Mathematical Statistician and Engineering Applications
ISSN: 2326-9865
DOI: https://doi.org/10.17762/msea.v70i1.2305

Enabling Fine-Grained Multi-Keyword Search Supporting Classified Sub-Dictionaries Over Encrypted Cloud Data

Satvik Vats

Asst. Professor, Department of Computer Science, Graphic Era Hill University, Dehradun, Uttarakhand India 248002

Article Info
Page Number: 243-252
Publication Issue:
Vol. 70 No. 1 (2021)

Abstract
Huge numbers of papers are transferred to the cloud for easier access and lower administration costs as cloud computing gains popularity. Individuals can utilize cloud computing to store their data on distant servers and grant public users access to the data via the cloud servers. Outsourced data are normally encrypted before being posted to the clod since they are likely to include sensitive privacy information. The difficulty of sifting through the encrypted data, however, severely restricts the utility of the outsourced data. In this research, we introduce the fine-grained multi-keyword search algorithms over encrypted clod data to overcome this problem. We made three original contributions. In order to provide exact keyword search and a tailored user experience, we initially present the relevance source plus preference factor upon keywords. Furthermore, we create a useful and very effective multi-keyword search strategy. The suggested system can enable complex logic searches using mixed "AND," "OR," and "NO" keyword operations. Thirdly, in order to improve the efficiency of index building and query generation, we also use the classified sub-dictionaries technique. Finally, we examine the security of the suggested strategies with regard to of document confidentiality, index alongside trapdoor privacy protection, as well as trapdoor unlinkability. We confirm that the suggested method performs better when it comes to of functionality, query complexity, yet efficiency than the ones currently in use plus can reach the same security level through rigorous tests on a real-world dataset. We confirm that the suggested method performs better with regard to functionality, query complexity, and efficiency than the existing ones and can reach the same security level through rigorous tests on a real-world dataset.

Article History
Article Received: 25 January 2021
Revised: 24 February 2021
Accepted: 15 March 2021

1. Introduction

In order to allow effective multi-keyword ranked search in cloud computing systems, this study suggests a lightweight search strategy. For effective multi-keyword ranked search, it employs a polynomial function to conceal the encrypted keywords and search patterns, and a privacy-preserving approach to safeguard the confidentiality of the searched multi-keywords. The experiment's findings show that the suggested technique can provide a very effective cloud computing solution for encrypted multi-keyword ranking searches. Cloud computing is expanding in popularity and significance in our everyday lives as users may now remotely outsource their data to the cloud and take use of the on-demand services offered by the shared computer resources. Users gain various advantages from cloud computing, including flexible data access and reduced storage load.

To guarantee that the data is private when transferred to the cloud, sensitive data must first be encrypted by the data owner. It is necessary to provide efficient ranked multiple keyword

search, which accepts a variety of input terms and simultaneously accomplishes high efficiency in user search behaviors. It is crucial to do effective keyword searches on encrypted data. There have recently been approaches for cloud computing that use several keywords to search, but there is still much space for improvement in terms of how effective these methods are. We use polynomial functions in this research to do multi-keyword searches on encrypted data stored in clouds.

We provide a multi-keyword search strategy that takes use of the quantity of query keywords included in the document index to assess how similar the query and the content are. This method does away with the pre-defined binary index vector used in the current multiple keyword search strategy also allows for efficient index updating, thereby enabling it to scale to a large number of search terms. We integrate the present safe inner product methodology with our polynomial function-based method, which is taken from the secure k-nearest neighbour (kNN) technology, to counter an attacker with strong computational resources. The outcomes demonstrate how successful and efficient our strategy is for carrying out ranked multiple keyword searches. Datacenters are a key component of the main cloud infrastructure providers like Amazon, Google, as well as Microsoft Azure. These companies offer utility computing services to software service providers, who then offer application services to end users through the Internet.

The suggested secure inner product computation process, which is developed from a secure-nearest neighbour approach, along with the modifications on it to meet multiple privacy criteria under the known-background threat model are the most crucial features in this work. More and more data is being outsourced by data owners to cloud service providers, who they don't completely trust, in order to take use of the computational and storage capabilities provided by vendors of cloud infrastructure. Prior to uploading sensitive data to the cloud, it must be encrypted, to preserve users' privacy, and certain data must be shared with reputable partners. To accomplish this feature, the query graph is also expressed as a binary vector, and each data graph has a binary vector assigned to it as a sub-index. The inner product of the query vector plus the data vector may be able to exactly quantify the amount of query features included in the data graph, allowing one to eliminate negative data graphs that are devoid of the query graph.We introduce a safe inner product calculation approach derived from the secure nearest neighbour method, in order to tackle the difficulty of providing graph semantics without violating privacy, and then demonstrate our improvements on it.

This essay offers definitions of dependability, a general term that encompasses qualities like dependability, availability, safety, integrity, and maintainability, among others. Confidentiality issues are raised by security, and the initial definitions are then discussed and expanded upon by new ones. The goal is to record a minimal level of agreement on concepts across a range of disciplines to encourage productive technical discussions. The metrics of reliability and security are a significant topic for which a consensus has not yet developed. The goal of the study is to tie together the interconnected threads of security and reliability.

It provides a fundamental set of concepts for the full spectrum of computing and communication systems, from standalone logic gates to computer networks with human

operators and consumers. The ability of a system to produce behaviour is due to its structure. A system is made up of various parts connected so they can work together. The collection of the (external) states of a system's atomic constituents makes up its overall state. A system provides a service when it behaves in a way that the user(s) perceive.

The provider's service interface is the area of the system border where services are delivered. It is possible to think of function and service as being made up of function and service components.

2. Literature Survey

With the help of a geographically dispersed cloud service architecture, mobile cloud computing is a viable method for moving modules for data plus computing from different gadgets. The goal of this work is to balance the computing loads across several cloud domains by proposing a service decision-making mechanism for inter-domain service transfer. By reducing the amount of service rejections, the system aims to maximise benefits for both the cloud system and the consumers. Numerous simulation outcomes demonstrate that the suggested decision making system, when compared to the greedy approach, can significantly increase system rewards and decrease service interruptions. The goal of the suggested resource allocation decision model is to maximise resource distribution among mobile cloud service domains and improve the QoE for mobile users. According to a simulation-based study, the service rejection probability with inter-domain service transfer is 20% lower than with the greedy approach. In order to determine if the mobile service would need to be approved, refused, or moved, the system reward model is crucial for linking the home cloud domain. This study examines the effectiveness of a proposed SMDP-based inter-domain resource allocation plan while taking system gain, resource costs for computation, and communication costs into account. Based on the suggested SMDP model, it calculates the likelihood of new services and inter-domain transfer services being dropped. In order to maximise system benefits, future research will examine the best system resources [1].

This paper suggests a number of protocols that ensure data privacy while keeping each participant's communication and computing complexity to a minimal constant. It is predicated that all channels are vulnerable to eavesdropping attempts and that the discrete logarithm issue is computationally challenging if both the related integer numbers and the ordering of the integer groups are big prime values. It also evaluates how well our protocol performs in comparison to other multi-party computation systems already in use that Benet al. (Fair Play MP) have put into place. In this research, a unique safe product and sum computation protocol is presented that effectively implements a multivariate polynomial evaluation while protecting privacy without using secure communication channels. The approach, however, reveals every polynomial product portion, which provides attackers with further information. Future research will concentrate on developing privacy-preserving data release mechanisms and minimising information leakage during computation and transmission [2].

This paper introduces A newly developed searchable encryption method containing updates that only show the access pattern, asymptotically fast search times, linearity, yet a very tiny and asymptotically efficient index size, along with implementation that is possible without

storing anything on the client (aside from the key). The building method is very effective for cloud storage since it is depending upon the innovative notion of inferring the index for effective access from the access pattern itself. The number of keywords is linear, it is asymptotically optimum, and client storage (apart from the key) is not required. In addition to remaining semantically safe encrypted, 84% of all keywords were never looked for. This method presents a fresh, workable substitute for secure cloud storage that is more secure while yet being almost as effective as deterministic encryption. In terms of absolute performance measures, it is quite effective [3].

This study develops a set of privacy regulations and suggests Secured Multi-keyword Search (SMS) over encrypted cloud data (ECD). It makes use of coordinate matching to determine whether the search query and the data are comparable, using inner data correspondence to formally quantify similarity assessment. When an unauthorised user tries to access the data, an alarm system creates alerts, and the Ranked result displays the top k retrieval results. A new computer approach called cloud computing offers greater security in systems with multiple users. By arranging files that match in a ranked order according to relevance criteria, ranked search enhances system usability. To provide just expected data, asymmetric encryption with grading of the query data is offered. Data owners have access to a huge user base who can access their outsourced data [4].

This paper proposes a fix for the issue of ranking keywords securely while still being effective through encrypted cloud data. It makes effective use of the currently available cryptographic primitive, order-preserving symmetric encryption (OPSE), which employs the most up-to-date searchable symmetric encryption (SSE) security specification. Numerous experimental findings support the effectiveness of the suggested approach, bringing it one step closer to being implemented in practise for the provision of privacy-preserving data hosting services in cloud computing. This paper proposes a ranked searchable symmetric encryption scheme that uses statistical measure and text-mining to embed weight information of each file before outsourcing the encrypted file collection. To protect sensitive weight information, OPSE is modified to provide efficient ranked search functionalities. The ability to conduct limited searches, data owners can outsource their data while still maintaining searchable encryption. By include file ID as an additional seed in the process of choosing the final ciphertext, the one-to-many order-preserving mapping is modified from the original OPSE. This adaption performs well when there are few plaintext duplicates, but if there are numerous copies, it may still show skewness or peaky characteristics [5].

3. Proposed System

The fine-grained multi-keyword search (FMS) problem for encrypted cloud data has been examined in this research, and two FMS schemes have been suggested. The relevance scores and term choice factor are both included in the FMS 1 in order to improve both the user experience and more accurate search. With useful functions, the FMS 2 delivers secure and effective search. In addition, we have suggested the FM-SCS, which is an improved system supporting classified sub-dictionaries, to boost effectiveness. We evaluate the protection of the

suggested approach in regards to document confidentiality, privacy protection for the index as well as trapdoor, including trapdoor unlink ability.

We confirm through extensive testing on a real-world dataset, the suggested method's performance can be brought up to parity in security with the existing ones and higher with respect to functionality, query complexity, as well as efficiency. People can keep their data on distant servers in this hypothetical system and grant public users access to that data via cloud servers. Data from outsourced sources are frequently encrypted before being transferred to the cloud since they are likely to include sensitive privacy information. Prior to sending private data to cloud services, encryption is frequently required. The difficulty of finding through encrypted data, however, would dramatically reduce the utility of the data. For easy and dependable the relevant search users' ability to obtain data, the data owner outsources her data to the cloud. Utilizing symmetric encryption, the original data is scrambled by the data owner to safeguard data privacy. For each document that is outsourced, the data owner creates a set of keywords to increase search efficiency. A secret key and the keywords are then used to construct the matching index. The data owner then uploads the encrypted files and associated indexes to the cloud, as well as sending the symmetric key and secret key to search users. According to the outsourced data, the data owner produces several keywords. The cloud server then stores these keywords once they have been encrypted.

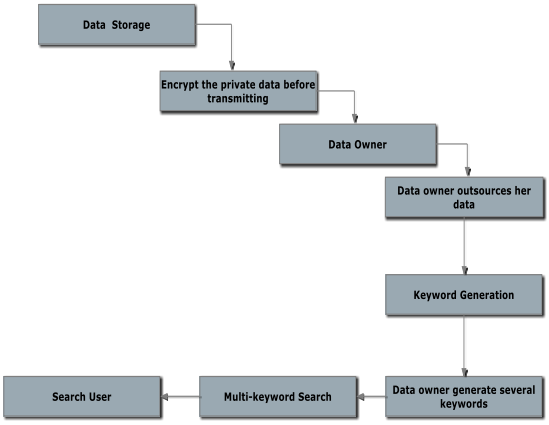

Fig 1: System Architecture

If a search user requires access to the data that has been outsourced, it can choose a few pertinent keywords and send the cypher text of those keywords to the cloud server. The cloud server then compares the outsourced encrypted keywords using the ciphertext, and finally provides the search user with matched results. a multi-keyword text search system that takes keyword relevance scores into account while employs a multidimensional tree strategy to provide effective search results. In order to ensure maximum security, provide a multi-keyword top-k retrieval strategy that encrypts the index/trapdoor using completely holomorphic encryption. I suggest the use of a multi-keyword ranked search that uses a coordinate machine as the keyword matching criteria. The following three steps are taken by a search user to query the outsourced documents from the cloud server.

The secret key and symmetric key are first given to the search user by the data owner. Second, the search user creates a trapdoor based on the search terms and transmits it to the cloud server using the secret key. Then, using the symmetric key, she retrieves the matching document collection from the cloud server. A user must first get authorisation before using a keyword search tool to access data. This means that only the authorised individual may retrieve the data. A randomly generated key is used to give authorisation. Each user's copy of the key is distinct. This key should be kept in mind when conducting the search. The user can input key words, which are the combination of one or more keywords. That is AND, OR, BOTH, and he receives a search result that is ranked. In the current system, we provide a conjunctive keyword to search for and retrieve the data. Here, a sophisticated tire-tree is employed to store this group of terms and search them individually. Also defined are the terms AND, OR, and BOTH. For "fuzzy keyword construction," we employ the "wild card" approach and the "gramme based" method. The implementation of the term conjunction in each of these techniques results in a highly effective ranking result.

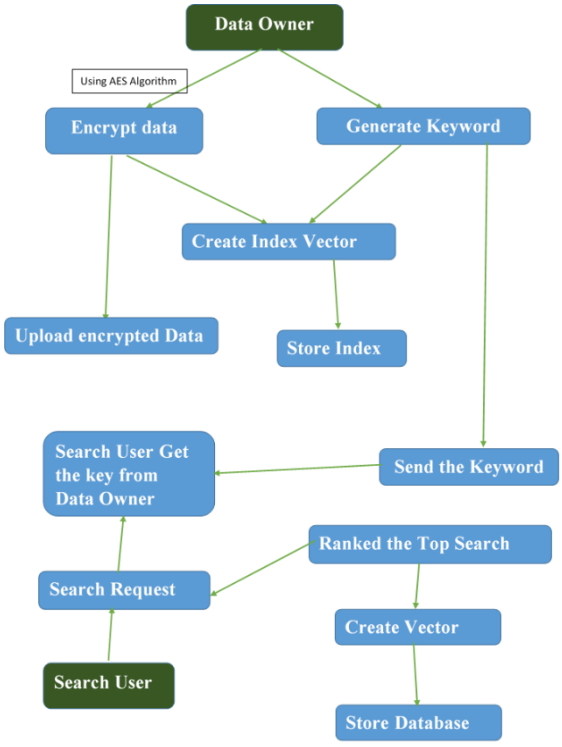

Fig 2: Flow Diagram

Likewise, we consider search users to be reliable parties who share the same secret key and symmetric key. With the data owner, search users already have mutual trust. For the sake of simplicity, The safe transmission of the symmetric key as well as the secret key amongst the data owner & search users is not taken into consideration; instead, it may be done with the help of common authentication and secure channel setup protocols in accordance with an earlier security context that both the data owner and search users have access to. Additionally, in order to sharpen the focus of our presentations, we do not regard the following issues as distinct ones: the access control issue involving the management of users' decryption privileges and the data

collection issue involving adding fresh documents, editing current papers, as well as removing current documents.

In our FMSCS schemes, simply the data owners who use the corresponding sub-dictionaries are required to update their indexes when it comes to changes to sub-dictionaries or the addition of new sub-dictionaries; the majority of other data owners are not required to perform any update operations. These dictionary updating procedures are incredibly light-weight. Furthermore, even though all indexes do not need to be generated from scratch, our scheme can still be more effective because all indexes must be subjected to the corresponding extended operations. In contrast, only the indices of partial data owners need to be extended in our schemes. The following benefits of the suggested strategy are listed:

- It can accomplish document confidentiality, index and trapdoor privacy protection, and trapdoor unlinking capability.
- It accomplishes practical functionality while providing secure and effective search.
- There is no need to recreate every index, but it is important to do the relevant extended operations on every index.
- The indexes of partial data owners are the only ones that this approach needs to extend.

4. Results

Document outsourcing to the cloud has been made possible by cloud computing, however this has hampered the utility of the papers because it is impossible to search through the encrypted data. For the purpose to address this issue, we construct fine-grained multi-keyword search algorithms over encrypted clod data in this study. It introduces security analysis, a realistic and effective multi-keyword search strategy, relevance source and preference variables, and a technique for using classified subdictionaries. The effectiveness of the suggested method is verified by tests on the real-world dataset. In order to increase the effectiveness of ranked keyword search algorithms, this study suggests a searching technique.

Large amounts of unstructured text are categorised and searched conceptually using a combination of concept-based and keyword-based searching algorithms. The suggested search strategies significantly increase the effectiveness of ranked keyword search, according to experimental relevance score analysis results. Two fine-grained multi-keyword search (FMS) methods have also been put forth, one of which takes relevance scores and keyword preference factors into account to improve search accuracy and user experiences.

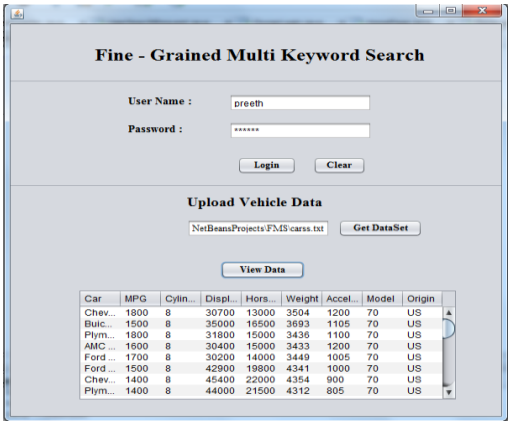

Fig 3: Upload Vehicle Data

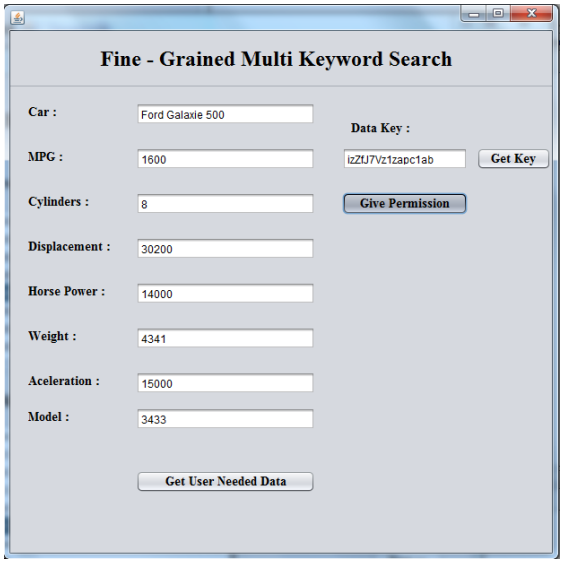

Fig 4: Vehicle Details

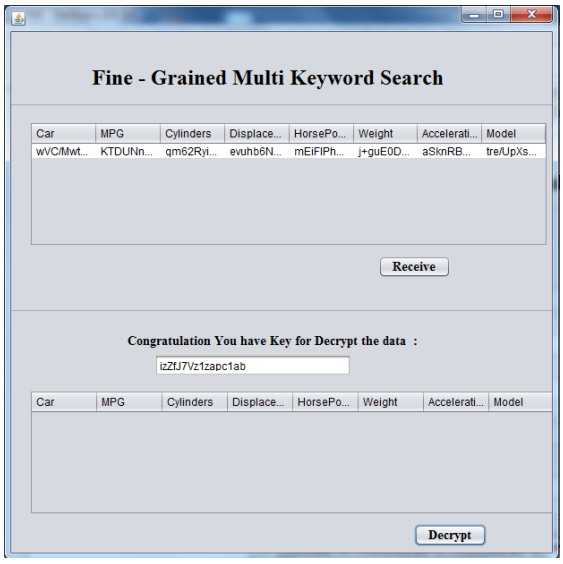

Fig 5: Data Encryption

5. Conclusion

We suggested a search strategy in this research to increase the effectiveness of ranked keyword search algorithms. We introduced the current architecture for searchable encryption, which is incredibly ineffective for achieving effective ranked search. Furthermore, we suggested combining concept-based and keyword-based search strategies. Large quantities of unstructured text may be conceptually categorised and searched using these kinds of approaches. Compared to conventional searches, this type of searching approach is more efficient and dependable and is more likely to provide results that are pertinent. The recommended search strategies significantly increase the effectiveness of ranked keyword search, according to the findings of our experimental relevance score analysis. In this article, we looked at the fine-grained multi-keyword search (FMS) problem with encrypted cloud data and offered two FMS schemes. The relevance scores and term choice factor are both included in the FMS 1 in order to improve both the user experience and more accurate search. With useful functions, the FMS 2 delivers secure and effective search. In addition, we have suggested the FM-SCS, which is an improved system supporting classified sub-dictionaries, to boost effectiveness. We evaluate the security of the suggested approach with respect to document confidentiality, privacy protection for the index and trapdoor, and trapdoor unlinkability.

Reference

1. H. Liang, L. X. Cai, D. Huang, X. Shen, and D. Peng, “An smdp-based service model for interdomain resource allocation in mobile cloud networks,”IEEE Transactions on Vehicular Technology, vol. 61, no. 5, pp. 2222–2232, 2012.
2. M. M. Mahmoud and X. Shen, “A cloud-based scheme for protecting source-location privacy against hotspot-locating attack in wireless sensor networks,” IEEE Transactions on Parallel and Distributed Systems, vol. 23, no. 10, pp. 1805–1818, 2012.
3. Q. Shen, X. Liang, X. Shen, X. Lin, and H. Luo, “Exploiting geo-distributed clouds for e-health monitoring system with minimum service delay and privacy preservation,”IEEE Journal of Biomedical and Health Informatics, vol. 18, no. 2, pp. 430–439, 2014.
4. T. Jung, X. Mao, X. Li, S.-J. Tang, W. Gong, and L. Zhang, “Privacy-preserving data aggregation without secure channel: multivariate poly-nomial evaluation,” inProceedings of INFOCOM. IEEE, 2013, pp.2634–2642.
5. 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.
6. 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, 2014.
7. https://support.google.com/websearch/answer/173733?hl=en.
8. D. X. Song, D. Wagner, and A. Perrig, “Practical techniques for searches on encrypted data,” inProceedings of S&P. IEEE, 2000, pp. 44–55.
9. R. Li, Z. Xu, W. Kang, K. C. Yow, and C.-Z. Xu, “Efficient multi-keyword ranked query over encrypted data in cloud computing,”Future Generation Computer Systems, vol. 30, pp. 179–190, 2014.

10. H. Li, D. Liu, Y. Dai, T. H. Luan, and X. Shen, “Enabling efficient multi-keyword ranked search over encrypted cloud data through blind storage,” IEEE Transactions on Emerging Topics in Computing, 2014, DOI10.1109/TETC.2014.2371239.
11. 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.
12. A. Boldyreva, N. Chenette, Y. Lee, and A. Oneill, “Order-preserving symmetric encryption,” in Advances in Cryptology-EUROCRYPT. Springer, 2009, pp. 224–241.
13. W. Sun, B. Wang, N. Cao, M. Li, W. Lou, Y. T. Hou, and H. Li, “Verifiable privacy-preserving multi-keyword text search in the cloud supporting similarity-based ranking,”IEEE Transactions on Parallel and Distributed Systems, vol. DOI: 10.1109/TPDS.2013.282, 2013.
14. J. Yu, P. Lu, Y. Zhu, G. Xue, and M. Li, “Towards secure multi-keyword top-k retrieval over encrypted cloud data,”IEEE Transactions on Dependable and Secure Computing, vol. 10, no. 4, pp. 239–250, 2013.
15. A. Arvanitis and G. Koutrika, “Towards preference-aware relational databases,” inInternational Conference on Data Engineering (ICDE). IEEE, 2012, pp. 426–437.