International Journal for Innovative
Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

IJIEMR Transactions, online available on 5th Jan 2023. Link :http://www.ijiemr.org/downloads.php?vol=Volume-12&issue=Issue 01

DOI: 10.48047/IJIEMR/V12/ISSUE 01/17

Title A SURVEY ON SECURE AND EFFICIENT RANK KEYWORD SEARCH IN CLOUD

Volume 12, ISSUE 01, Pages: 173-184

Paper Authors

Ramachandra Rao G, Dr Kavitha M S, Dr S.Karthik

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

A SURVEY ON SECURE AND EFFICIENT RANK KEYWORD SEARCH IN CLOUD

Ramachandra Rao G[1], Dr Kavitha M S[2], Dr S.Karthik[3]
[1]Research Scholar, Anna University, Chennai.
Email: grrsoft30@gmail.com
[2]Supervisor, Associate Professor, Department of CSE, SNS college of Technology, Coimbatore.
Email:drmskavitha@yahoo.com
[3]Professor, Department of CSE, SNS college of Technology, Coimbatore.
Email:profskarthik@gmail.com

Abstract

We develop a secured and efficient query rank (both single word or multi-word) search in cloud domain. All the methods presented are lies within the scope of cloud computing and it involves the goals required to design a secured ranked keyword search. The study provides a detailed overview of various models associated with cloud that involves the computation of access control search, single/multi-keyword search. The models selected are deeply associated with cloud and its associated limitations are presented in the study.

KEYWORDS: Rank search, Query, Keyword, Cloud

I Introduction

Cloud computing is the realization of the long-held vision of computing as a utility, in which users share a large pool of configurable computing resources and access their data and applications via the internet whenever and wherever they like [2]. The concept of cloud computing has been considered for quite some time. Users of cloud computing have the option of putting their data in a remote server. There are many advantages to this new computer architecture, including less time spent on storage management, the ability to access data from anywhere in the world, and the elimination of the need to regularly upgrade costly hardware, software, or personnel.

As cloud computing grows in popularity, it will likely become the norm for sensitive data such as emails, medical records, financial records, government documents, and so on to be housed in one easily accessible area . Since we must ensure the security of our customer personal information and financial transactions when sending data abroad, we must encrypt all such data prior to transmission. Due to the potential for a significant number of data files to be outsourced, decrypting encrypted data becomes a very difficult task. One more perk of cloud computing is that it lets data owners share their outsourced data with many users, each of whom can pull only the data files they need at any given time. In most cases, the users should use search terms to look for something. Due to the fact that users can select which files they want to retrieve using this search approach, plaintext search settings [3] have been using it frequently. When applied to encrypted cloud data, the traditional methods used for searching plaintext files are inadequate. Because of the need of privacy when dealing with keywords and the difficulty users have in conducting keyword searches when data is encrypted.

According to the work in [20], a normal CP-ABE plot that protects privacy might be used. The proposed system offers various benefits over the current system, including short code communications and private keys of a standard size. Additionally, decoding only requires four blending computations. This work fulfils a vast number of requests while maintaining specific security and confidentiality. According to the conventional approach, the recommended plot's safety is reduced to judgmental n-BDHE and DL suspicions. The proposed plot also maintains authority check without any protection spills.

Keyword searches can be performed on encrypted data with common searchable encryption methods without the requirement to decrypt the material first. But this approach just allows for a basic Boolean keyword search, and it doesn't consider the importance of the files. There are two potential issues when these solutions are instantly put into a massive cloud environment that is utilized for collaborative data outsourcing: One drawback of encrypted cloud data is that users who do not have extensive familiarity with the data must manually go through all of the files that are retrieved in response to a search in order to locate the ones that are most relevant to their needs. In today pay-as-you-go cloud environment, this could necessitate a great deal of post-processing work, and returning all files based on the presence or absence of the keyword would result in excessive network traffic. To rephrase, if you were to return all files based on whether or not they contained the term, you would do just that. One of the main issues with current searchable encryption techniques, especially in the context of cloud computing, is that there are not enough checks to make sure that data can be found correctly.

However, the relevance of files in response to a certain search query has been ranked quantitatively and alphabetically using a number of ranking methodologies by state-of-the-art members of the IR community. Although the IR community has long understood the significance of ranked search in the context of plaintext searching, this issue has yet to be resolved in the context of encrypted data search.

Secure Ranked Keyword Search

The system architecture must meet both of these requirements simultaneously to ensure the safety and effectiveness of the ranked keyword search verification process.

- **Efficiency**: The simplicity with which data owners can create the verification data is a crucial component of efficiency. Furthermore, the cloud server should not demand unreasonable fees in order to return the verification data. Information seekers can easily confirm the accuracy of the search result.
- **Security**: Data owner identities and the confidentiality of their verification data used to validate cloud-stored information must be safeguarded by the proposed solution.
- **Detectability**: The cloud server untrusted behaviour can be prevented if the recommended method is found. The approach should detect any dishonest activity on the cloud server with a high degree of certainty as soon as it occurs.
- **Search Efficiency**: Our top concern is improving search efficiency; therefore, we are investigating both a tree-based index structure and a fast search algorithm. Cloud servers will compile encrypted indexes without being able to decrypt the sensitive information they represent. Users who are authorized to access the data only need to encrypt their search terms once.
- **Ranked multi-keyword**. It was determined that multi-keyword search was superior to multi-owner search. The technique that has been presented ought should be able to conduct out multi-keyword searches on files that have been encrypted using keys that are unique to each data owner. In addition, the cloud service must be able to deliver the best k results based on the order in which the

search results from various data owners are rated.

- **Data owner scalability**. Information owners have the option of using an infinite data storage capacity. The proposed solution should provide for plug-and-play scalability for the data owners, allowing new data owners to join the system without disrupting the experience of existing data owners or consumers.
- **Data user revocation**. Valid searches would be restricted to users who have been granted access to the data under the proposed structure. If a user loses permission to access the data, they will no longer be able to do a successful search for the encrypted information kept in the cloud.

II. RANKED SEARCH METHODS

When it comes to symmetric searchable encryption (SSE), Song et al. [4] were the pioneers who proposed a solution with a search time that scales linearly with the amount of the dataset. Goh came up with the idea for formal security standards for SSE and built a methodology off of Bloom taxonomy.

Two techniques, SSE-1 and SSE-2, were proposed by [5] for achieving the fastest search time. Their SSE-1 and SSE-2 systems are secure against selected-keyword attacks (CKA1) and adaptive CKA2 respectively. The earliest research projects only used a single keyword for their boolean search algorithms, making their complexity relatively low. Since then, many contributions have been made using different danger models to develop advanced search tools including keyword searches, multi-keyword boolean searches, ranked searches, etc [6].

Users of a boolean multi-keyword search can find relevant papers by inputting a string of keywords. In order for a web page to be returned in a conjunctive keyword search [7], it must match every phrase in the query. The pages that contain any of the query terms are obtained via disjunctive keyword search methods.

On the other hand, conjunctive keyword search techniques would only return results that contained all of the query terms. It has been proposed that conjunctive and disjunctive searches be supported by predicate search algorithms. None of these methods for doing a multi-keyword search provides any more reliable results than a simple keyword search would have yielded.

The most pertinent results to your query can be easily located using a ranked search. Limit your responses to the top k most relevant documents to reduce your network load. Prior studies [8] only allowed for a single term to be utilized in the ranked search, despite their use of order-preserving methods. Nonetheless, ranked search functionality was successfully developed.

Cao et al. created the first multi-keyword ranked search system that protected user anonymity. Coordinate matching ranks documents according on how closely their content matches a query. However, this method lacks precision since it disregards the relative importance of the various keywords. Moreover, the search efficiency offered by the approach grows linearly with the cardinality of the document collection being searched.

In order to bypass this problem, Sun et al. [9] designed a method that allows for the secure execution of numerous keyword searches at the same time and that also takes into consideration rankings defined by the degree to which search results are identical. Using the vector space model, the authors created a hierarchical index, with ranks obtained using the cosine measure and the TF IDF. Sun and his team search method is faster than linear search, but it is less precise.

In order to safely group together documents that are relevant to one another during a multi-keyword search, Orencik et al. [10] proposed using the local sensitive hash, or LSH, function. In general, the LSH technique performs well for similar queries, although it cannot provide precise ranking data. The issue of secure ranked search with numerous owners was the focus of Zhang et al. [11]. These efforts do not currently support dynamic operations.
An individual who stores data in the cloud may need to make revisions to previously uploaded documents. In addition, there are a variety of methods for implementing dynamic searchable encryption. In the study conducted by Song et al. [4], each document is indexed separately and is treated like a string of words of a constant length. Although it is inefficient, this technique allows for simple updating procedures to be carried out.

In order to create a more precise index, Goh proposed using keywords to create a sub-index (Bloom filter) for each page. Implementing the dynamic operations is a breeze when both the Bloom filter and the associated document are maintained up to date. Goh method is efficient, but it is time-consuming and prone to false positives because of the linear search it employs. Kamara et al developed an encrypted inverted index for efficient and rapid analysis of real-time data.

Implementing this concept, however, will be difficult. Then, as a follow-up enhancement, Kamara and Papamanthou introduced a brand-new search strategy that makes use of a tree-based index. This procedure can handle real-time modifications to the document data that is part of the leaf nodes. Still, their method is limited to single-term Boolean searches.

For simple keyword queries, Wang et al. proposed a safe ranked keyword search strategy. Using this method, the k most relevant files from the search is obtained. Users can narrow the scope of their search results by entering numerous keywords into the multi-keyword ranked search box.

The multi-keyword ranked search technique over encrypted cloud data (MRSE), put forth by Cao et al. in [12], uses the inner product, the product of the file vectors and the query vectors, to rank the documents. However, they disregard the significance of each sentence on its own. Chuah and Hu devised multi-keyword fuzzy search algorithms to handle user errors and format incompatibilities. The suggestion for a secure ranked multikeyword search strategy in a multi-owner paradigm by Zhang et al. allowed the cloud server to do a multikeyword search without accessing private data (PRMSM). They also gave the data owner complete control over any changes made to the encryption key. The researcher PRMSM allowed for both of these enhancements. On the other hand, query efficiency is usually not a major concern in these approaches. Evaluation of search performance is a crucial step in bettering the user experience.

Ranking search results greatly enhances system usability and auxiliary ensures the correctness of file retrieval by providing search result relevance ratings rather than broadcasting undifferentiated results. [19]. In order to build a protected searchable index and progress a one-to-many order-preserving mapping technique, the work studies the relevance score statistical measure methodology from information retrieval.

Anonymity during a search for many keywords in a text document was addressed by Sun et al. The MD-algorithm was used to create a tree-based index structure for the vector index, which expedited searches at the expense of precision. The index size was shrunk, which allowed for this to happen. The greedy depth-first search (GDFS) approach was proposed by Xia et al. [13] in order to enhance search performance. Unfortunately, the prospect of there being many data owners is not considered in these investigations.

Dong et al. [14] explored a real-world scenario in which a large number of users transfer data across the medium of an untrusted third party. The authors devised a novel technique based on proxy cryptography to make encrypted material searchable by many users simultaneously. As opposed to competing searchable encryption approaches, their approach allows many users to access and edit the same dataset in real time. As opposed to the old ways, this one was a huge success. All of the information provided indicates that their plan is completely safe and sound.

Mylar, introduced by Popa et al.is a platform that combines system techniques with novel cryptographic primitives to simplify data transfer, computation over encrypted data, and application code verification. The goal was to simplify these actions for the end user. We tested six alternative apps and determined that Mylar was the most effective multi-user web tool for facilitating information sharing and dissemination.

Near-duplicate detection (NDD) was using encrypted in-network storage that allowed for a number of users and searchable encryption keys. Unfortunately, the issue of multi-keyword ranked searches is not well-served by those solutions. Due to the fact that this is not the case, we cannot employ their methods to fix the issue at hand.

The cloud has the capacity to join encrypted indexes from different data owners without knowing the contents of the indexes thanks to a multi-source encrypted indexes merge (MEIM) technique developed by Yao et al. [34] PHRs were analyzed, and file-based searches were preferred over vector-based ones because the latter required more work to be done to evaluate each attribute.

Multi-keyword search

The capacity to conduct a search query over encrypted content using several terms is a significant feature of multi-keyword searchable encryption. Multi-keyword searchable encryption has drawn a lot of research attention. because of its fundamental importance to the study of cryptography. Buyrukbilen and Bakiras [15] described a technique that enables ranked results to be supplied from queries that employ multiple keywords. The similarity relevance multi-keyword search was examined by Yu et al. [16]. They improved the search by encrypting the top k results using a methodology that works in vector space.

Zhang et al. [17] presented a strategy for multi-keyword ranked search in a setting with a large number of users. Even if a user does not have access to the keys that the underlying data owners use to encrypt their documents and keywords, they are still able to perform a search on the database. Documents and search terms can be encrypted with the owner private keys. To find the most precise search results, the authors suggested employing a function that preserves additive order. But dynamic operations support is missing from these advancements.

Li et al. [18] developed a ciphertext policy that uses attributes to encrypt messages. Their technique is equally effective at preventing collusion and protecting the privacy of sensitive documents.

III. Limitations

Research on secure keyword search over encrypted data as a method to aid efficient data recovery from encrypted data has increased in recent years. This is possible because it is possible to execute a safe keyword search over encrypted data without jeopardizing the security of the data. Each step of these processes is based on the assumption that the cloud server has a inquisitive yet honest personality. Unfortunately, when used in real-world applications, the cloud server could be hacked and behave dishonestly. If a cloud server was hacked in any of the following ways, users might get skewed search results:

1) The cloud server might potentially fake search results, for one. Due to the pay as you go nature of cloud storage, the provider may prioritize some ads over others or return unexpectedly big files in order to maximize revenue.
2) To prevent generating performance bottlenecks during peak times, the cloud server may only communicate a subset of

the overall search results. The credibility of search engine results has been investigated in a number of research.

However,These methods cannot be utilised to independently verify the top-k ranked search results due to the enormous number of data owners available in a cloud computing environment. We give two examples to show that this is so.

1) All existing systems are founded on the same core idea, which holds that data owners have advanced knowledge of how their information will appear in search results. However, in real-world applications, there are often a large number of individual data owners, each of whom is only aware of a subset of the overall order. Traditional techniques of search result approval are insufficient for these data owners because they lack information about the overall order.

2) Only a minority of data owners will have files that fulfill a top-k ranked keyword search. This is due to the fact that there are numerous varieties of information. To verify the accuracy of a sizable data set using conventional methods, you must return a substantial amount of information.

When there are several data owners, it is far more difficult to design an effective system than when there is only one user. We typically construct a tree-like index structure for each data owner encrypted data. This allows us to protect user anonymity while simultaneously making searches more efficient. To allow the cloud to search each index, data consumers must first build a trapdoor for each data owner.

For each given query condition, this is essential. Since there is a linear correlation between the number of trapdoors and the number of data owners, it is not surprising that this is inefficient. Forcing all data owners to use the same key when encrypting files is a simple solution to this limitation. But if even one of the owners is tainted, the whole enterprise can ruin.

Forward security is a new concern that arises when users search encrypted data using keywords. In order to guarantee user continued access to the data, the owner of the data will typically make changes to the files containing the data. Existing dynamic ranked keyword search algorithms that utilize the aforementioned index structures present a significant update overhead when performing dynamic activities such as altering, removing, or adding files.

Because of its accurate nature, the inverted index is not recommended for usage with real-time changes. If you're working with a sizable dataset, you should know that using a tree-based index will require frequent updates to a significant number of intermediate nodes. Most existing methods for ranking search results dynamically leave systems open to attacks like file injection.

IV. Conclusion

We develop a secured and efficient query rank (both single word or multi-word) search in cloud domain. All the methods presented are lies within the scope of cloud computing and it involves the goals required to design a secured ranked keyword search. The study provides a detailed overview of various models associated with cloud that involves the computation of access control search, single/multi-keyword search. The models selected are deeply associated with cloud and its associated limitations are presented in the study.

V. References

[1] Guan, Z., Du, X ,Liu, X.Wu, L. Abedin,& Guizani, M. (2019, May). Achieving secure and efficient cloud search services: Cross-lingual multi-keyword rank search over encrypted cloud data. In *ICC 2019-2019 IEEE International Conference on Communications (ICC)* (pp. 1-6). IEEE.

[2] Pham, H., Woodworth, J., & Amini Salehi, M. (2019). Survey on secure search over encrypted data on the cloud. *Concurrency and Computation: Practice and Experience*, *31*(17), e5284.

[3]M.Karthick, Naresh, R., Sayeekumar, G. M., & Supraja, P. (2019). Attribute-based hierarchical file encryption for efficient retrieval of files by DV index tree from cloud using crossover genetic algorithm. *Soft Computing*, *23*(8), 2561-2574.

[4] Song, D. X., Wagner, D., & Perrig, A. (2000, May). Practical techniques for searches on encrypted data. In *Proceeding 2000 IEEE symposium on security and privacy. S&P 2000* (pp. 44-55). IEEE.

[5] Curtmola, R., Garay, J., Kamara, S., & Ostrovsky, R. (2011). Searchable symmetric encryption: improved definitions and efficient constructions. *Journal of Computer Security*, *19*(5), 895-934.

[6] Swaminathan, A., Mao, G. M., Gou, H., Varna, A. L., He, S., ... & Oard, D. W. (2007, October). Confidentiality-preserving rank-ordered search. In *Proceedings of the 2007 ACM workshop on Storage security and survivability* (pp. 7-12).

[7]Golle, P., Staddon, J., & Waters, B. (2004, June). Secure conjunctive keyword search over encrypted data. In International conference on applied cryptography and network security (pp. 31-45). Springer, Berlin, Heidelberg.

[8] Nejdl, W, Zerr, S, Olmedilla, D., & Siberski, W(2009, March). Zerber r: Top-k retrieval from a confidential index. In *Proceedings of the 12th International conference on extending database*

technology: advances in database technology (pp. 439-449).

[9]Sun, W., Wang, B., Cao, N., Li, M., Lou, W., Hou, Y. T., & Li, H. (2013, May). Privacy-preserving multi-keyword text search in the cloud supporting similarity-based ranking. In *Proceedings of the 8th ACM SIGSAC symposium on Information, computer and communications security* (pp. 71-82).

[10] Kantarcioglu, M., Orencik, C., & Savas, E. (2013, June). A practical and secure multi-keyword search method over encrypted cloud data. In *2013 IEEE Sixth International Conference on Cloud Computing* (pp. 390-397). IEEE.

[11] Zhang, W., Zhou, S, Xiao, S., Lin, Y., Zhou, T., &. (2014, June). Secure ranked multi-keyword search for multiple data owners in cloud computing. In *2014 44th Annual IEEE/IFIP International Conference on Dependable Systems and Networks* (pp. 276-286). IEEE.

[12]N., Wang Cao, , C., Li, M., Ren, K., & Lou, W. (2013). Privacy-preserving multi-keyword ranked search over encrypted cloud data. *IEEE Transactions on parallel and distributed systems*, *25*(1), 222-233.

[13]Fu, Z., Wu, X., Guan, C., Sun, X., & Ren, K. (2016). Toward efficient multi-keyword fuzzy search over encrypted outsourced data with accuracy improvement. *IEEE Transactions on Information Forensics and Security*, *11*(12), 2706-2716.

[14]Dong, C., Russello, G., & Dulay, N. (2011). Shared and searchable encrypted data for untrusted servers. *Journal of Computer Security*, *19*(3), 367-397.

[15]Bakiras, S. &Buyrukbilen, S., (2013, November). Privacy-preserving ranked search on public-key encrypted data. In *2013 IEEE 10th International Conference on High Performance Computing and Communications & 2013 IEEE International Conference on Embedded and Ubiquitous Computing* (pp. 165-174). IEEE.

[16] Y., Xue,Yu, J., Lu, P., Zhu, G., & Li, M. (2013). Toward secure multi keyword top-k retrieval over encrypted cloud data. IEEE transactions on dependable and secure computing, 10(4), 239-250.

[17] Lin, Y., Xiao ,Zhang, W.,S., Wu, J., & Zhou, S. (2015). Privacy preserving ranked multi-keyword search for multiple data owners in cloud computing. *IEEE Transactions on Computers*, *65*(5), 1566-1577.

[18]Y., Luan,Li, H., Liu, D., Dai,T. H., & Shen, X. S. (2014). Enabling efficient multi-keyword ranked search over encrypted mobile cloud data through blind storage. *IEEE Transactions on Emerging Topics in Computing*, *3*(1), 127-138.

[19] K. K. Kumar, S. G. B. Kumar, S. G. R. Rao and S. S. J. Sydulu, "Safe and high secured ranked keyword search over an outsourced cloud data," 2017 International Conference on Inventive Computing and Informatics (ICICI), 2017, pp. 20-25, doi: 10.1109/ICICI.2017.8365348.

[20] R. K. Yarava, G. R. C. Rao, Y. Garapati, G. C. Babu and S. D. V. Prasad, "Analysis on the Development of Cloud

Security using Privacy Attribute Data Sharing," 2022 First International Conference on Electrical, Electronics, Information and Communication Technologies (ICEEICT), 2022, pp. 1-5, doi:10.1109/ICEEICT53079.2022.9768608