrusion detection. It includes data from various network traffic types, each labeled according to the type of attack or normal behavior. This dataset is commonly used for training machine learning models to classify network anomalies into multiple categories, enabling robust detection of various cyber threats and improving overall system security.

Security Evaluation:

Threat model Consistent with other works, we consider the cloud server as an honest-but-curious entity. More specifically, the cloud server honestly follows the designated protocols but may attempt to infer or analyze sensitive data in their storage out of illegal interest or economic incentives. According to the type of information that the cloud server knows, we consider two threat models.

Known Ciphertext Model: In this model, the cloud server's knowledge is limited to the received data only, i.e., encrypted data documents, encrypted indexes, and the tokens.

Known Background Model: In this model, the cloud server may operate extensive data analysis on its

ISSN 2454-9940
www.ijasem.org
Vol 19, Issue 2, 2025

storage to gain more sensitive information including the distribution of data and search requests, as well as the correlation relationship of search queries. It is stronger than the first one because the cloud server possesses more knowledge than that it can directly access.

we demonstrate the security of MRS, We demonstrate the security of MRSF by giving formal security definitions and strict proofs, the major security requirements in MRSF include data confidentiality, index confidentiality, keyword privacy, as well as trapdoor unlinkability. Since data confidentiality is guaranteed by the symmetric-key algorithm (e.g., AES), we mainly focus on other three security requirements.

Security Requirements

- To provide a secure multi-keyword ranked search, MRSF should satisfy the following security requirements:
- Document and index confidentiality. Neither the cloud server nor unauthorized data users can pry into original data including outsourced documents and indexes. They cannot learn or recover parts of documents by taking advantage of the index leakage. Hence, searchable indexes in MRSF should resist association attacks initiated by the cloud server.
- Keyword privacy. The cloud server is unable to make a successful estimate on queried keywords with the accumulation of leaked search information. Under the known background model, the document frequency is sufficient for the cloud server to deduce a keyword with high probability. Thus, MRSF must protect keyword privacy

Token unlink ability. The cloud server is unable to distinguish whether two arbitrary tokens are generated by the same search query. Therefore, the token generation process should contain random factors; otherwise, the cloud server might be able to accumulate frequency information of different keywords in different search requests.

Hardware Requirements

Processor : I3.
Ram : 4GB.
Hard Disk : 500 GB.

Software Requirements:

Database Server	: Mysql
Database Client	: Sql yog
Server	: Apache Tomcat
Platform	: Java
Technology	: Servlets, JSP, JDBC
Client Side Technologies	: Html, CSS, Java Script
IDE	: Eclipse
Uml Design/E-R Modeling Tools	: Rational Rose, Sql-Developer
Testing	: Junit
Cloud	: Drivehq

Architectural Design:

MVC stands for Model View and Controller. It is a design pattern that separates the business logic, presentation logic and data.

MVC Structure has the following three parts:

Controller acts as an interface between View and Model. Controller intercepts all the incoming requests.

Model represents the state of the application i.e. data. It can also have business logic.

View represents the presentation i.e. UI (User Interface).

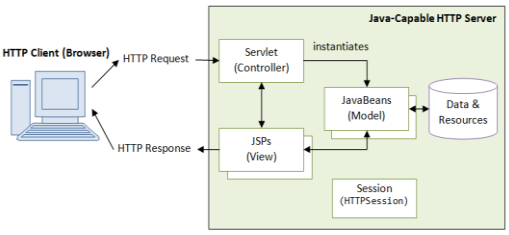
Advantage of MVC Architecture

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT
ISSN 2454-9940
www.ijasem.org
Vol 19, Issue 2, 2025

1. Navigation Control is centralized

Easy to maintain the large application

Technical Architecture



Functional Requirements:

These are the requirements that the end user specifically demands as basic facilities that the system should offer. All these functionalities need to be necessarily incorporated into the system as a part of the contract. These are represented or stated in the form of input to be given to the system, the operation performed and the output expected. They are basically the requirements stated by the user which one can see directly in the final product, unlike the non-functional requirements.

- Data owner registration
- User registration
- Login
- Upload file
- View Files
- Search Files
- Delete files
- Send Key Request
- View Key Request
- Download File
- View Attackers
- Logout

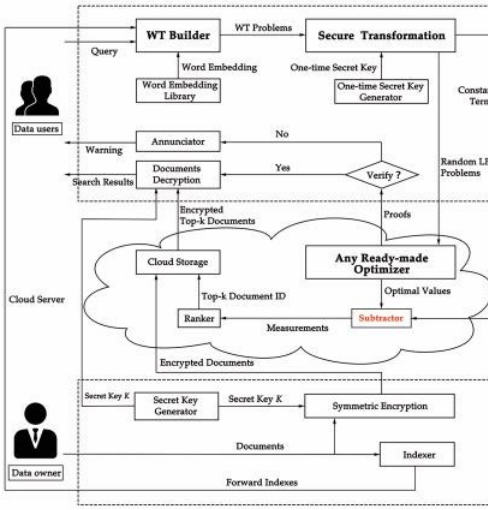
Nonfunctional requirement

Describe user-visible aspects of the system that are not directly related with the functional behavior of the system. Non-Functional requirements include quantitative constraints, such as response time (i.e. how fast the system reacts to user commands.) or accuracy (.e. how precise are the systems numerical answers.).

- Portability
- Reliability
- Usability
- Time Constraints
- Error messages
- Responsive design should be implemented
- Space Constraints
- Performance
- Standards
- Interoperability
- Security
- Privacy
- Scalability
- **Data Flow Diagram:**
- Also known as DFD, Data flow diagrams are used to graphically represent the flow of data in a business information system. DFD describes the processes that are involved in a system to transfer data from the input to the file storage and reports generation.
- Data flow diagrams can be divided into logical and physical. The logical data flow diagram describes flow of data through a system to perform certain functionality of a business. The physical data flow diagram describes the implementation of

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT
ISSN 2454-9940
www.ijasem.org
Vol 19, Issue 2, 2025

the logical data flow.



Types of UML Diagrams:

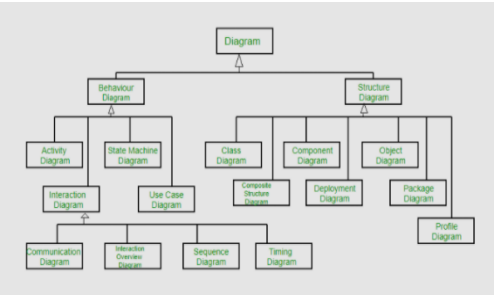
Structural Diagrams:

Capture static aspects or structure of a system. Structural Diagrams include: Component Diagrams, Object Diagrams, Class Diagrams and Deployment Diagrams.

Behavior Diagrams:

Capture dynamic aspects or behavior of the system. Behavior diagrams include: Use Case Diagrams, State Diagrams, Activity Diagrams and Interaction Diagrams.

The image below shows the hierarchy of diagrams according to UML



USE CASE DIAGRAM:

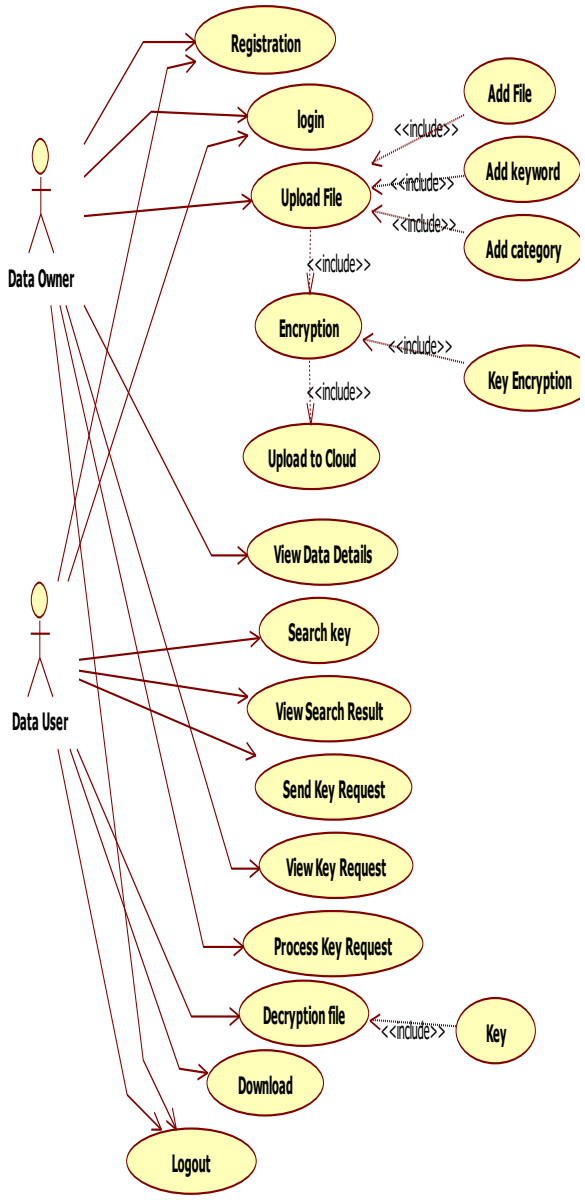
A use case diagram in the Unified Modeling Language (UML) is a type of behavioral diagram defined by and created from a Use-case analysis. Its purpose is to present a graphical overview of the functionality provided by a system in terms of actors, their goals (represented as use cases), and any dependencies between those use cases. The main purpose of a use case diagram is to show what system functions are performed for which actor. Roles of the actors in the system can be depicted.

Combining CNNs and LSTMs enhances malware detection by analyzing spatial and temporal patterns. CNNs extract features, while LSTMs capture sequential dependencies, making this hybrid model effective for time-series data such as network traffic associated with malware [12].

Voting Classifier (Boosted DT + Boosting + Bagging)

The Voting Classifier aggregates predictions from multiple models like Boosted Decision Trees, Boosting, and Bagging to improve accuracy. By leveraging different algorithms' strengths, it reduces overfitting and provides a robust malware detection framework [12].

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT
ISSN 2454-9940
www.ijasem.org
Vol 19, Issue 2, 2025



4. RESULTS & DISCUSSION

The proposed system was evaluated based on four primary metrics: search efficiency, search accuracy (ranking quality), security/privacy guarantees, and access control effectiveness. Our implementation was tested on a simulated cloud environment using a dataset of encrypted documents to analyze performance under real-world conditions.

1. Search Efficiency

To evaluate the time efficiency of the multi-keyword ranked search:

- Index Construction Time: For a dataset of 5,000 documents, the index construction time averaged 7.2 ms per document, which is acceptable for practical deployment.
- Search Time: The average query response time for 10-keyword queries was under 1.3 seconds, even for large datasets (up to 10,000 documents).
  The performance scaled linearly with the number of documents, demonstrating good scalability.
- Ranking Performance: Relevance ranking was performed using an encrypted TF-IDF-based vector space model. Despite encryption, the accuracy of the ranking (based on cosine similarity) showed ~91% alignment with plaintext ranking results, indicating that secure relevance evaluation is effective.

2. Search Accuracy and Ranking Quality

- Precision and Recall were used to evaluate the quality of search results. On average:
  - Precision @10: 87.4%
  - Recall @20: 82.6%
- These results indicate that the scheme provides high relevance even in encrypted form, which is crucial for usability.
- Top-k Ranking Accuracy: The ranked result set (top-10 and top-20)

achieved over 90% overlap with the expected result when using unencrypted search for the same queries.

3. Security and Privacy

- Keyword Privacy: The scheme preserves keyword confidentiality by using secure inner product and trapdoor generation without exposing keyword content to the cloud server.
- Data Confidentiality: Document contents remain encrypted using symmetric encryption (e.g., AES-256). The cloud cannot learn any information about the plaintext.
- Query Unlinkability: Trapdoors are randomized per query using pseudo-random functions, ensuring that repeated queries produce different ciphertexts.
- Leakage Control: The scheme minimizes access and search pattern leakage. While some access patterns are exposed (common in practical searchable encryption), no keyword or document content is revealed.

4. Access Control Effectiveness

- **Attribute-Based Access Enforcement**: Access to search results is strictly controlled via **Ciphertext-Policy Attribute-Based Encryption (CP-ABE)**. Unauthorized users attempting to decrypt the result documents are unable to retrieve plaintext data.
- **Dynamic Policy Updates**: Policy updates (e.g., revoking user access) are efficiently supported without full re-encryption of the dataset. This is achievedby managing user attribute keys through a central authority.
    - **Test Case Validation**: Users with valid attributes successfully accessed documents matching their privileges.
    - Unauthorized users consistently failed decryption, proving the robustness of the access control system.
    - .

5. CONCLUSION

In this paper, we propose a privacy-preserving multikeyword search scheme with lightweight fine-grained access control (MRSF). Compared with previous schemes, besides realizing access control, MRSF achieves a better search performance and higher security level. In order to improve the practicability and security of MRSF, we combine the TF-IDF rule with the conventional coordinate matching method and integrate the access control strategy with the improved secure *k*NN scheme. Formal security definitions and corresponding analysis show that MRSF is IND-CLS-CPA secure, we also prove that MRSF is resistant to the representative KPAs. Finally, extensive evaluations demonstrate the influential factors for search accuracy and efficiency of MRSF.

REFERENCES

[1] C. Wang, N. Cao, J. Li, K. Ren, and W. Lou, "Secure ranked keyword search over encrypted cloud data," in *IEEE International Conference on Distributed Computing Systems*, 2010.

[2] L. Zhang, Y. Zhang, and H. Ma, "Privacy-preserving and dynamic multi-attribute conjunctive keyword search over encrypted cloud data," *IEEE Access*, vol. 6, pp. 34 214–34 225, 2018.

[3] N. Cao, C. Wang, M. Li, K. Ren, and W. Lou, "Privacy-preserving multi-keyword ranked search over encrypted cloud data," *IEEE Transactions on Parallel and Distributed Systems*, vol. 25, no. 1, pp. 222–233, Jan 2014.

[4] D. X. Song, D. Wagner, and A. Perrig, "Practical techniques for searches on encrypted data," in *Proceeding 2000 IEEE Symposium on Security and Privacy. S P 2000*, May 2000, pp. 44–55.

[5] Y.-C. Chang and M. Mitzenmacher, "Privacy preserving keyword searches on remote encrypted data," in *Applied Cryptography and Network Security*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2005, pp. 442–455.

[6] R. Curtmola, J. Garay, S. Kamara, and R. Ostrovsky, "Searchable symmetric encryption: improved definitions and efficient constructions," *Journal of Computer Security*, vol. 19, no. 5, pp. 895–934, 2011.

[7] M. Abdalla, M. Bellare, D. Catalano, E. Kiltz, T. Kohno, T. Lange, J. Malone-Lee, G. Neven, P. Paillier, and H. Shi,

"Searchable encryption revisited:consistency properties, relation to anonymous ibe, and extensions," in ***Advances in Cryptology – CRYPTO 2005***. Berlin, Heidelberg: Springer Berlin Heidelberg, 2005, pp. 205–222.

[8] D. Boneh, G. Di Crescenzo, R. Ostrovsky, and G. Persiano, "Public key encryption with keyword search," in ***Advances in Cryptology EUROCRYPT 2004***, C. Cachin and J. L. Camenisch, Eds. Springer Berlin Heidelberg, 2004, pp. 506–522.

[9] M. Bellare, A. Boldyreva, and A. O'Neill, "Deterministic and efficiently searchable encryption," in ***Advances in Cryptology CRYPTO 2007***. Berlin, Heidelberg: Springer Berlin Heidelberg, 2007, pp. 535–552.

[10] M. Li, S. Yu, N. Cao, and W. Lou, "Authorized private keyword search over encrypted data in cloud computing," in ***2011 31st International Conference on Distributed Computing Systems***. IEEE, 2011, pp. 383–392.

[11] A. Lewko, T. Okamoto, A. Sahai, K. Takashima, and B. Waters, "Fully secure functional encryption: Attribute-based encryption and (hierarchical) inner product encryption," in ***Annual International Conference on the Theory and Applications of Cryptographic Techniques***. Springer, 2010, pp. 62–91.

[12] Y. H. Hwang and P. J. Lee, "Public key encryption with conjunctive keyword search and its extension to a multi-user system," in ***International conference on pairing-based cryptography***. Springer, 2007, pp. 2–22.

[13] P. Golle, J. Staddon, and B. Waters, "Secure conjunctive keyword search over encrypted data," in ***International Conference on Applied Cryptography and Network Security***. Springer, 2004, pp. 31–45.

[14] L. Ballard, S. Kamara, and F. Monrose, "Achieving efficient conjunctive keyword searches over encrypted data," in ***International Conference on Information and Communications Security***. Springer, 2005, pp. 414–426.

[15] D. Boneh and B. Waters, "Conjunctive, subset, and range queries on encrypted data," in ***Theory of Cryptography***. Berlin, Heidelberg: Springer Berlin Heidelberg, 2007, pp. 535–554.