

## **An Efficient Multi-Keyword Searching with User Confidentiality in Information Networks**

**Suman Sidh<sup>1</sup>, Mr. Sabir Ali<sup>2</sup>, Dr. Sunita Chaudhary\*<sup>3</sup>**

<sup>1</sup>Marudhar Engineering College, Bikaner, Rajasthan

<sup>2</sup>Associate Professor, Marudhar Engineering College, Bikaner, Rajasthan

<sup>3</sup>Professor, Marudhar Engineering College, Bikaner, Rajasthan ORCID ID:- 0000-0001-8913-4897

\*Email:er.sunita03@gmail.com

### **Abstract**

In Information Networks, owners can store their documents over distributed multiple servers. It facilitates users to store and access their data in and from multiple servers by sitting anywhere and on any device. It is a very challenging task to provide efficient search on distributed documents also provide the privacy on owner's documents. The existing system provides one possible solution that is privacy preserving indexing (PPI). In this system, documents are distributed over multiple private servers which are collectively controlled by cloud/public server. When user wants some documents, they query to public cloud, which then returns the candidate list that is private server list to users. After getting list, user can search the documents on specific private server but in this system, documents are stored in plain text form on private server that is privacy is compromised. But proposed system enhances this existing system to make it more secure and efficient. First documents are stored in encrypted form on the private servers and then use Key Distribution Center (KDC) for allowing decryption of data receive from private server, at client side. The proposed system also implements TF-IDF, which provides the ranking of results to users.

Keywords:- Key Distribution Center, TF-IDF, privacy preserving indexing , public cloud, encrypted.

### **INTRODUCTION**

During the time of registration, data customers are both reluctant and capable of using fogs, despite the fact that they value a large proportion of the normal community (for example, incurring substantial harm suitability and open data). Late analysis and mechanical efforts to reclaim data power for cloud customers have resulted in a number of multi-space cloud phases, most notably rising data structures. A data owner in a data system may have absolute ownership throughout his data because of being able to read a variety of primary relationships that it can apparently trust or perhaps even dispatch, particularly without the assistance of another person, an individual server. The structured data does not require constant trust amongst servers, i.e. a proprietor just needs to trust his own server. Data platforms are expanding in a variety of applications.

In one instance, staff in an intranet attempt will save and maintain their own biography on truly administered computers (e.g. IBM You Server [1], [2]). While employees may have specific security issues and may be given the ability to track action courses in field reports, they may be expected to exchange specific details from the corporate level company in order to facilitate future composite measures [2]. For example, a lot of interpersonal relationships have passed on (for example, Diaspora [3], Status [4], and Individual [5]), which are recently developed and have become rational, depending on the plan to disassociate the personal internet connections and the organizational meaning linked.

As with the unified casual communication, the passed interpersonal partnerships do not permit ordinary social clients to send a single server in order to pick their own unique social data and to hold self-described data managed rules for the careful mutual information[6] (e.g. , Facebook and LinkedIn).

Various data structures are being applied to electronic healthcare through the Internet society as a whole (e.g., NHIN Direct Wander [7] open source), acceptable reporting assignments [8], and others. A registered consumer may only need to broaden the spectrum to include available servers (e.g., a virtual machine) and data processing on any of these platforms under full customer supervision. Inside the numerous servers, there are disconnected and related rooms. The sharing of ideas and the containment of an environment is appealing to a variety of technologies.

A trusted game plan is a safety protection document for careful insurance tracking and data-sharing Organisations in controlled dispersed files [9], [10], [11], or short PPI files. APPI is an index advantage that is encouraged to serve various data clients and searchers in a third party social affair (e.g. open cloud). A search engine would take part in a two-style search procedure to find reports of interest: firstly, it addresses a problem of corresponding phrases against a PPI server and gives back in the system a summary of trusted owners (e.g. p0 andp1). By then the search engine will contact its server and sales for customer approval and endorsement for every contestant owners once in a while before seeking them locally. Note that control and support occurs within the data organizer, but not on the PPI server. The PPI system is exceptional because 1) the information in the PPI system is protected with cipher text (i.e. not encrypted) to allow and modify data to provide a valuable level of utility. This applies differently to the current secured data work in the cloud [12], [13], [14]. Unless encryption is used, PPI sticks customer security to obscure sensitive ground-truth data by adding uproars. 2) Only ground-grained data (for example, the responsibility for a proprietor's looked-for expression) is secured in the PPI server while the main substance still remains private and safe on each server.

The ebb and flow investigation and technological initiatives at restoring data control back to open server clients have provided a number of open server multi-space levels, most notably developing knowledge frameworks. The data proprietor can hold total control over their data within an information system by allowing them to explore a variety of pro-associations, which they obviously can trust or can even dispatch a single server directly controlled with no other person. The structure of data needs no trust relationship among servers, which simply means that the data owner have no other option and needs to fully trust and relay on their servers.

Data frameworks create a range of districts of application. In one scenario, agents will store and maintain their own special records on long administered machines in the intranet effort. Although members have their own individual privacy issues and may be able to monitor the plans in the adjacent documents, they may be forced to exchange some details for the advancement of future collaborative initiatives at corporate level gathering. For another, a few casual groups have flowed from late ascent to continual remarkability and rely on the arrangement of disconnecting social information and on the hand fullness of casual contact over a wide range. Under no circumstances like the combined high-profile casual communication through the social network like Facebook and LinkedIn that enables a traditional social consumer to send an independent server to protect its specific data over social platform and make self-reported individuals able to control rules on privacy-conscious exchange of information.

Various cases of frames of information integrate electronic healthcare around the all-inclusive internet community, circulate documents to access controls and others. A cloud provider may have a specified functional design programs in any of these systems (i.e., a virtual machine) and an individual customer data organization area. Spaces arranged within different servers are segregated and adhered to. Different application requirements are fulfilled by knowledge sharing and exchanges around a confining stage.

A candidate game plan includes a data security document that ensures privacy when accessing managed circular documents or short PPI for the purposes of careful requests and information-sharing. In the illustration. 1. The PPI is a list benefits in a third-country (e.g. an open cloud) substance which sends the overall data to various data customers or searchers. A search engineer will take advantage of a two-organized look technique to locate reports of interest: first she addresses requests for important catchphrases against the PPI server, which returns the applicant proprietors' one time (e.g.,  $p_0$  and  $p_1$ ).

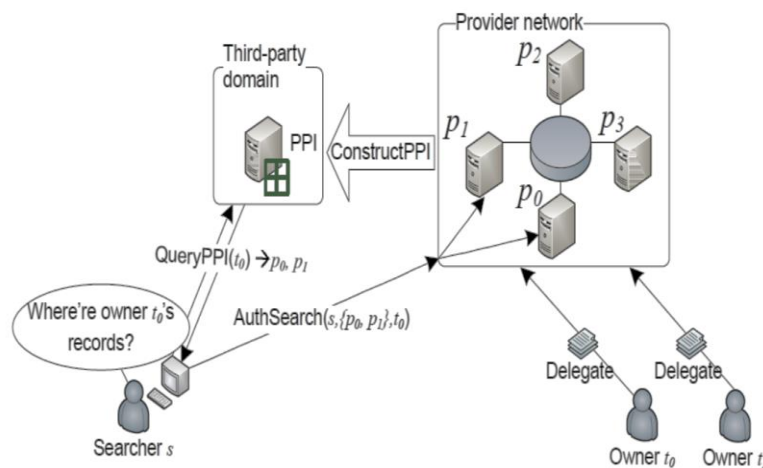


Figure 1: PPI system

### PROBLEM STATEMENT

Efficient search on decentralized documents and the security of privacy of ownership is essential in emerging knowledge networks. A privacy protection index (PPI) technique is available for this issue.

In the presence of multifunctional keyword papers, the under-studied problem with PPI techniques is how to ensure differentiated privacy protection. We use ePPI, which offers quantitatively varying privacy for distributed document searches, to protect this privacy, as well as AES and TF-IDF to include more protection for shared documents and effective searching with ranking performance.

In the presence of a multi-keyword document search, the project aims at providing a protected distributed document search, which provides a quantitatively differentiated data security and also provides results ranking.

- Distributed document search
- Quality and Quantitative privacy protection
- Provide privacy and security
- Maintaining Efficiency and Confidentiality of data
- Data should be stored in encrypted format on private servers and it should provide the ranking of results for time efficient search

## IMPLEMENTATION

**Step1:-** First run MKS\_Monitoring Server

**Step2:-** Run MKS\_PrivateServer1/MKS\_PrivateServer2 (we can run one or many server at one time)

Common Running process description

- i. Browse File (.txt files only)
- ii. Read File
- iii. Click on processing tab.
- iv. Click on button

1. Remove stopwords

2. Generate index(5 words extracted from file will be generated and showed in text box which will act as keyword at user side.)

3. Create table

4. Add index

**Note-:** If we want to add more than one file then don't click on show table button at immediately

Choose another file and do the same processing.

5. Click on show table

6. Click on upload button.

**Step 3:-** On monitoring server click the each button one by one.(make sure that you performed all the operations of private servers1 / private servers 2 (upload)).

**Step 4:-** Run MKS\_User

1.Enter keyword (String) in text field which are generated at the private server processing time and click on search button. you will get the matrix.

2. click on connect to server and do the registration first ,you will get AES Password, remember the password and then do the login, OTP will appear. save/write the OTP.

3.click on File download Tab.

I.click on download button & enter the password.

II.click on decrypt button and enter the password.

III.click on ranking tab (according to keyword file will be preferred to end user)

IV.click on graph button.(graph will be shown to you)

### ANALYSIS

This new method would protect user privacy by ensuring that the used ranking methods were kept separate from the other data, i.e., specifically, the two-way encryption is applied to data queries so the data remain undisturbed and complete. Here are several additional things to consider when measuring time: There are the different procedures such as file upload time, questions, search time, and time of encryption, as well as the time it takes to generate tokens, and the time it takes to rate which can be measured.

By utilizing the non-grouping approaches for ad hoc style harvesting, such as PPI, would result in more user privacy in terms of their data and resulting in better ad-relevant performance.

### SIMILARITY MEASUREMENT

The first table above evaluates the two different systems (the one you already have and the one you are thinking of creating) with respect to their respective similarities. The project completed successfully completed four iterations and each iteration returned a different result that was greater than the previous one. Even if the current system doesn't seem to be identical to the proposed system, it suggests that the proposed values are in need of further expansion.

Table 1 : Similarity Table

	Existing	Proposed
D1	0.47	0.93
D2	0.79	0.96
D3	0.39	0.99
D4	0.42	0.95

This proposed method achieves a better degree of similarity score than the reference system for 4 text in Table 1.

## TIME MEASUREMENT

If you expand the chart to show time versus various processes like uploading, indexing, querying, and token generation, you will be able to see that all processes in the same timeframe generally take the same amount of time. Work the project twice, take the data, and plot the results.

Table 2: Time Measurement Table

	File Upload	Query Search	Encryption Time	Token Generation	Ranking
D1	3.09	0.98	2.04	0.24	0.85
D2	5.94	0.39	3.80	0.49	0.73

By evaluating the data in Table 2. It is shown that the proposed system would follow the trajectory over time.

## CONCLUSIONS

The proposed framework includes connecting the local server with the cloud server for user data sharing. Relevant data or knowledge requires some authentication. This authentication is processed by the encryption method. For the critical performance of secure calculations, the Associate in MPC Care Technology suggests that hidden sharing schemes should be used economically. This way, the machine user can use PPI and encryption technique to access the necessary data in a ranking order.

We also shortened special scanning systems in this research study for encoded data across the cloud. For different search systems, a thorough examination of security and information use is provided. Data safety, data preservation, search, scalability, functionality, index security, query anonymity, results location, index classification, search anonymity, query confidentiality, search unlink ability, and trapdoor unlink ability are regarded to be a portion of the main issues to be addressed with the hunting strategy. All of the search methods listed in this paper have limitations as well. According to the research, Public Key Encryption can provide secrecy, and security details can be double-adjusted as the index using different techniques like the fudge-like parameter look or the tree.

## REFERENCES

- [1] Maryam Hozhabr, Parvaneh Asghari, Hamid Haj Seyyed Javadi, Dynamic secure multi-keyword ranked search over encrypted cloud data, *Journal of Information Security and Applications*, Volume 61, 2021, 102902, ISSN 2214-2126, <https://doi.org/10.1016/j.jisa.2021.102902>.
- [2] X. Liu, G. Yang, W. Susilo, J. Tonien, X. Liu and J. Shen, "Privacy-Preserving Multi-Keyword Searchable Encryption for Distributed Systems," in *IEEE Transactions on Parallel and Distributed Systems*, vol. 32, no. 3, pp. 561-574, 1 March 2021, doi: 10.1109/TPDS.2020.3027003.
- [3] Wang, H., Fan, K., Li, H. et al. A dynamic and verifiable multi-keyword ranked search scheme in the P2P networking environment. *Peer-to-Peer Netw. Appl.* 13, 2342–2355 (2020). <https://doi.org/10.1007/s12083-020-00912-7>
- [4] Li M., Jia C., Shao W. (2020) Blockchain Based Multi-keyword Similarity Search Scheme over Encrypted Data. In: Park N., Sun K., Foresti S., Butler K., Saxena N. (eds) *Security and Privacy in Communication Networks. SecureComm 2020. Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering*, vol 336. Springer, Cham. [https://doi.org/10.1007/978-3-030-63095-9\\_23](https://doi.org/10.1007/978-3-030-63095-9_23), 12 December 2020.

- [5] Shabnam Kasra Kermanshahi, Joseph K. Liu, Ron Steinfeld, Surya Nepal, Shangqi Lai, Randolph Loh, "Multi-client Cloud-based Symmetric Searchable Encryption," in IEEE Transactions on Dependable and Secure Computing 2019, doi: 10.1109/TDSC.2019.2950934.
- [6] Y. Miao, R. Deng, X. Liu, K. R. Choo, H. Wu and H. Li, "Multi-authority Attribute-Based Keyword Search over Encrypted Cloud Data," in IEEE Transactions on Dependable and Secure Computing 2019, doi: 10.1109/TDSC.2019.2935044.
- [7] N. H. Sultan, N. Kaaniche, M. Laurent and F. A. Barbhuiya, "Authorized Keyword Search over Outsourced Encrypted Data in Cloud Environment," in IEEE Transactions on Cloud Computing 2019, doi: 10.1109/TCC.2019.2931896.
- [8] J. Sun, S. Hu, X. Nie and J. Walker, "Efficient Ranked Multi-Keyword Retrieval With Privacy Protection for Multiple Data Owners in Cloud Computing," in IEEE Systems Journal 2019, doi: 10.1109/JSYST.2019.2933346.
- [9] X. Liu, G. Yang, Y. Mu and R. Deng, "Multi-user Verifiable Searchable Symmetric Encryption for Cloud Storage," in IEEE Transactions on Dependable and Secure Computing 2018, doi: 10.1109/TDSC.2018.2876831.
- [10] Mohammad Hassan Ameri, MahshidDelavar, JavadMohajeri, Mahmoud Salmasizadeh, "A Key-Policy Attribute-Based Temporary Keyword Search scheme for Secure Cloud Storage", IEEE Transactions On Information Forensics And Security, Volume: 13, Issue: 10, Oct. 2018.
- [11] Y. Yang, X. Liu, X. Zheng, C. Rong and W. Guo, "Efficient Traceable Authorization Search System for Secure Cloud Storage," in IEEE Transactions on Cloud Computing 2018, doi: 10.1109/TCC.2018.2820714.
- [12] H. Wang, X. Dong and Z. Cao, "Multi-value-Independent Ciphertext-Policy Attribute Based Encryption with Fast Keyword Search," in IEEE Transactions on Services Computing 2017, doi: 10.1109/TSC.2017.2753231.
- [13] C. Guo, X. Chen, Y. Jie, F. Zhangjie, M. Li and B. Feng, "Dynamic Multi-phrase Ranked Search over Encrypted Data with Symmetric Searchable Encryption," in IEEE Transactions on Services Computing 2017, doi: 10.1109/TSC.2017.2768045.
- [14] Y. Miao, J. Ma, X. Liu, X. Li, Q. Jiang and J. Zhang, "Attribute-Based Keyword Search over Hierarchical Data in Cloud Computing," in IEEE Transactions on Services Computing 2017, doi: 10.1109/TSC.2017.2757467
- [15] Yuzhe Tang and Ling Liu, Fellow, IEEE, "Privacy-Preserving Multi-Keyword Search in Information Networks," IEEE Transactions On Knowledge And Data Engineering, Volume 27, Issue 9, 2015
- [16] Ning Cao, Cong Wang, Ming Li, Kui Ren, and Wenjing Lou, "Privacy-Preserving Multi-keyword Ranked Search over Encrypted Cloud Data", Proc. IEEE Infocom, Volume 3, Issue 8, 2014
- [17] Yuzhe Tang , Ling Liu , Arun Iyengar , Kisung Lee, Qi Zhang, " E-PPI: Locator Service in Information Networks with Personalized Privacy Preservation", IEEE Transactions On Knowledge And Data Engineering, Volume 7, Issue 6, 2015
- [18] Yuzhe Tang and Shuigeng Zhou, "LHT: A Low-Maintenance Indexing Scheme over DHTs", The 28th International Conference on Distributed Computing Systems, Volume 10, Issue 3, 2008
- [19] Randy Baden, Adam Bender, Neil Spring, Bobby Bhattacharjee, "Persona: An Online Social Network with User-Defined Privacy", ACM SIGCOMM, Volume 9, Issue 6, 2009
- [20] Ming Li et al., "Toward Privacy-Assured and Searchable Cloud Data Storage Services", IEEE Transactions on Network, volume 27, Issue 4, July/August 2013
- [21] A. Singhal, "Modern Information Retrieval: A Brief Overview," IEEE Data Eng. Bull., vol. 24, no. 4, pp. 35- 43, Mar. 2001.
- [22] D. Song, D. Wagner, and A. Perrig, "Practical Techniques for Searches on Encrypted Data," Proc. IEEE Symp. Security and Privacy, 2000
- [23] Jianfeng Wang et al., "Efficient Verifiable Fuzzy Keyword Search over Encrypted Data in Cloud Computing" , Journal of Computer Science and Information system, volume 10, Issue 2, April 2013
- [24] Wei Zhou et al., "K-Gram Based Fuzzy Keyword Search over Encrypted Cloud Computing "Journal of Software Engineering and Applications, Scientific Research , Issue 6, Volume 29-32, January 2013
- [25] Cong Wang et al., "Enabling Secure and Efficient Ranked Keyword Search over Outsourced Cloud Data", IEEE Transactions on parallel and distributed systems, vol. 23, no. 8, August 2012
- [26] Qin Liuy, GuojunWangyz, and JieWuz, "Secure and privacy preserving keyword searching for cloud storage services", ELSEVIER Journal of Network and computer Applications, March 2011
- [27] Ming Li et al., " Authorized Private Keyword Search over Encrypted Data in Cloud Computing, IEEE proc. international conference on distributed computing systems, June 2011, pages 383-392.