ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 1, January 2017

Survey on Recent Keyword Search Techniques on Outsource Encrypted Big Data

Neha Mahajan, V.M.Barkade

M.E. Student, Department of Computer Engineering, Rajarshi Shahu College of Engineering, Pune, India

Professor, Department of Computer Engineering, Rajarshi Shahu College of Engineering, Pune, India

ABSTRACT: Cloud data servers are the attractive outsource data storage options for number of academics and industries, because of various reasons like efficient data storage management, economical cost and fast deployment. In this era, cloud service providers can provide the access to some part of storage servers on demand of requested users, who can access it from anywhere on any time and any kind of device. But with such successful data outsourcing solution, various data privacy and confidentiality issues are arises. Because such data is available on internet. It can be illegally access by unauthorized users and they can alter it. Therefore Data encryption technique is one of the solutions to maintain the data privacy and integrity. In which authorize data owner encrypt the data before outsourcing. But sometime, on such encrypted data, efficient data utilization is not possible. It requires, some standard techniques, which can provide the efficient multikeyword search over outsourced encrypted data. In this survey, we investigate recent multikeyword searching techniques over encrypted data and make their comparative analysis on the basis of technique/ approach used their advantages and disadvantages.

KEYWORDS: Privacy preserving, data mining, outsource data, keyword search.

I. INTRODUCTION

Cloud computing becomes interesting approach in various applications like academics and the industries. Cloud computing has several features to attract users such as resource management, economical cost and easy and fast deployment. Because of the huge economic advantages of cloud computing, a most of the organization deployed their cloud centers. Examples of such cloud centersare, the Elastic Compute Cloud of Amazon, the App Engine of Google, the Azure of Microsoft, and Blue Cloud of IBM. Even though such lot of advantages, users has some worries about outsourcing their data on cloud servers. Owners data is sometime sensitive data such as personal records, financial records password record etc., because once the data is outsourced, owner can not directly control the data. It is available online sometime. The Cloud Service Provider (CSP) can maintain the security of such sensitive data by using some techniques like fireware, virtualization, and Intrusion Detection System (IDS).

In this case, CSP gains total control over such data. But there is no guarantee or full trust on the employee of CSP. They can leak or modify the data. That is they can reveal the sensitive information of data owners. So to overcome the problem of security of outsource data, data encryption is one of the best solution. But such data encryption and search in traditional approaches is not practical, where all encrypted files are downloaded and then decrypt it at local side to get the desire file. This is very time consuming and tricky process, also an unrealistic process. It may increase the communication and computation cost of searching the data in encrypted outsourced data. To overcome these issues, there is need to revise such system, which will be results in, secure and privacy preserving search approach on encrypted data. Such system may consist of data owner, cloud server and data users. In this system, data owner upload encrypted files on cloud server. For data user, authorization is perform with trapdoor generation and submit to cloud server. After this, on receiving search query from data user, cloud server provides search results without revealing the original contents of sensitive data.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 1, January 2017

Here we categories different search techniques based on different criteria as follow:

1. Searchable Encryption

With this technique user can store their data in encrypted format to improve the security of data, also other user can search in this encrypted secure data. This can also workout on large scale data. Various encryption techniques can be used to achieve this, such as AES, ECC etc.

2. Ranked Single keyword search

In this technique, single random keyword is an input to cloud server. In return cloud server generates the most related file that matched with input keyword. Ranked keyword search generated the results rank wise instead of just providing matched results. It will reduce the cost of searching also provide the most related results to improve the user experience.

3. Multi keyword Search

To make searching system more practical, system can support to multi keyword search in the place of single keyword search. With this system, input search query may contain more than one keyword, which improves the accuracy of search query. Multiple keywords has the capability to explain the search query accurately.

4. Ranked Multi-keyword Search

In this technique, multiple random keywords is an input to cloud server. As a reply cloud server generates the most related file that matched with input keywords. Ranked keyword search generated the results rank wise instead of just providing matched results.

5. Fuzzy keyword search

This technique allows the minor errors in search query keywords. It improves the robustness and increase the usability of fuzzy keyword search system. It also generates the result for misspelled keywords.

6. A conjunctive keyword search

In this approach, a query request has multiple keywords and for each keyword trapdoor is generated. The final result is the intersection of results of each keyword in search request.

7. Similarity keyword search

This search system fetches all possible similar data with search query request from cloud server. Because of security and privacy concerns, it is very challenging to achieve similarity keyword search approach in practical. During development of such system, efficiency, security and robustness should be considered.

8. Attribute based keyword search

The attribute based keyword search system is useful to support multiple data owners to share their information with other multiple data owners in large scale cloud storage systems. Such system securely shares the data among multiple data owners

In this paper further we will see: Section II talks about related work studied till now on topic. Section III discusses existing system. Section IV describes proposed system and this paper is concluded in section V.

II. RELATED WORK

In this section discuss the existing method developed for cloud computing. now we discuss different methods developed by the researchers, the different methods are as follows:

In paper [1] authors analyzed cipher text search in the cases of cloud storage.they found the keeping up the semantic relationship between various plain files over the related encrypted documents and give the configuration procedure to enhance the execution of these mantic searches. authors additionally propose the mrse-hci design to adapt to the necessities of information explosion, online data recovery and semantic search. in the meantime, an evident component is additionally proposed to ensure the accuracy of search results.

In paper [2] authors introduced the comparability between various documents into cipher text search. They also derived the MRSEHCI architecture and compatible algorithm. Moreover, they examine the search productivity and security in two well threat models. An exploratory stage is worked to assess the search productivity, exactness and rank security.

In paper [3] authors designed and implemented dynamic symmetric searchable encryption system which effectively as well as privately seeks server-held encrypted databases with several billions of record-keyword sets.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 1, January 2017

Fundamental hypothetical development underpins single-keyword searches and offers asymptotically ideal server index size, totally parallel searching, and minimal leakage. Authors execution effort conveyed to the fore a few cases overlooked by before coarse-grained theoretical performance examinations, including low level space use, I/O parallelism and good put. We as needs be present a few enhancements visit hypothetically ideal development that model the prototype's qualities characteristics to beat these components.

In paper [4] authors get through the indicate examination the query about cipher text in big information environment. Outline cipher text query framework model BDES in huge information environment, control the virtual keyword numerical, checking the BDES performance finally. Agreeing to creators the next step, research direction predominantly about the algorithm of cipher text query, including algorithm of fluffy keywords and semantic query, put forward more flexible and efficient cipher text query algorithm.

In paper [5] authors have focused on practical elements and incremental redesign in cloud. They made utilization of document replication and partition with one-to-many preserving encryption to achieve privacy-preserving assure. With a set of reasonable practical factors, the proposed system can give a proficient efficient privacy-preserving multi-keyword ranked search service. Also, when information owners transfer new information in cloud, the proposed system can effectively perform incremental upgrade without rebuilding entire indices.

In paper [6] authors has concentrated on how the client can viably store their personal documents on cloud while taking care of privacy of their files and at whatever and wherever important they can recover them by forwarding a query consisting of numerous keywords. Security will be accomplished by scrambling the questions. Because of the clients query, framework will coordinate the keywords from query to the files utilizing "keyword-matching principle". Best positioned reports will get brought which will comprise of keywords indicated by client in query. Checking the rank of the recovered file should be possible by computing what number of docs contains the predetermined keywords how often.

In paper [7] authors proposed a light-weight search approaches that backings efficient multi-keyword ranked search in cloud computing framework. Essential scheme utilizes the polynomial function to hide the encrypted keyword and search patterns down multi-keyword ranked search. We then enhance the essential scheme and propose a privacy-preserving system which uses the safe inner product strategy for ensuring the security of the looked multi-keywords. Careful examination on the protection certification of our proposed system is given, and broad tests depending on this present real-world dataset are additionally directed.

Table 1: Survey Table

Paper Name	Proposed work	Advantages	Disadvantage
Privacy-Preserving Utility Verification of the Data Published by Non-interactive Differentially Private Mechanisms	Propose a privacy-preserving utility verification mechanism based upon cryptographic technique for DiffPart-a differentially private scheme designed for set-valued data.	Improve the security and efficiency of the system	Association rule mining over huge data may increase the execution time.
On k-Anonymity and the Curse of Dimensionality	View the k-anonymization problem from the perspective of inference attacks over all possible combinations of attributes.	the curse of high dimensionality also applies to the problem of privacy preserving data mining.	loss for high-dimensional data
ℓ-Diversity : Privacy Beyond K-Anonymity	Show with two simple attacks that a k-anonymized dataset has some subtle, but severe privacy problems.	ℓ-diversity is practical and can be implemented efficiently	Limited till this work

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 1, January 2017

III. PROPOSED SYSTEM

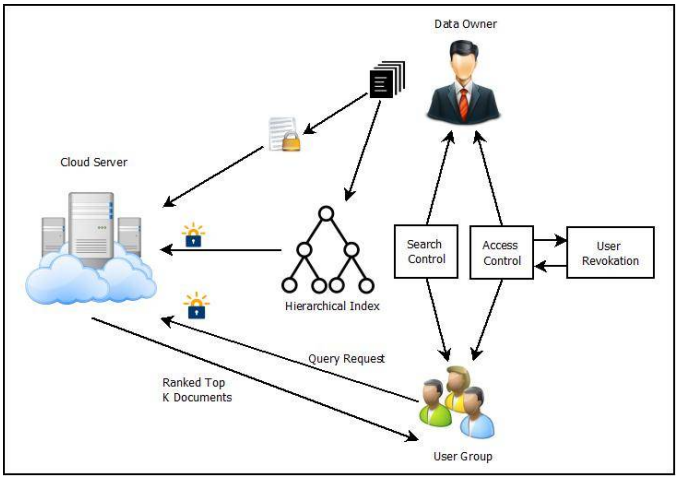

Fig 1: Proposed system architecture

Figure demonstrates the system architecture of the proposed system. The proposed system has of different modules; the main modules are listed below:
Cloud Server: Cloud server is the storage medium on which the data is stored. Data stored on the cloud is in encrypted form by which only authenticated users can to use the data in this way security is provided for data stored on cloud.

Data Owner: Data owner are the user which uses the cloud storage to store the data on cloud for the sharing purpose with the different user present in the group. He is the one responsible for encryption of data while storing it on cloud.

User Group: This is the group of users who are willing to use the data stored on the cloud. They have made a request in order to access the data.

Access control module: At the time when user enters in the group or leaves the group this module is responsible for providing the access keys when one user enters in the group also responsible for the key revocation at the time when the user leaves the group.

Search Control: This module is brought in to action when user searches for specific keywords on the cloud. This module is able to retrieve the files with the top k ranked documents acceding to the user search.

IV. CONCLUSION

This survey makes the comparative study of some recent multi keyword search techniques on large scale encrypted cloud data. Such techniques maintain the privacy and security of data during search. In comparative analysis, we have compare then on various criteria such as, key idea of approach, their advantages and disadvantage. From this survey we identify the limitations of existing system, which can be explored in future. We also identify some challenges of secure data utilization like, privacy, search efficiency, scalability, accuracy, ranking of results etc. This survey will help various researchers to address the issue of privacy preserving cloud bid data utilization.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 1, January 2017

REFERENCES

[1] Chen, Chi, et al. "An efficient privacy-preserving ranked keyword search method." IEEE Transactions on Parallel and Distributed Systems 27.4 (2016): 951-963.
[2] Chen, Chi, et al. "A hierarchical clustering method for big data oriented ciphertext search." Computer Communications Workshops (INFOCOM WKSHPS), 2014 IEEE Conference on. IEEE, 2014.
[3] Cash, David, et al. "Dynamic Searchable Encryption in Very-Large Databases: Data Structures and Implementation." IACR Cryptology ePrint Archive 2014 (2014): 853.
[4] Yanzhu Liu, Zhi Li, Wang Guo and Wu Chaoxia, "Privacy-preserving multi-keyword ranked search over encrypted big data," Third International Conference on Cyberspace Technology (CCT 2015), Beijing, 2015, pp. 1-3.
[5] Ching-Yang Tseng, ChangChun Lu and Cheng-Fu Chou, "Efficient privacy-preserving multi-keyword ranked search utilizing document replication and partition," 2015 12th Annual IEEE Consumer Communications and Networking Conference (CCNC), Las Vegas, NV, 2015, pp. 671-676.
[6] D. D. Rane and V. R. Ghorpade, "Multi-user multi-keyword privacy preserving ranked based search over encrypted cloud data," Pervasive Computing (ICPC), 2015 International Conference on, Pune, 2015, pp. 1-4.
[7] Y. Ren, Y. Chen, J. Yang and B. Xie, "Privacy-preserving ranked multi-keyword search leveraging polynomial function in cloud computing," 2014 IEEE Global Communications Conference, Austin, TX, 2014, pp. 594-600.