International Journal of Scientific Research in Science, Engineering and Technology
Print ISSN: 2395-1990 | Online ISSN : 2394-4099 (www.ijsrset.com)
doi : https://doi.org/10.32628/IJSRSET

# Survey on Muilt-Keyword Search in Encrypted Data with Privacy Preservation

**Radhika S Landge[1]*, Prof. Nitin R. Chopde[2]**

[1] M.E Scholar, Department of Computer Science & Engineering, G. H. Raisoni College of Engineering & Management, Amravati, Maharashtra, India

[2]Assistant Professor, Department of Computer Science & Engineering, G. H. Raisoni College of Engineering & Management, Amravati, Maharashtra, India

**Article Info**
Volume 8 Issue 2
Page Number: 299-305

**Publication Issue :**
March-April-2021

**Article History**
Accepted : 05 March 2021
Published : 12 March 2021

## ABSTRACT

Cloud computing is computing that is on demand. This is computation focused on the Internet. On request, PCs and various gadgets are provided with shared properties, information and data. It also provides the administration over the phone. In distributed computing, specialist organizations have the capacity to provide cloud storage as needed by customers. They allow customers to store and retrieve information on request from anywhere and on a gadget on a cloud server. This cloud server information management gives rise to such a number of security problems when data is accessed over the network. Information is stored in an encoded ordered structure for security purposes. In this, once it is migrated to the cloud server, the customer has no immediate control over details. We explore the concept behind single watchword hunting over encoded information in this paper and multi catchphrase positioning in addition. In a scrambled framework, cloud data owners need their records with the ultimate aim of safeguarding privacy. It is necessary to build efficient and solid systems for ciphertext search in this way. One test is that during the time spent encryption, the link between documents will typically be secret, which will prompt debasement of vital investigation accuracy execution.

Keywords : Cloud computing, Encryption, Inner product similarity, Single Keyword Search, Multi-keyword search, ranking

## I. INTRODUCTION

As we move into the age of big data, terabytes of data are produced every day worldwide. The idea of "Utility Computing" invented by MIT computer scientist and Turing award winner John McCarthy was preferably recognised as the cloud computing concept over a network in the late 1960's. Industries were searching for some form of significant solution, and for businesses such as IBM, utility computing ended up being something of a big business. In fact, Martin Greenberger pointed to the notion that "advanced future arithmetical machines" are now being used not only institutionalized for scientific calculation and analysis, but also for organizational processes such as accounting and inventory. In addition, he expected his work in which machines would be universal in due time, much like the big power firms running cables everywhere. In order to minimize data processing costs and storage facility spending, companies and consumers who own a significant amount of data often tend to outsource their precious data to cloud facilities. As a consequence, data volume is undergoing a drastic rise in cloud storage facilities. While cloud server providers (CSPs) say that their cloud services are equipped with strong security measures, security and privacy are major barriers to preventing cloud computing services from being more widely accepted [1]. The encryption of data is a standard way of minimizing information leakage. However, this would make the use of data on the server side, such as searching for encrypted data, a very difficult job. In recent years, several ciphertext search schemes have been proposed by researchers through the incorporation of cryptography techniques. These techniques have been proven with proven protection, but their techniques involve large operations and have high time complexity. For the big data scenario, therefore, previous approaches are not suitable where the amount of data is very high and online data processing is needed for applications. Furthermore, in the above processes, the relationship between documents is hidden. The relationship between documents reflects the characteristics of the documents and it is therefore important to preserve the relationship in order to fully convey a document. For instance, it is possible to use the relationship to express its classification. If a document is independent of any other document other than sports-related documents, then it is simple for us to say that that document belongs to the sports group. This significant property has been hidden in conventional methods because of blind encryption. It is therefore beneficial to suggest a methodology that can preserve and use this relationship to accelerate the search process. On the other hand, data search results that return to the users which contain damaged data or have been skewed by the malicious administrator or attacker due to software/hardware failure and storage corruption. Therefore, for users to verify the accuracy and completeness of the search results, a verifiable mechanism should be given. Over the past decade, there has been a rise in the demand for outsourcing of data over a wide variety of networks due to a revolutionary shift in the industry. A common technology cloud computing infrastructure has been adapted to manipulate this vast volume of data in a cost-effective manner, removing the burden of data management. Companies in this data-driven environment prefer to store their data on a cloud consisting of important consumer data assets such as addresses, personal health records, etc. In the development of information technology that offers versatile, ubiquitous access to demand access and capital expenditure saving, cloud computing turns out to be the most important paradigm in the advances in information technology.

## II. LITERATURE REVIEW

Protection and privacy preservation of keyword search was suggested by Qin Liu et al. in [1]. It provides confidentiality of keywords, privacy of data and semantic protection through public key encryption. The primary challenge of this search is that there are more communication and computing costs of encryption and decryption.

In [2], Allowed Private Keyword Search (APKS) was proposed by Ming Li et al. It offers privacy of keywords, privacy of indexes and queries, authorization and revocation of fine-grained searches, multi-dimensional keyword search, scalability and performance. This search approach improves the efficacy of the search using the hierarchy of attributes, but not all attributes are hierarchical in reality.

In [3], Cong Wang et al proposed Stable and Efficient Ranked Keyword Search that resolves overhead processing, privacy of data and keywords, minimal communication and overhead computation. It is not helpful for many searches of keywords, there is also a little bit of overhead in index construction.

Secured fuzzy keyword search with symmetrical searchable encryption was suggested by Kui Ren et al. [4] (SSE). It does not support fuzzy search with public key-based, searchable encryption, nor can it perform semantic search for multiple keywords. Fuzzy searchable index updates are not successfully enforced.

Ming Li et al. [5] suggested the system of Searchable Cloud Storage for Privacy Assured. It is implemented by means of SSE, Scalar-Product-Preserving Encryption and Symmetric Encryption Order-Preserving. It supports the standards for privacy and accessibility. This scheme does not allow searchable encryption based on the public key.

Wei Zhou et al. [6] suggested Graded Search with a K-gram based fuzzy keyword. The k-gram fuzzy keyword index for files D is generated in this owner and tuple <I, D> is submitted to the search server (SS) which is inserted for size control in the bloom filter. The encrypted file D is uploaded to the server for storage. The problem is that, however, the size of the fuzzy keyword collection centered on k-gram depends on the value of the jacquard coefficient.

J. Stable Channel Free Public Key Encryption with Keyword Search (SCF-PEKS) system proposed by Baek et al. in [7]. Cluster servers generate their own public and private key pairs in this approach, but this process suffers from KGA's external attacker.

H. S. Rhee et al. [8] suggested Trapdoor in Public-Key Encryption with Keyword Search distinguishability (IND-PEKS). This outsourcing is done in the form of SCF-PEKS. It uses KGA to evaluate the frequency of the occurrence of the keyword trapdoor from outside attackers.

Peng Xu et al. [9] suggested PEFKS Public-Key Encryption with Fuzzy Keyword Scan, which generates fuzzy keyword trapdoor Tw and precise keyword trapdoor Kw for W. User requests for CS from Tw. CS then tests Tw with a fuzzy keyword index and sends the Fuzz Test algorithm performed by CS to the superset of matching cypher texts. The ExactTest algorithm of the user method for checking ciphertext with Kw and retrieving the encrypted data. For a large database, the process of producing a fuzzy keyword index and an exact keyword index is difficult.

Ning et al. [10] proposed Multi Keyword Ranked Search Protection of Privacy (MRSE). It is useful over encrypted data for the known cypher text model and the context model. It provides low computing and overhead communication. For multi-keyword searching, the matching coordinates are chosen. The downside is that MRSE has a slight standard deviation that limits the confidentiality of the keyword.

| Sr.No | Author(s) | Concept used | Evaluation Parameter | Claims by Concerned Author(s) | Our Findings |
|---|---|---|---|---|---|
| 1. | Yuzhe Tang and Ling Liu | Privacy preserving Indexes | Semantic Meanings | e-MPPI for providing the distributed document search along with quantitatively differentiated privacy preservation | An MPC-reduction technique based on the efficient use of secret sharing schemes. We also discovered common-term vulnerability and |

| | | | | | |
|---|---|---|---|---|---|
| | | | | | proposed a term-mixing solution. |
| 2. | Ning Cao, Cong Wang, Ming Li, KuiRen, and Wenjing Lou | Privacy preserving multi-keyword ranked search over encrypted cloud data (MRSE) | Coordinate Matching, Inner Product Similarity | We establish a set of strict privacy requirements for such a secure cloud data utilization system. Among various multi-keyword semantics, we choose the efficient similarity measure of “coordinate matching”, i.e., as many matches as possible, to capture the relevance of data documents to the search query. | MRSE using secure inner product computation. |
| 3. | Y. Tang, L. Liu, A. Iyengar, K. Lee, and Q. Zhang | Anidentity-mixing protocol against the attack in e-PPI. | Effectiveness the e-PPI in terms of delivering quantitative privacy protection. Performance of our index construction protocol. | Proposed e-PPI construction protocol is the first without any trusted third party and/or trust relationships between providers. | The construction protocol for ϱ-PPI without any trusted party involved. |
| 4. | Yuzhe Tang, Shuigeng Zhou | a Low maintenance Hash Tree, for efficient data indexing over DHTs. | maintenance cost, performance for exact-match queries | A Low maintenance Hash Tree, for efficient data indexing over DHTs. LHT employs a novel naming function and a tree summarization strategy to gracefully distribute its index structure. | LHT can save up to 75%(at least 50%) maintenance cost, and achieves better performancein exact-match and range query processing. |
| 5. | Randy Baden, Adam Bender | Attribute-based encryption (ABE), Online social networks (OSNs) | Privacy in OSNs | Persona provides an effective means of creating applications in whichusers, not the OSN, define policy over access to private data. | Persona hides user data with attribute-based encryption (ABE), allowing users to apply fine-grained policies over who may view their data. |
| 6. | K.S.Sureh, Mrs.SaritaChowdary, T. Balachary | private-key cryptography and symmetric Encryption | Privacy & Security | The personal health record system needs security against attackers and hackers. Scalable and Secure sharing includes basic securities to protect the information from unauthorized access and loss. | Paper proposed the new approach for existing PHR system for providing more security using symmetric encryption which plays an important role because these are unique. |

| | | | | | |
|---|---|---|---|---|---|
| 7. | Y. Tang, T. Wang, and L. Liu | SS-PPI, a novel privacy-preserving index abstraction | Privacy Protection and Execution Efficiency | Focus is on addressing the privacy concerns of content providers; that is, the search should not reveal the specific association between contents and providers | It incorporates access control policies in the privacy preserving index, which improves both search efficiency and attack resilience; |
| 8. | M. Bawa, R. J. Bayardo, Jr, R. Agrawal, and J. Vaidya | Privacy-preserving Index a distributed access-control enforcing search protocol | Privacy-preserving index (PPI) | The new index provides strong and quantifiable privacy guarantees that hold even if the entire index is madepublic. | Content providers maintain complete control in defining access groups and ensuring its compliance |
| 9. | A. Ben-David, N. Nisan, and B. Pinkas | Secure Multi-Party Computation | Number of computation players, Size of the circuit and General run time | The BMR protocol is modified in a novel way and considerably improved its performance by using the Ben-Or-Goldwasser-Wigderson (BGW) protocol for the purpose of constructing gate tables. | The performance of fast machines is dramatically reduced if even a single player is using a weak machine. The reason for this is that in every communication round the fast players have to wait until the weakest player finishes its computation and sends its results. |
| 10. | S. Zerr, E. Demidova, D. Olmedilla, W. Nejdl, M. Winslett, and S. Mitra | r-confidential Zerber indexing | Response Size for the DFM Index, Efficiency in Query Answering, | A tunable r-confidentiality measure, as the degree of information from inaccessible documents an index can leak, given an adversary compromises the index and possesses some background knowledge on the corpus and/or language statistics. | Zerber, an r-confidential global inverted index for sensitive documents. Zerber relies on a centralized set of largely untrusted index servers and offers resistance against inappropriate information disclosure even if k-1 index servers are compromised. |

## III. PROPOSED WORK

For the purpose of privacy protection, cloud storage owners tend to externalize documents in an encrypted form. Therefore, the production of effective and reliable ciphertext search techniques is important. One problem is that in the encryption process, the relationship between documents will typically be hidden, leading to a major deterioration in the performance of search accuracy. There has also been a dramatic increase in the amount of information in data centers. This would make it much more difficult to design ciphertext search schemes that can provide vast quantities of encrypted data with effective and accurate online information retrieval.

In the context of ontology, the next generation network called the Semantic Web can assist the user to retrieve the helpful data stored on the cloud and make the data accessible to the user concealed behind the cloud. The purpose of the proposed ranking algorithm is to provide users with relevant data from the result collection.

In order to accommodate more search semantics and also to meet the demand for fast cypher text search within a big data environment, a hierarchical clustering approach is suggested. The suggested hierarchical method clusters the documents on the basis of the minimum threshold of importance and then divides the resulting clusters into sub-clusters until the maximum cluster size limit is reached. This method will achieve a linear computational complexity in the search process against an exponential increase in the size of document set. A structure called a minimum hash sub-tree is built to check the validity of the search results. In order to understand inherent relationships between concepts using ontologies, we expand this definition of semantic similarity. With multi-keyword and ontology, we suggest rating algorithms.

In order to accommodate more search semantics and also to meet the demand for fast cypher text search within a big data environment, a hierarchical clustering approach is suggested. The suggested hierarchical method clusters the documents on the basis of the minimum threshold of importance and then divides the resulting clusters into sub-clusters until the maximum cluster size limit is reached. This method will achieve a linear computational complexity in the search process against an exponential increase in the size of document set.

A structure called a minimum hash sub-tree is built to check the validity of the search results. In order to understand inherent relationships between concepts using ontology, we expand this definition of semantic similarity. We also proposed a multi-keyword and ontological rating algorithm.

## IV. CONCLUSION

This paper explores different methods of searching for data storage in the encrypted cloud. The protection and data utilization problems associated with all available search techniques are systematically presented in cloud storage. Thus, keyword privacy, data privacy, index privacy, query privacy, fine-grained search, scalability, performance, result ranking, index confidentiality, query confidentiality, query Unlinkability, semantic protection and trapdoor Unlinkability have been described as the main problems to be addressed for safe data utilization. Most of the search techniques concentrate primarily on security and others on data use. The drawbacks of all the methods for searching are also discussed. Protection can be provided by public-key encryption and efficient data utilization through fuzzy keyword search through the above survey. We believe that this study will allow researchers to shape their problem in the area of cloud storage data utilization.

## V. REFERENCES

[1]. Qin Liuy, Guojun Wangyz, and Jie Wuz,"Secure and privacy preserving keyword searching for cloud storage services", ELSEVIER Journal of Network and computer Applications, March 2011

[2]. 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

[3]. 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

[4]. Kui Ren et al., "Towards Secure and Effective Data utilization in Public Cloud", IEEE Transactions on Network, volume 26, Issue 6, November / December 2012

[5]. Ming Li et al.,"Toward Privacy-Assured and Searchable Cloud Data Storage Services", IEEE Transactions on Network, volume 27, Issue 4, July/August 2013

[6]. 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,January2013

[7]. J. Baek et al., "Public key encryption with keyword search revisited", in ICCSA 2008, vol. 5072 of Lecture Notes in Computer Science, pp. 1249 - 1259, Perugia, Italy, 2008. Springer Berlin/Heidelberg.

[8]. H. S. Rhee et al., "Trapdoor security in a searchable public-key encryption scheme with a designated tester," The Journal of Systems and Software, vol. 83, no. 5, pp. 763-771, 2010.

[9]. Peng Xu et al., Public-Key Encryption with Fuzzy Keyword Search: A Provably Secure Scheme under Keyword Guessing Attack",IEEE Transactions on computers, vol. 62, no. 11, November 2013

[10]. Ning Cao et al.," Privacy-Preserving Multi-Keyword Ranked Search over Encrypted Cloud Data", IEEE Transactions on parallel and distributed systems, vol. 25, no. 1, jan 2014

[11]. D. X. D. Song, D. Wagner, and A. Perrig, "Practical techniques for searches on encrypted data," in Proc. S & P, BERKELEY, CA, 2000, pp. 44.

[12]. C. Wang, N. Cao, K. Ren, and W. J. Lou, Enabling Secure and Efficient Ranked Keyword Search over Outsourced Cloud Data, IEEE Trans. Parallel Distrib. Syst., vol. 23, no. 8, pp. 1467-1479, Aug. 2012.

[13]. W. Sun, B. Wang, N. Cao, M. Li, W. Lou, Y. T. Hou, and H. Li, "Privacy-preserving multi-keyword text search in the cloud supporting similarity-based ranking," in Proc. ASIACCS, Hangzhou, China, 2013, pp. 71-82.

[14]. R. X. Li, Z. Y. Xu, W. S. Kang, K. C. Yow, and C. Z. Xu, Efficient multi-keyword ranked query over encrypted data in cloud computing, Futur. Gener. Comp. Syst., vol. 30, pp. 179-190, Jan. 2014.

[15]. Gurdeep Kaur, Poonam Nandal, "Ranking Algorithm of Web Documents using Ontology", IOSR Journal of Computer Engineering (IOSR-JCE) eISSN: 2278-0661, p- ISSN: 2278-8727Volume 16, Issue 3, Ver. VIII (May-Jun. 2014), PP 52-55

**Cite this article as :**

Radhika S Landge, Prof. Nitin R. Chopde, "Survey on Muilt-Keyword Search in Encrypted Data with Privacy Preservation", International Journal of Scientific Research in Science, Engineering and Technology (IJSRSET), Online ISSN : 2394-4099, Print ISSN : 2395-1990, Volume 8 Issue 2, pp. 299-305, March-April 2021.
Journal URL : https://ijsrset.com/IJSRSET218279