ISSN 2278-3091
Volume 9 No.2, March -April 2020
International Journal of Advanced Trends in Computer Science and Engineering
Available Online at http://www.warse.org/IJATCSE/static/pdf/file/ijatcse72922020.pdf
https://doi.org/10.30534/ijatcse/2020/72922020

Effective and Secure Approach for Multi-Keyword Quest Graded over Authenticated Data

Shobharani D A[1], Parikshith Nayaka S K[2], Swasthika Jain T J[3] , Dr. Dayanand Lal N[4]
[1]Assistant Professor, GITAM School of Technology, Bengaluru, India, shobharanida@gmail.com
[2]Assistant Professor, GITAM School of Technology, Bengaluru, India, pari2sn@gmail.com
[3]Assistant Professor, GITAM School of Technology, Bengaluru, India, sjain@gitam.edu
[4]Assistant Professor, GITAM School of Technology, Bengaluru, India, dnarayan@gitam.edu

ABSTRACT

IT market is growing, but capital demand is also rising, demanding greater processing power and storage space. Cloud storage enables companies to access their private data and distribute it. Nonetheless, safety is a big concern, and data owners are expected to encrypt data until it is outsourced to cloud. A number of encryption methods have been introduced in this but looking for authenticated data is again a difficult job. A number of schemes have been suggested lately but they don't understand the terminology between the papers. A hybrid hierarchical cluster and latent functional index approach is introduced to help further search semantics in order to address this here. The suggested approach clusters the documents depending on the threshold number, and then splits the clusters into sub clusters before the cluster size maximum is exceeded and the documents are indexed. The new approach often has a benefit over the conventional system in the accuracy and safety of the records that have been recovered.

Key words: Cloud storage, data exchange, multi keyword, index rating, hierarchical clustering, LSI process.

1. INTRODUCTION

Cloud infrastructure facilitates distributed platforms and as well as offering capacity focused on cost-effective usage and tools for different computations, cloud computing growth is rising exponentially. Because of large cloud resources, storage vendors are able to hold their data in the cloud rather than serving customers individually, and since cloud technology often offers a platform for accessing records, they need to be protected. Cloud Service Provider (CSP) can defend it from foreign threats but cannot shield it from internal employees.

One approach to this dilemma is to encrypt data until it's saved in the cloud. Cloud preserves data protection but reduces the flexibility and accuracy of queries. Recently, several search models, such as Boolean model and Vector space model focused on single keyword and multi keyword respectively, have been introduced to guarantee user protection with the

potential to scan over encrypted data. Both such systems have been able to scan keyword-based data but in various ways it has specific definitions and different terms have the same meaning. Therefore, when queried, it is really necessary to get the exact answer. For this reason, a novel model is suggested called Effective and secure approach for multi-keyword quest graded over authenticated data (ESAMKQGA). The proposed model is focused on Latent Semantic Indexing (LSI) to index documents focused on their meanings and hierarchical clustering is centered on these LSI principles, minimizing space and time when looking for documents.

Whenever the consumer wishes to inquire for firm information, users will give the database keywords they need to the data provider. The later job of provider must create trapdoor and return to the requesting customer. Users will then transfer the trapdoors they have obtained into the cloud. When the cloud collects the user's trapdoors, it will use an effective search algorithm to find the user's requested documents based on the specified trapdoors and keywords, and the documents returned to the user by the cloud will be in encrypted form. Finally, user will obtain an encrypted document depending on the quest conducted and user must use the provider's private key to decrypt the obtained documents.

1.1 Literature Review

Song et al.[1]: We initially suggested in this paper an authenticated data quest on the Boolean pattern. They used symmetric key here for both encryption as well as decryption. Yet it does present certain protection issues. To resolve that they suggested yet another approach named encryption of the public key. We used separate key for encryption and decryption in this, but it also presents several challenges, i.e. it is solely focused on Boolean model and uses only one keyword to check for encrypted info.

Pang et al.[2]: In this paper they first suggested a vector space model dependent scheme to scan for encrypted files. In this they used cosine similarity measure to calculate the significance between the data requested and the files retrieved and have estimated meaning for the files retrieved. So, only certain records that reach the cosine mark are obtained for the application. This approach improves efficiency, but it also presents several disadvantages. This will only endorse single keyword, but multi-keyword questions cannot be added.

Cao et al.[3]: In this paper they suggested multi keyword search methodology for authenticated data to prevent the issue of single keyword search. They used the teamwork matching principle here, and generated the production dependent on the number of matched keywords. Several systems of identical nature have been suggested. Yet the biggest downside to both such systems are that it generates the documents that are not only slightly suited to the demand and that scheme is always focused on Boolean pattern. In fact, this scheme is not in a position to combine protection and efficiency.

Jiadi Yu et al.[4]: This paper's conceptual model deals with the value principles and significance between the queried papers, and also reflects on protection concerns. These issues may be solved by adding two encryption rounds each. They looked at privacy concerns during the first trip and centered on safe retrieval of top-k rated documents along with the related scores for obtained documents during the second trip. The pattern also guaranteed a strong degree of data protection.

Ruihui Zhao et al.[5]: In this paper they suggested a methodology named searchable encryption to help the problem of effective keyword dependent search and rank system. It allows custom search for consumer choice in this scheme and delivers the desired results for the proper person. The open directory project is used to eliminate the confusion of the same question from multiple people. It allows multi keyword ranking search with minimal overhead and returns the documents to top k places. Here the people are granted preferences, so it is important to provide the consumer with the highest priority with customized search and performance data.

Chi Chen et al.[6]: In this article, a hierarchical clustering approach was introduced to preserve search similarity for the established question of preserving consistency among the papers. The recommended solution clustered the documents based on predefined meaning and then separated the clusters into sub-clusters before the predefined meaning was achieved. In the quest process the minimum hash sub tree is used to validate the results. Our studies have found that the quest outcome decreases linearly, however the output of the conventional approach exponentially decreases.

2. ISSUE DISCLOSURE

To provide a proficient and consistent privacy preserving encrypted data search on cloud, a proposed scheme called Effective and secure approach for multi-keyword quest graded over authenticated data is implemented. Here users are allowed to search using multiple keywords. This scheme uses the concept of both hierarchical clustering and latent semantic indexing method. In hierarchical clustering all documents are clustered and minimum desired cluster is resulted when queried. In latent semantic indexing, all documents are provided with some index values based on number of repeated keywords. Later users are allowed to retrieve top k ranked documents.

2.1 Suggested Approach

The suggested program consists basically of three stages

a. Report transfer phase
b. Trapdoor Generation method
c. Document recovery phase

Firstly, the vendor is liable for encrypting the words in the paper and delivering the data for transmission to the server and the server will keep the conditions in the contract. Here Advanced Encryption Standard is used It is calculated depending on the frequency of occurrence of the words in the document and that count would be a related index for that text. Finally, together with the authenticated data, the vendor encrypts this version of the individual record and outsources the authenticated database.

The person who does or allows the search request must authenticate with the company in order for the user to do so, because this user must go through the authentication process and access using the credentials. The provider must submit the hidden data to the customer during the authentication process to do the decryption which is saved in the cloud.

When the consumer authenticates with the provider, the provider gives the user a trapdoor and transfers noise-containing trapdoors to the server. Cloud ranks the documents based on index and keyword that given for quest. Most appropriate and related one submitted to the recipient. Using a key the user will then decrypt the relevant document.

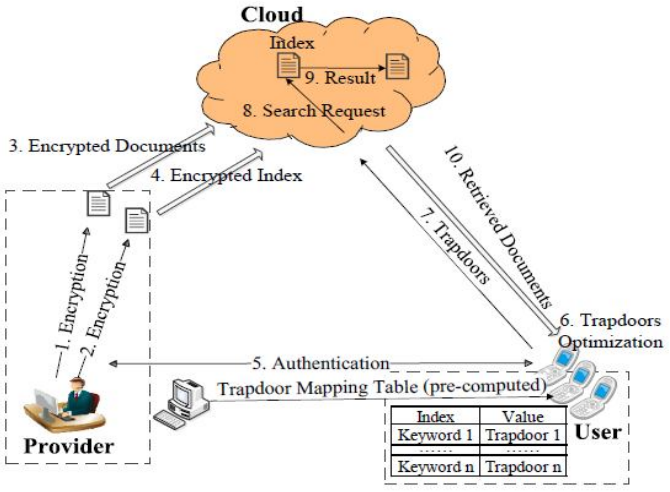


Figure 1: Suggested Approach Architecture

There are so many round trips in a network in conventional system, which creates a pause in quest and heavy network activity, which contributes to quite expensive. The keywords in the type of encryption must be submitted twice as consumer does an exploration requires. This also adds to a quest wait in all these situations, and as well as further data usage, both of which cannot be fair for Smartphone devices. The conventional method has challenges such as noise, and quest time inefficiency.

2.2 Design and Implementation

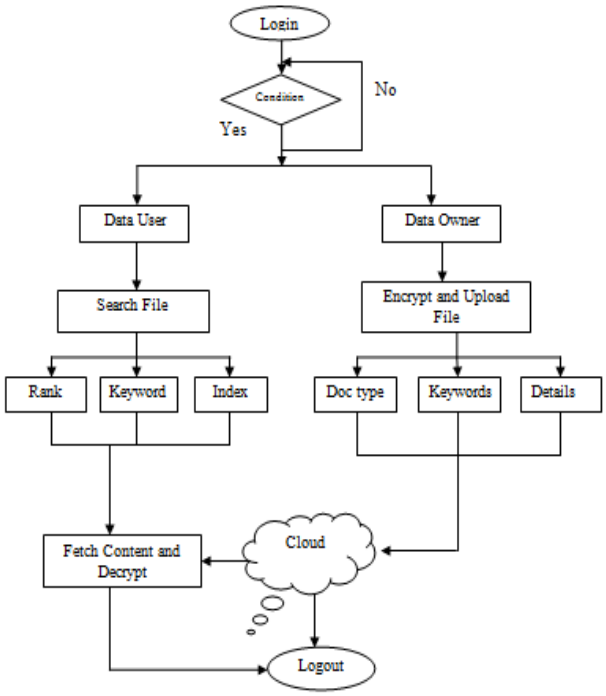


Figure 2: Work Flow of Proposed System

The figure above illustrates the suggested process from login to gathering documents from the cloud phase by phase:
Phase 1: First data owner will sign by supplying valid records.
Phase 2: Providing user name and password, owner will login.
Phase 3: When signing in, the upload frame would be forwarded to a server.
Phase 4: Owner selects the script, keywords are extracted for selected item.
Phase 5: The selected documents will be produced with public key and secret key.
Phase 6: It should be transferred to the cloud by encrypting documents together with keywords.
Phase 7: When any user wants to access info, they need to sign in and register.
Phase 8: Consumer can use multiple keywords to search the data.
Phase 9: Similar records are shown alongside index value for the specified keywords.
Phase 10: Consumer can pick top-k rated documents by looking at the index score.
Phase 11: User may need to authenticate with username and decryption key for chosen papers.
Phase 12: The key to decrypt the requested document will be sent via mail.
Phase 13: Software users may access the document for which they are being asked by utilizing the decryption key.

In proposed system we are building trapdoors and implementing search algorithm in order to decrease the latency in the scan as well as the network traffic. For the trapdoor generation we use keywords which are pre-computed. The trapdoor must search for a user-requested trapdoor routing table for the keyword. Because computing trapdoors are achieved before the request is made, we save network round trips and the suggested method often offers powerful algorithms by reducing network traffic to send trapdoors there.

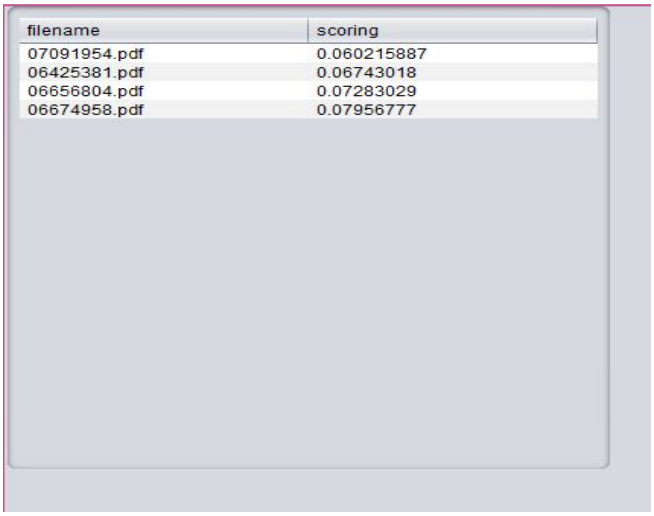
Trapdoor generation algorithm

----------The requirements----------
Keywords = KW
Hash function in Fast accumulation = HA()
Function for mapping is GT()
Noise set is p = (q1,q2,q3....qn)
----------The calculation of trapdoor ----------
Calculate the index for compressed trapdoor 1/t
Extract the term t from KW
If term t is already present in the trapdoor mapping table then extract the term without any noise in it
Else
We have to choose from the noise set p = (q0,q1.....qn)
Return 1/t

Figure 3: Trapdoor generation for Keywords

Apache lucene algorithm

----------Requirements----------
Keywords and trapdoors 1/t1,1/t2.....1/tn
Document index of encrypted are B =j1.....jn
The amount of documents: X
----------Calculation----------
Top documents are selected greatest match of the user need = {Y1,Y2,Y3.....YX)
Total = ones (0, L)
H varies 0 to L
V varies 1 to L
Total[j] ->Total[j] + binary search (t1,t2....tn)
Y -> indices [0 : X -1]
Return Y



filename	scoring
07091954.pdf	0.060215887
06425381.pdf	0.06743018
06656804.pdf	0.07283029
06674958.pdf	0.07956777

Figure 4: Files along with their index values

3. CONCLUSION

Efficient search over authenticated data network reduces search times as well as improves resource productivity as contrasted with conventional network. I began this project by evaluating the shortcomings in the conventional encrypted structure and what are the cloud service bottlenecks that were attempted to resolve in the proposed scheme named Effective and secure approach for multi-keyword quest graded over authenticated data using an apache lucene algorithm that seeks semantics between files and is well ranked in Similar to current structures the program suggested performs well. More research on dynamic operations on files will be in the future.

REFERENCES

[1] Song, D. Wagner, A. Perrig, Practical Techniques for Search on Encrypted Data, IEEE Symposium on Security and Privacy, pages 44-55. IEEE, 2010.
[2] Pang, J. Shen, and R. Krishnan, Privacy-Preserving Similarity- Based Text Retrieval, Transactions on Internet Technology (TOIT), Volume 10, Number 1, Article 4. ACM, 2011.
https://doi.org/10.1145/1667067.1667071
[3] Cao, C. Wang, Privacy-Preserving Multi-keyword Ranked Search over Encrypted Cloud Data, Transactions on Parallel and Distributed Systems, pages 222-233, Volume 25, Issue 1. IEEE, 2013.
https://doi.org/10.1109/TPDS.2013.45
[4] Jiadi, Bing Wang, Privacy-preserving multi-keyword text search in the cloud supporting similarity-based ranking, Symposium on Information, Computer and Communications Security (SIGSAC), pages 71-82. ACM, 2013.
[5] Ruihui Zhao†, Hongwei Li†, Yi Yang†, Yu Liang†, Privacy preserving personalized search over encrypted cloud data supporting multi keyword ranking, Volume 12, Issue 1. IEEE, 2014.
https://doi.org/10.1109/WCSP.2014.6992161
[6] Wenhai Sun, Bing Wang, Verifiable Privacy-Preserving Multi- Keyword Text Search in the Cloud Supporting Similarity-Based Ranking, Transactions on Parallel and Distributed Systems, pages 3025-3035, Volume 25, Issue 11. IEEE, 2014.
https://doi.org/10.1109/TPDS.2013.282
[7] Chi Chen, X. Zhu, P. Shen, An Efficient Privacy-Preserving Ranked Keyword Search Method, Transactions on Parallel and Distributed Systems, pages 951-963, Volume 27, Issue 4. IEEE, 2016.
https://doi.org/10.1109/TPDS.2015.2425407
[8] Sun, Bing Wang, Privacy-preserving multi-keyword text search in the cloud supporting similarity-based ranking, Symposium on Information, Computer and Communications Security (SIGSAC), pages 71-82. ACM, 2013.
https://doi.org/10.1145/2484313.2484322
[9] Ahmad K. Al Hwaitat , Mais Haj Qasem and ,Rim A. Fabozzi, Security of Data Access in Fog Computing using Location-based Authentication, IJATCSE, Volume 9 No. 1, 2020
https://doi.org/10.30534/ijatcse/2020/37912020
[10] Asmita Poojari, Nagesh HR, Kiran Kumar V G and Shantharama Rai C, A Novel Key Scheduling Algorithm for Lightweight Cryptographic Applications, IJATCSE, Volume 9 No. 1, 2020
https://doi.org/10.30534/ijatcse/2020/96912020
[11] Abdallah Ghourabi and Mohamed Jelidi , Experimental Evaluation of a Hybrid Intrusion Detection System for Cloud Computing, IJATCSE, pages 3065- 3073, volume 8 No.6, 2019
https://doi.org/10.30534/ijatcse/2019/65862019
[12] S Kuragod, P Nayak, M Kotari – 2016, "Implementation of IBE with outsourced revocation technique in cloud computing", International Journal of Innovative Research in Electrical, Electronics, Instrument and Control Engineering (IJIREEICE) Vol. 4, Issue 5, PP. 190-193 DOI 10.17148/IJIREEICE.2016.4548
[13] Sujatha Manni1, Parikshith Nayak -2015, "Apriori Based Muti-Keyword Search Over Encrypted Cloud Data", International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering (IJIREEICE) Vol. 3, Special Issue 1, PP. 234-236 DOI 10.17148/IJIREEICE