Intelligent Multi-Keyword Searching Paradigm with Security enabled Cloud Architecture using Advanced Crypto Feature Strategy

KETURU MADAN MOHAN
Research Scholar,
Department of Computer Science and Engineering,
Jawaharlal Nehru Technological University Hyderabad,
(JNTUH), Kukatpally,,Hyderabad, 500085,
Telangana State, India..
Mail-id: madan.keturu@gmail.com,
ganga.madan@gmail.com.

DR.P.PREMCHAND
Professor,
Department of Computer Science and Engineering
University College of Engineering,
Osmania University, Hyderabad-500007,
Telangana State, India.
Mail-id: profpremchand.p@gmail.com.

Abstract—Now-a-days each and every one required to maintain the data globally, which will be suitable for searching and processing over anywhere in the world at any time. More and more clients would like to store their data into cloud servers along with the rapid development of cloud computing. New security problems have to be solved in order to help more clients process their data in public cloud. This paper is intended to provide a secure and dynamic data maintenance over cloud environment with Multi-Keyword Data Searching paradigm by using Advanced Crypto Feature (ACF) strategy. Security is the main constraint in cloud computing environment, which depicts the nature of the surveillance mechanisms in remote server based data maintenance scheme. In this paper the maximum security utility maximization with powerful Advanced Crypto Feature (ACF) is used, which process the data with 256-bit unbreakable encryption mechanism. Along with this we are using two data searching schemes such as kNN and Greedy Depth-first Search, which provides faster and easier data retrieval process with intelligent Ranking procedures. These two search algorithms are hybrid mentioned together and named it as Efficient Deep-Search Algorithm (EDSA).For all the entire cloud based data maintenance scheme is used to maintain the resources over remote place with high security features and easy accessibility.

***Index Terms*—Cloud Computing, Data Security, user Privacy, Advanced Crypto Feature, ACF.**

I. INTRODUCTION

Due to the increasing popularity of cloud computing, more and more data owners are motivated to outsource their data to cloud servers for great convenience and reduced cost in data management. However, sensitive data should be encrypted before outsourcing for privacy requirements, which obsoletes data utilization like keyword-based document retrieval. In this paper, a secured multi-keyword ranked search scheme over encrypted cloud data is proposed, which simultaneously supports dynamic update operations like deletion and insertion of documents. Specifically, the vector space model and the widely-used ACF model are combined in the index construction and query generation.

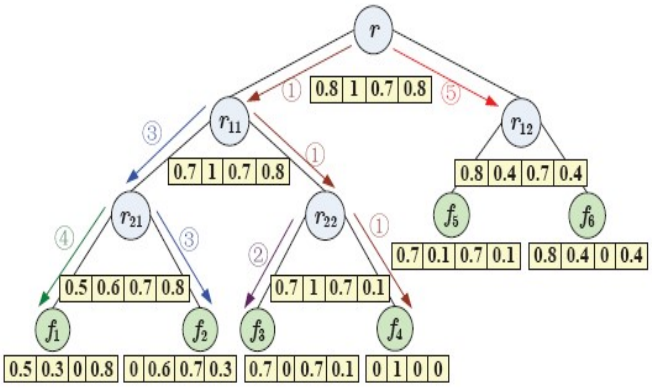

Fig.1 Tree-Index Structured Document Accumulation Model

A special tree-based index structure is constructed and proposes a Greedy Depth first Search algorithm to provide efficient multi-keyword ranked search. The secure kNN algorithm is utilized to encrypt the index and query vectors, and meanwhile ensure accurate relevance score calculation between encrypted index and query vectors. In order to resist statistical attacks, phantom terms are added to the index vector for blinding search results. Due to the use of our special tree-based index structure, the proposed scheme can achieve sub-linear search time and deal with the deletion and insertion of documents flexibly. Extensive experiments are conducted to demonstrate the efficiency of the proposed scheme.

This paragraph illustrates the details of authors and copyright owners who owned this paper and having rights for the implementation mentioned in paper. Author 1, M.Tech, (Ph.D.), Research Scholar, Department of Computer Science and Engineering, Jawaharlal Nehru Technological University Hyderabad ,(JNTUH), Kukatpally, Hyderabad, 500085, Telangana State, India. Author 2, M.Tech., Ph.D., Professor, Department of Computer Science and Engineering, University College of Engineering, Osmania University, Hyderabad-500007, Telangana State, India.

II. Related Study

In the year of 2014, the authors "B. Wang, S. Yu, W. Lou, andY. T. Hou [1]" proposed a paper titled "Privacy-preserving multikeyword fuzzy search over encrypted data in the cloud [1]", in that they described such as: enabling keyword search directly over encrypted data is a desirable technique for effective utilization of encrypted data outsourced to the cloud. Existing solutions provide multi-keyword exact search that does not tolerate keyword spelling error, or single keyword fuzzy search that tolerates typos to certain extent. The current fuzzy search schemes rely on building an expanded index that covers possible keyword misspelling, which lead to significantly larger index file size and higher search complexity. In this paper, we propose a novel multi-keyword fuzzy search scheme by exploiting the locality-sensitive hashing technique. Our proposed scheme achieves fuzzy matching through algorithmic design rather than expanding the index file. It also eliminates the need of a predefined dictionary and effectively supports multiple keyword fuzzy search without increasing the index or search complexity. Extensive analysis and experiments on real-world data show that our proposed scheme is secure, efficient and accurate. To the best of our knowledge, this is the first work that achieves multi-keyword fuzzy search over encrypted cloud data.

In the year of 2012, the authors "C. Wang, K. Ren, S. Yu, and K. M. R. Urs [2]" proposed a paper titled "Achieving usable and privacy-assured similarity search over outsourced cloud data [2]", in that they described such as the data produced by individuals and enterprises that need to be stored and utilized are rapidly increasing, data owners are motivated to outsource their local complex data management systems into the cloud for its great flexibility and economic savings. However, as sensitive cloud data may have to be encrypted before outsourcing, which obsoletes the traditional data utilization service based on plaintext keyword search, how to enable privacy-assured utilization mechanisms for outsourced cloud data is thus of paramount importance.

Considering the large number of on-demand data users and huge amount of outsourced data files in cloud, the problem is particularly challenging, as it is extremely difficult to meet also the practical requirements of performance, system usability, and high-level user searching experiences. In this paper, we investigate the problem of secure and efficient similarity search over outsourced cloud data. Similarity search is a fundamental and powerful tool widely used in plaintext information retrieval, but has not been quite explored in the encrypted data domain [2]. Our mechanism design first exploits a suppressing technique to build storage-efficient similarity keyword set from a given document collection, with edit distance as the similarity metric. Based on that, we then build a private trie-traverse searching index, and show it correctly achieves the defined similarity search functionality with constant search time complexity. We formally prove the privacy-preserving guarantee of the proposed mechanism under rigorous security treatment. To demonstrate the generality of our mechanism and further enrich the application spectrum, we also show our new construction naturally supports fuzzy search, a previously studied notion aiming only to tolerate typos and representation inconsistencies in the user searching input [2].

In the year of 2010, the authors "J. Li, Q. Wang, C. Wang, N. Cao, K. Ren, and W. Lou [3]" proposed a paper titled "Fuzzy keyword search over encrypted data in cloud computing [3]", in that they described such as Cloud Computing becomes prevalent, more and more sensitive information are being centralized into the cloud. For the protection of data privacy, sensitive data usually have to be encrypted before outsourcing, which makes effective data utilization a very challenging task.

Although traditional searchable encryption schemes allow a user to securely search over encrypted data through keywords and selectively retrieve files of interest, these techniques support only exact keyword search. That is, there is no tolerance of minor typos and format inconsistencies which, on the other hand, are typical user searching behavior and happen very frequently. This significant drawback makes existing techniques unsuitable in Cloud Computing as it greatly affects system usability, rendering user searching experiences very frustrating and system efficacy very low. In this paper, for the first time we formalize and solve the problem of effective fuzzy keyword search over encrypted cloud data while maintaining keyword privacy [3]. Fuzzy keyword search greatly enhances system usability by returning the matching files when users' searching inputs exactly match the predefined keywords or the closest possible matching files based on keyword similarity semantics, when exact match fails. In our solution, we exploit edit distance to quantify keywords similarity and develop an advanced technique on constructing fuzzy keyword sets, which greatly reduces the storage and representation overheads. Through rigorous security analysis, we show that our proposed solution is secure and privacy-preserving, while correctly realizing the goal of fuzzy keyword search [3].

In the year of 2011, the authors "B. Zhang and F. Zhang [4]" proposed a paper titled "An efficient public key encryption with conjunctive-subset keywords search [4]", in that they described such as the problem of conjunctive with subset keywords search function, discuss the drawbacks about the existed schemes, and then give out a more efficient construction of Public Key Encryption with Conjunctive-Subset Keywords Search (PECSK) scheme. A comparison with other schemes about efficiency will be presented. We also list the security requirements of our scheme, then give out the security analysis.

III. System Summary

A. Existing System

The existing techniques on keyword-based information retrieval, which are widely used on the plaintext data, cannot be directly applied on the encrypted data. Downloading all the data from the cloud and decrypt locally is obviously impractical. All these multi keyword search schemes retrieve search results based on the existence of keywords, which cannot provide acceptable result ranking functionality. However, sensitive data should be encrypted before outsourcing for privacy requirements, which obsoletes data utilization like keyword-based document retrieval.

Disadvantages

- Lack of Data Security and Missing Data Integrity
- Trust Failure nature in maintenance and Complex Retrieval Process
- Slow in process.

B. Proposed System

In the proposed system, a special tree-based index structure is constructed and we propose a "Greedy Depth-first Search" algorithm to provide efficient multi-keyword ranked search. The proposed scheme can achieve sub-linear search time and deal with the proper verification of documents flexibly. Extensive experiments are conducted to demonstrate the efficiency of the proposed scheme. A power data security algorithm called Advanced Crypto Feature Strategy is implemented, which process the data with 256-bit unbreakable encryption mechanism.

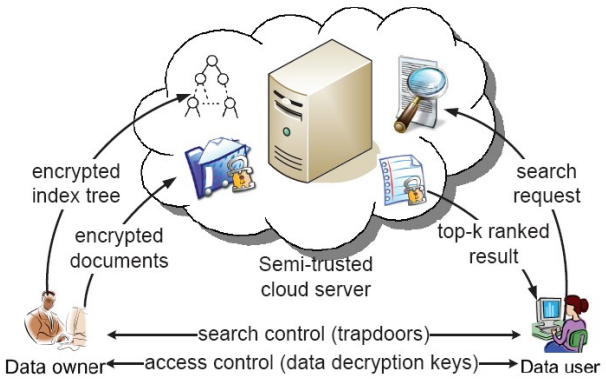

Fig.2 Proposed System Architecture

Advantages

- Algorithm to provide efficient multi-keyword ranked search.
- The secure kNN algorithm is utilized to encrypt the index and query vectors.
- Propose a "Greedy Depth-first Search" algorithm based on this index tree.
- Algorithm achieves better-than-linear search efficiency but results in precision loss.
- The ACF algorithm is suitable for security and similar search as well as provide exact ranking.

IV. System Implementation

This paper adopts the new methodologies and those methodologies are implemented by using the following modules, which are all will be briefly explained below.

A. User Authorization and Authentication

Proxy defines the security by means of user authorization and authentication. Proxy signature is a signature scheme, in which an original signer can delegate his/her signing capability to a proxy signer, and then the proxy signer generates a signature on behalf of the original signer. From a proxy signature, a verifier can be convinced of the original signer's agreement on the signed message. Researchers have proposed three different kinds of proxy signature algorithms: full delegation, partial delegation and partial delegation by warrant. The former two are eliminated by partial delegation with warrant which is proved to be more secure and practical, so we also use partial delegation with warrant in our protocol design. Let A be an original signer who has an authentic key pair (PrKA and PuKA), and B be a proxy signer who has an authentic key pair (PrKB and PuKB). Let mw be A's warrant information for the delegation, which has semantic means including the original signer's identity, some information about the proxy signer (for example the identity), period of delegation validity, the qualification of messages on which the proxy signer can sign, etc. Let dA Sign PrKA; be A's signature on the warrant mw using his/her private key PrKA. A transmits dA to the proxy signer B.

B. Trusted Data Maintenance and Data Security

This trusted data maintenance module describes the unencrypted dynamic multi-keyword ranked search (UDMRS) scheme which is constructed on the basis of vector space model and KBB tree. Based on the UDMRS scheme, two secure search schemes kNN and Greedy Depth-first Search are constructed against threat models, respectively. In the Data Security norms, we just consider the cloud provider is semi-trusted: honest but curious, which means that cloud servers would follow our proposed protocol in general, but would try to find out as much secret information as possible based on each group member's inputs. In general, we assume cloud servers are interested in data contents and group member's security information rather than other secret information. Cloud servers might collude with some malicious members for the purpose of getting data contents and group members' private information. Our scheme should satisfy the security

requirements of backward secrecy and forward secrecy. The former one ensures that the revoked user cannot decrypt new cipher texts. The later one ensures that the newly joined user can also access and decrypt the previously published data. This two security requirements are usually used in some cloud based data sharing scenarios. A potential adversary may be a former group member or any one out of the group. We assume that an adversary can be a passive attacker who could be a man-in-the-middle to monitor the communications among the group members and cloud servers. A former group member can collude with cloud servers and try to access data contents shared in his/her former group. An active adversary is able to impersonate an legitimate group member to gain some right. In general, we say that our scheme is secure if no adversary can succeed with any possible attacks mentioned above.

C. Content Based Data Search

The Content Based Data Search module eliminates the problem of unwanted confusions and problems over data mining scenario, which is achieved by means of structured data maintenance module. The structured data storage scheme fully describes about the flow of data structure maintenance and the concept of implicit data mining and document maintenance schema. Once the resource owner uploads the respective document it checks for the reference schema of the existing document for reference, if the document is presented in the database server then the following document is sequenced under the existing document otherwise it creates a new schema for the following document, so that the data into the database server is maintained in the structured manner. All the data into the server is based on the cluster format and provides the frequent access for the user search between the server and the data client (user).

D. Page Ranking Scenario

The PageRank model is a ranking methodology used by Search engines to rank the resulting in their search results. PageRank is a way of measuring the importance of website pages. According to Google: PageRank works by counting the number and quality of links to a page to determine a rough estimate of how important the website is. The underlying assumption is that more important websites are likely to receive more links from other websites. It is not the only algorithm used by Google to order search engine results, but it is the first algorithm that was used by the company, and it is the best-known.

V. Result and Discussion

This section briefly describes the results of the proposed approach and the algorithms are depicted with proper accuracy ratio. The proposed system algorithms and its associated accuracy levels are measured as well as portrait over the following section in graphical manner. The following figure, Fig.3 illustrates that the searching efficiency of the proposed approach Efficient Deep-Search Algorithm (EDSA), which is nothing but an integration of kNN and Greedy Depth-first Search algorithms.

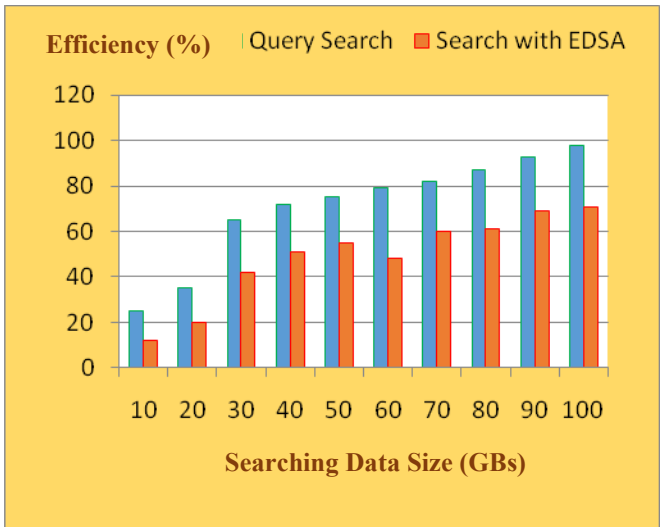

Fig.3 Searching Efficiency of EDSA and regular Query Search Algorithms

The following figure, Fig.4 illustrates that the Storage efficiency of the proposed approach called Multi-Keyword based Rank Enabled ACF and regular cloud storage schemes.

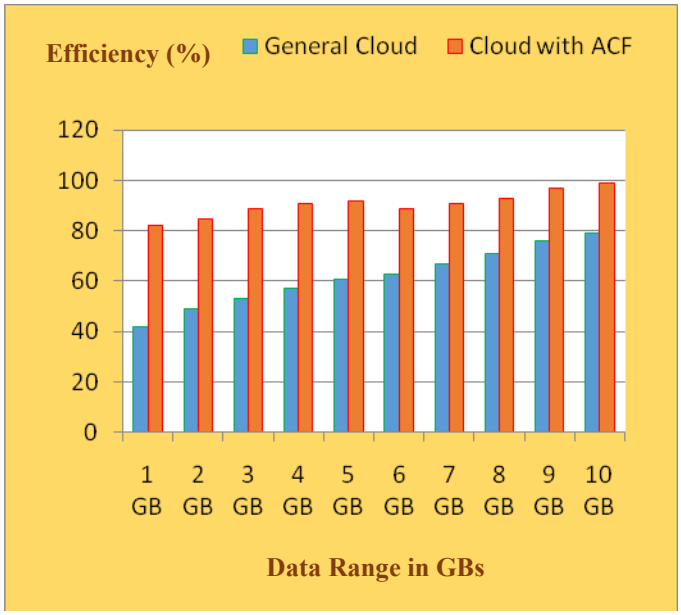

Fig.4Data Storage Efficiencyof General Cloud vs. Cloud with ACF

The following figure, Fig.5 illustrates that the cost efficiency of the proposed approach called duplication free ACF and regular cloud storage schemes.

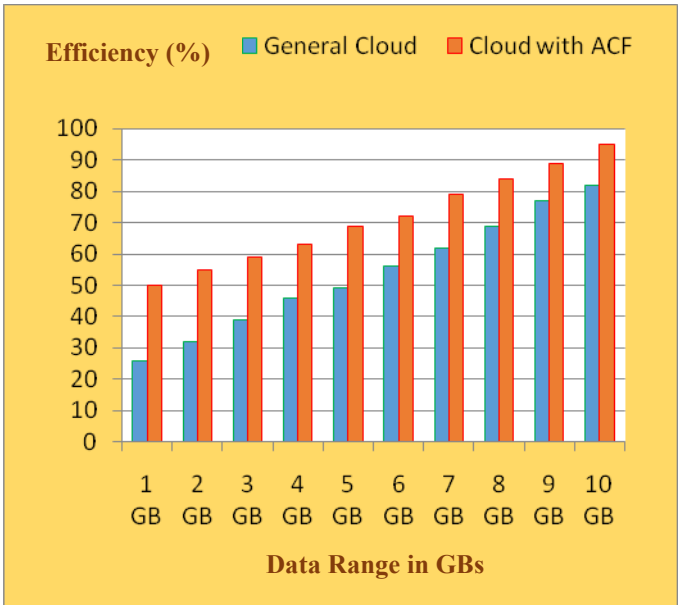

Fig.5Cost Efficiency of General Cloud vs. Cloud with ACF

The following figure, Fig.6 illustrates that the algorithm ACF crypto accuracy levels of the proposed approach and regular cloud storage encryption schemes.

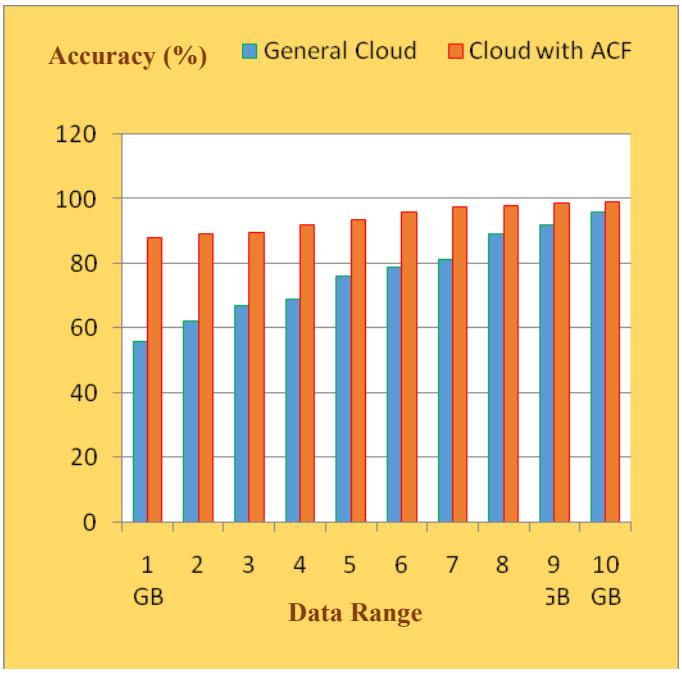

Fig.6Crypto Accuracy Levels of General Cloud vs. Cloud with ACF

VI.

VII. Conclusion and Future Scope

In this system, a secure, efficient and dynamic search scheme is proposed, which supports not only the accurate multi-keyword ranked search but also the dynamic deletion and insertion of documents. We construct a special keyword balanced binary tree as the index, and propose a Greedy Depth-first Search algorithm to obtain better efficiency than linear search. In addition, the parallel search process can be carried out to further reduce the time cost. The security of the scheme is protected against two threat models by using the secure kNN algorithm. Experimental results demonstrate the efficiency of our proposed scheme. There are still many challenge problems in symmetric Encryption schemes. In the proposed scheme, the data owner is responsible for generating updating information and sending them to the cloud server. Thus, the data owner needs to store the unencrypted index tree and the information that are necessary to recalculate the storage values. Such an active data owner may not be very suitable for the cloud computing model.

It could be a meaningful but difficult future work to design a dynamic searchable encryption scheme whose updating operation can be completed by cloud server only, meanwhile reserving the ability to support multi-keyword ranked search. In addition, as the most of works about searchable encryption, our scheme mainly considers the challenge from the cloud server. Actually, there are many secure challenges in a multi-user scheme. Firstly, all the users usually keep the same secure key for trapdoor generation in a Symmetric Encryption scheme. In this case, the revocation of the user is big challenge. If it is needed to revoke a user in this scheme, we need to rebuild the index and distribute the new secure keys to all the authorized users. Secondly, Symmetric Encryption schemes usually assume that all the data users are trustworthy. It is not practical and a dishonest data user will lead to many secure problems. For example, a dishonest data user may search the documents and distribute the decrypted documents to the unauthorized ones. Even more, a dishonest data user may distribute his/her secure keys to the unauthorized ones. In the future works, we will try to improve the Symmetric Encryption scheme to handle these challenge problems.

References

[1] B. Wang, S. Yu, W. Lou, andY. T. Hou, “Privacy-preserving multikeyword fuzzy search over encrypted data in the cloud,” in IEEE INFOCOM, 2014.
[2] C. Wang, K. Ren, S. Yu, and K. M. R. Urs, “Achieving usable and privacy-assured similarity search over outsourced cloud data,” in INFOCOM, 2012 Proceedings IEEE. IEEE, 2012, pp. 451–459.
[3] J. Li, Q. Wang, C. Wang, N. Cao, K. Ren, and W. Lou, “Fuzzy keyword search over encrypted data in cloud computing,” in INFOCOM, 2010 Proceedings IEEE. IEEE, 2010, pp. 1–5.
[4] B. Zhang and F. Zhang, “An efficient public key encryption with conjunctive-subset keywords search,” Journal of Network and Computer Applications, vol. 34, no. 1, pp. 262–267, 2011.

[5] D. Boneh, G. Di Crescenzo, R. Ostrovsky, and G. Persiano, "Public key encryption with keyword search," in Advances in Cryptology-Eurocrypt 2004. Springer, 2004, pp. 506–522.
[6] D. Boneh, E. Kushilevitz, R. Ostrovsky, and W. E. Skeith III, "Public key encryption that allows pir queries," in Advances in Cryptology-CRYPTO 2007. Springer, 2007, pp. 50–67.
[7] D. X. Song, D. Wagner, and A. Perrig, "Practical techniques for searches on encrypted data," in Security and Privacy, 2000. S&P 2000. Proceedings. 2000 IEEE Symposium on. IEEE, 2000, pp. 44–55.
[8] E.-J. Goh et al., "Secure indexes." IACR Cryptology ePrint Archive, vol. 2003, p. 216, 2003.
[9] Y.-C. Chang and M. Mitzenmacher, "Privacy preserving keyword searches on remote encrypted data," in Proceedings of the Third international conference on Applied Cryptography and Network Security. Springer-Verlag, 2005, pp. 442–455.
[10] R. Curtmola, J. Garay, S. Kamara, and R. Ostrovsky, "Searchable symmetric encryption: improved definitions and efficient constructions," in Proceedings of the 13th ACM conference on Computer and communications security. ACM, 2006, pp. 79–88.
[11] J. Li, Q. Wang, C. Wang, N. Cao, K. Ren, and W. Lou, "Fuzzy keyword search over encrypted data in cloud computing," in INFOCOM, 2010 Proceedings IEEE. IEEE, 2010, pp. 1–5.
[12] M. Kuzu, M. S. Islam, and M. Kantarcioglu, "Efficient similarity search over encrypted data," in Data Engineering (ICDE), 2012 IEEE 28th International Conference on. IEEE, 2012, pp. 1156–1167.
[13] C. Wang, K. Ren, S. Yu, and K. M. R. Urs, "Achieving usable and privacy-assured similarity search over outsourced cloud data," in INFOCOM, 2012 Proceedings IEEE. IEEE, 2012, pp. 451–459.
[14] B. Wang, S. Yu, W. Lou, and Y. T. Hou, "Privacy-preserving multikeyword fuzzy search over encrypted data in the cloud," in IEEE INFOCOM, 2014.
[15] P. Golle, J. Staddon, and B. Waters, "Secure conjunctive keyword search over encrypted data," in Applied Cryptography and Network Security. Springer, 2004, pp. 31–45.