iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 9 **Issue:** VII **Month of publication:** July 2021

DOI: https://doi.org/10.22214/ijraset.2021.36672

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue VII July 2021- Available at www.ijraset.com

Multi Keyword Search on Encrypted Data with Ranking

Revati M Wahul[1], Arpit Wadibhasme[2], Akshay Kumawat[3], Hemal Asapuri[4], Ankita Bidkar[5]
[1]*Assistant Professor, Department Of Computer Engineering, M.E.S College of Engineering (Wadia), Pune, Maharashtra, India*
[2, 3, 4, 5]*Students, Department Of Computer Engineering, M.E.S College of Engineering (Wadia), Pune, Maharashtra, India*

Abstract: ***To maintain the anonymity of users, cloud storage owners often outsource encrypted documents. As a consequence, it is important to establish efficient and precise cypher text search techniques. One issue would be that the connection between documents is typically obscured during the encryption process, resulting in a significant deterioration of search accuracy efficiency. Additionally, the volume of data stored in data centers has exploded. This will make it significantly more difficult to create cipher text search schemes capable of providing efficient and reliable online information retrieval on large quantities of encrypted data. The paper proposes a hierarchical clustering approach in order to accommodate additional search semantics and to satisfy the demand for fast cipher text search in a big data environment. The proposed hierarchical approach clusters documents according to their minimum importance levels and then sub-clusters them until the maximum cluster size is reached. This approach can achieve linear computational complexity throughout the search process, spite of the fact that its size of the record set grows exponentially. The minimum hash sub-tree structure is used in this paper to check the validity of search results. The results demonstrate that as the number of documents in the dataset increases, the proposed method's search time increases linearly, while the conventional method's search time increases exponentially. Additionally, the suggested method outperforms the standard method in terms of rank privacy and document relevance.***

Keywords: ***authorized search, searchable keyword, secret key, cloud owner key***

I. INTRODUCTION

In this paper, each documentation is represented by a sequence, which can be interpreted as a point in a high-dimensional space. All records could be classified as per their relationships with one another. In many other terms, points with short distances in a high-dimensional space can be categorized into one of many groups. By concentrating on important categories and disregarding irrelevant ones, search time can be greatly reduced. In comparison to the total number of documents in the database, the user is interested in a relatively small number of documents. Due to the small number of desired documents, a single category may be further subdivided into many subcategories. Rather than using the standard sequence search technique, a backtracking algorithm is used to search the target records. The cloud server will first search the categories for the tiniest possible sub-category. The cloud server will then select the k documents needed from the sub-category with the fewest documents. The user specifies the value of k in advance and sends it to the cloud server. If the current sub-category is unable to satisfy the k documents, the cloud server will revert to the parent category and choose the desired documents from its brother categories. This process will be repeated until the required number of k documents is reached or the root is reached. To validate the search result's legitimacy, a verifiable structure based on a hash function is developed. Any document can be hashed, with the hash result serving as a representation of the document. The hashed results of documents will be hashed again with the category information intended for them, and the result will be used to represent the current category. Likewise, each category will be represented by the hash value of the current category and its subcategories. A virtual root is created to represent all the data and categories. The virtual core is represented by the hash result of concatenating all of the categories in the first level. To ensure that the virtual root can be checked, it will be signed. Rather than verifying each document in the search result, the user only needs to check the virtual root.

A. Motivation

Another important utility feature is data exchange, which involves sending and receiving data files. A data consumer (e.g., a patient) should be able to access his or her top-k data files regarding a particular case from various data owners in a personal health record system (e.g., health monitors, hospitals, and doctors). Employees of an organization should also be able to search data files that have been outsourced by other employees. A privacy-preserving graded multi-keyword search in a multi-user model (PRMSM) has recently been suggested as a solution to the multi-keyword search problem in the multiple data owner's model. However, since PRMSM matches various cypher texts from different data owners even for the same question, it is inefficient and potentially costly for frequent queries.

B. Problem Statement

Multiple keyword searches over cypher text was not possible with a searchable encryption scheme that applied to encrypted data stored on the cloud. It also takes a long time to search.

II. LITERATURE SURVEY

Yan-Cheng Chang and Michael Mitzenmacher," Privacy Preserving Keyword Searched on Remote Encrypted Data"[1].As Present Consider the following scenario: a user U wishes to store his files in an encrypted format on a remote file server S. Later, user U wishes to quickly retrieve a subset of the encrypted files that contain (or are indexed by) unique keywords while concealing the keywords and maintaining the remote files' security. For instance, a user may wish to store encrypted old e-mail messages on a Yahoo or another large vendor's server and then retrieve specific messages using a mobile device. In this paper, we present solutions to this problem that are based on well-defined security requirements. Our schemes are useful in that they do not rely on public key cryptography. In fact, the encryption method used on the remote files has no effect on our system. Additionally, they are incremental in nature, allowing you to upload new files that are unaffected by previous requests but also searchable for future ones.

Yong Ho Hwang1 and Pil Joong Lee2," Public Key Encryption with Conjunctive Keyword Search and Its Extension to a Multi-user System"[2].As Present We examine the problem of public key encryption using conjunctive keyword search (PECK). The keyword searchable encryption enables a user to store his data on an untrusted server while still being able to search it selectively without leaking information. The PECK scheme provides a document search using many keywords in a public key environment. In the random oracle model, we first construct an efficient PECK scheme whose security is demonstrated over a decisional linear Diffie-Hellman assumption. Compared to other systems, ours has the smallest cypher text and private key sizes, as well as a comparable computing overhead. Second, we fix flaws in previous schemes' security proofs and show that they cannot guarantee complete security. Finally, we introduce a new model known as a multi-user PECK scheme, which can minimise processing and communication overhead while still handling storage in a server for multiple users efficiently.

Mihir Bellare, Alexandra Boldyreva, and Adam O'Neill," Deterministic and Efficiently Searchable Encryption"[3].As Present We examine the problem of public key encryption using conjunctive keyword search (PECK). The keyword searchable encryption enables a user to store his data on an untrusted server while still being able to search it selectively without leaking information. The PECK scheme provides a document search using many keywords in a public key environment. In the random oracle model, we first construct an efficient PECK scheme whose security is demonstrated over a decisional linear Diffie-Hellman assumption. Compared to other systems, ours has the smallest cypher text and private key sizes, as well as a comparable computing overhead. Second, we fix flaws in previous schemes' security proofs and show that they cannot guarantee complete security. Finally, we introduce a new model known as a multi-user PECK scheme, which can minimise processing and communication overhead while still handling storage in a server for multiple users efficiently.

Dawn Xiaodong Song David Wagner Adrian Perrig," Practical Techniques for Searches on Encrypted Data"[4].As Present It is desirable to store data on data storage servers such as mail servers and file servers in encrypted form to reduce security and privacy risks. However, in most cases, this means that functionality must be sacrificed in order to achieve security. For example, if a client wishes to retrieve only documents containing certain words, it was not previously known how to let the data storage server perform the search and answer the query without loss of data confidentiality. We describe our cryptographic schemes for searching encrypted data and provide security proofs for the resulting crypto systems in this paper. Our techniques have a number of crucial advantages. They are provably secure: they provide provable secrecy for encryption, in the sense that the untrusted server cannot learn anything about the plaintext when only given the cipher text; they provide query isolation for searches, meaning that the untrusted server cannot learn anything more about the plaintext than the search result; they provide controlled searching, so that the untrusted server cannot search for an arbitrary word without the user's authorization; they also support hidden queries, so that the user may ask the untrusted server to search for a secret word without revealing the word to the server. The algorithms we present are simple, fast (for a document of length n, the encryption and search algorithms only need O(n) stream cypher and block cypher operations), and introduce almost no space and communication overhead, and hence are practical to use today.

Ning Cao, Cong Wang, Ming Li, Kui Ren, and Wenjing Lou, "Privacy-Preserving Multi-keyword Ranked Search over Encrypted Cloud Data"[5].As Present Data owners are encouraged to outsource their complex data management systems from local sites to the commercial public cloud due to the advent of cloud computing, which provides greater flexibility and cost savings. However, in order to protect data privacy, confidential data must be encrypted before being outsourced, rendering conventional data usage based on plaintext keyword search obsolete.

As a result, allowing an encrypted cloud data search service is critical. Since there are so many data users and documents in the cloud, it's important to allow multiple keywords in the search request and return documents in the order of their importance to these keywords. Similar research on searchable encryption tends to concentrate on single keyword or Boolean keyword searches, and they seldom filter the data. We identify and solve the difficult problem of privacy-preserving multi-keyword ranked search over encrypted cloud data (MRSE) for the first time in this paper. We also provide a set of strict privacy standards for such a secure cloud data utilization scheme.

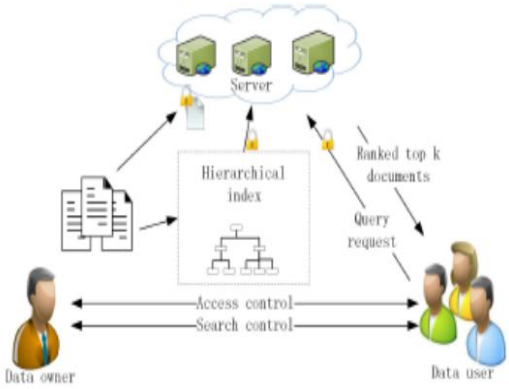
III. PROPOSED SYSTEM

The problem of maintaining close relationships between different plain documents over an encrypted domain is addressed in this scheme, and a clustering approach is proposed to solve the problem.

To speed up the server-side searching process, we proposed the MRSE-HCI architecture. The search time is reduced to a linear time rather than an exponential time as the record set grows exponentially. To boost rank privacy, we devise a search strategy. On top of the above clustering technique, this search strategy uses the backtracking algorithm.

As the amount of data grows, the proposed method's advantage in rank privacy becomes more evident. We provide a verification mechanism to ensure the correctness and completeness of search results by applying the Merkle hash tree and cryptographic signature to an authenticated tree structure.

IV. SYSTEM ARCHITECTURE



A. *Module*

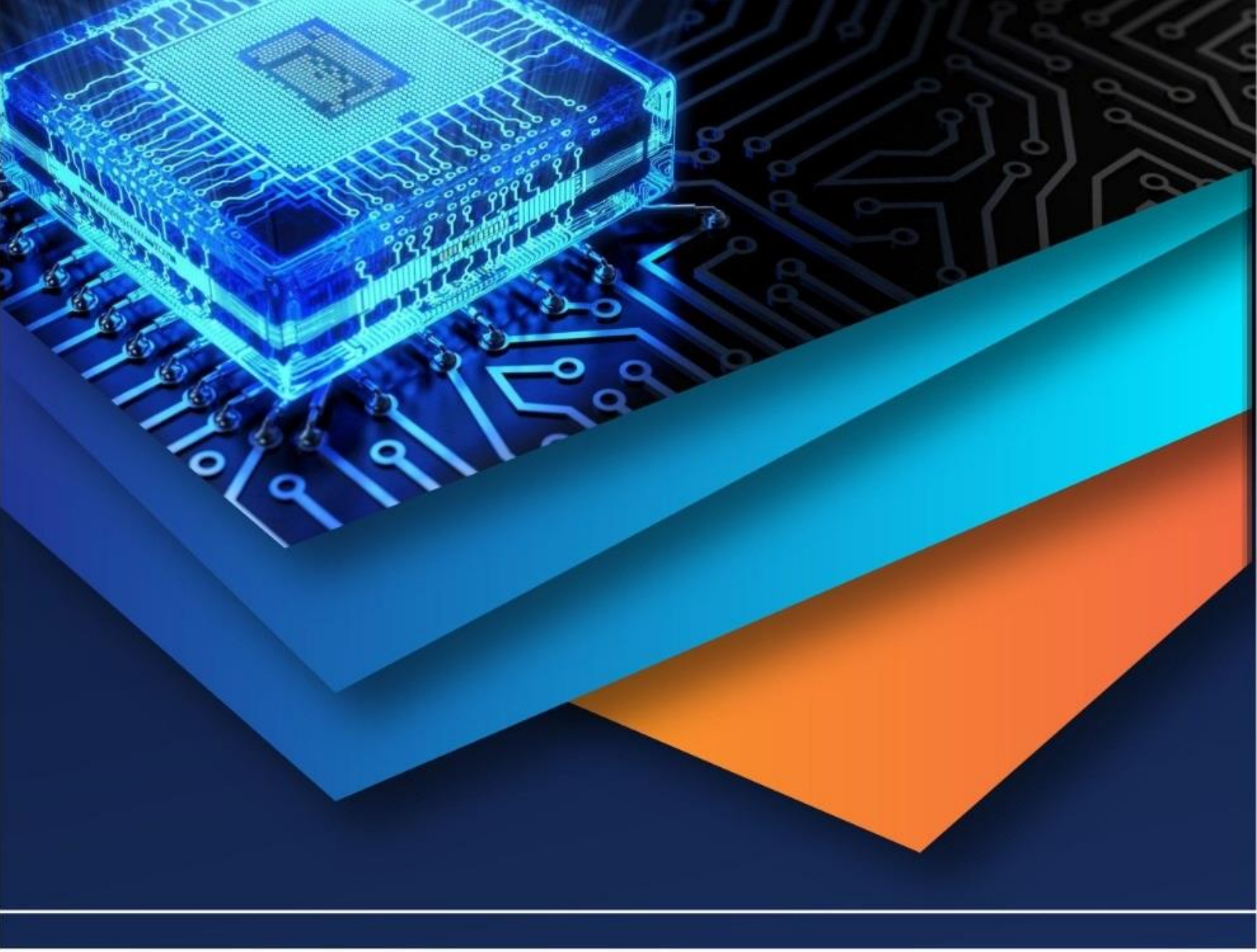
1) *Server:* A server is a computer or system that over a network distributes resources, data, services, or programmers to other computers known as clients. ..... Web servers, mail servers, and virtual servers are among the various types of servers.
2) *Hierarchical Index:* Your Data Frame would have two or more dimensions that can be used to describe each row if you use a hierarchical index. This produces a "Frozen List," which is a Pandas-specific construct for displaying a Data Frame's index mark.
3) *Query Request:* A query string is a part of a uniform resource locator (URL) that specifies values for parameters. A query string is a set of fields that a Web browser or other client program adds to a base URL.

V. CONCLUSIONS

We looked into cypher text search in the context of cloud storage in this paper. We investigate the problem of preserving semantic relationships between different plain documents and related encrypted documents, as well as a design method for improving semantic search efficiency. The MRSE-HCI architecture is also proposed to respond to the needs of data explosion, online information retrieval, and semantic search. At the same time, a verifiable mechanism is proposed to ensure that search results are accurate and complete. We also look at the search efficiency and protection in the context of two common threat models. The search performance, precision, and rank protection are all evaluated using an experimental platform. The results of the experiment show that the proposed architecture not only effectively solves the multi-keyword ranked search problem, but also improves search performance, rank protection, and document relevance.

REFERENCES

[1] D. Kornack and P. Rakic, "Cell Proliferation without Neurogenesis in Adult Primate Neocortex," Science, vol. 294, Dec. 2001, pp. 2127-2130, doi:10.1126/science.1065467.
[2] S. Grzonkowski, P. M. Corcoran, and T. Coughlin, " Security nalysis of authentication protocols for next-generation mobile and CE cloud services," in Proc. IEEE Int. Conf. Consumer Electron., 2011, Berlin, Germany, 2011, pp. 83–87.
[3] D. X. D. Song, D. Wagner, and A. Perrig, "Practical techniques for searches on encrypted data," in Proc. IEEE Symp. Security Priv., BERKELEY, CA, 2000, pp. 44–55.
[4] D. Boneh, G. Di Crescenzo, R. Ostrovsky, and G. Persiano, "Public key encryption with keyword search," in Proc. EUROCRYPT,Interlaken, SWITZERLAND, 2004, pp. 506–522.
[5] Y. C. Chang and M. Mitzenmacher, "Privacy preserving keyword searches on remote encrypted data," in Proc. 3rd Int. Conf. Applied Cryptography Netw. Security, New York, NY, 2005,pp. 442–455.
[6] R. Curtmola, J. Garay, S. Kamara, and R. Ostrovsky, "Searchable symmetric encryption: improved definitions and efficient constructions," in Proc. 13th ACM Conf. Comput. Commun. Security, Alexandria, Virginia, 2006, pp. 79–88.
[7] Yingling Liu ; Xindong Wu ; Xuegang Hua ; Jun Gao "Pattern matching with wildcards based on key character location" IRI '09. IEEE International Conference on 2009.
[8] S. Zerr, D. Olmedilla, W. Nejdl, and W. Siberski, "Zerber+r: Top-k retrieval from a confidential index," in Proc. of EDBT, 2009.
[9] N. Cao, C. Wang, M. Li, K. Ren, and W. Lou, "Privacy-preserving multikeyword ranked search over encrypted cloud data," IEEE Transactions on Parallel and Distributed Systems, vol. 25, no. 1, pp. 222–233, 2014.
[10] T. Jung, X. Mao, X. Li, S.-J. Tang, W. Gong, and L. Zhang, "Privacy preserving data aggregation without secure channel: multivariate polynomial evaluation," in Proceedings of INFOCOM. IEEE, 2013, pp.2634–2642.
[11] Y. Yang, H. Li, W. Liu, H. Yang, and M. Wen, "Secure dynamic searchable symmetric encryption with constant document update cost," in Proceedings of GLOBCOM. IEEE, 2014, to appear.
[12] C. Wang, N. Cao, J. Li, K. Ren, and W. Lou, "Secure ranked keyword search over encrypted cloud data," in Proceedings of ICDCS. IEEE,2010, pp. 253–262.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY