IJIRSET
International Journal of Innovative Research in
SCIENCE | ENGINEERING | TECHNOLOGY

e-ISSN: 2319-8753 | p-ISSN: 2320-6710

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 10, Issue 6, June 2021

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 7.512

9940 572 462 | 6381 907 438 | ijirset@gmail.com | www.ijirset.com

International Journal of Innovative Research in Science, Engineering and Technology (IJIRSET)

|e-ISSN: 2319-8753, p-ISSN: 2320-6710| www.ijirset.com |Impact Factor: 7.512|

|| Volume 10, Issue 6, June 2021 ||

| DOI:10.15680/IJIRSET.2021.1006432|

Searching Techniques over Encrypted Cloud Data

Miss. Nutan S. Shelke [1], Prof. Monika D. Rokade. [2]

PG Student, Department of Computer, SPCOE, Dumbarwadi (Otur) Pune, India[1]

Assistant Professor (ME Co-coordinator) : Department of Computer, SPCOE, Dumbarwadi (Otur) Pune, India[2]

ABSTRACT: Cloud storage can now accommodate massive amounts of data. The data must be encrypted in order to maintain its secret. When data is outsourced to the cloud, encryption solutions play a critical role in ensuring that only authenticated users have access. This research focuses on several search strategies for encrypted data in the cloud. A comparison of different types of searchable encryption systems is presented in this study, based on factors such as multiple keyword, number of scale data, search complexity, accuracy, and amount of storage.

KEYWORDS: Cloud, Ranked search, Fuzzy search, Conjunctive search, Multi-keyword search, Semantic search

I. INTRODUCTION

A cloud is a collection of interconnected servers and computers that may be accessed via the internet [1]. This pool of virtualized computer resources is referred to as a "pool of virtualized computer resources," and this pool of virtualized resources provides data abstraction. The key characteristics of cloud computing are service oriented, availability of resources (e.g. networks, servers, storage, applications and services), sharing resources, on-demand delivery of resources, virtualization, scalability of resources, pay-per-use, loose coupling, self-service, ease of use, and high fault tolerance [1,2].

In the cloud, a large amount of data can be stored. When data is outsourced to the cloud, encryption techniques are critical because they maintain the confidentiality of data access. In cloud computing, many approaches for keyword searching on encrypted data have been proposed. We can categorise searching strategies over encrypted data into five primary classes based on related work. 1. Searchable Index Schemes, in which the index file is created from the extracted keywords. 2. Fuzzy key word search, which allows for a match between the query keyword and the stored keywords, with the closest results being returned. 3. Conjunctive keyword search only returns documents that match the search query's keywords. 4. Multi-keyword ranked search, which retrieves documents containing one or more of the words supplied in the query. 5. Semantic search: this type of search is based on the original keywords having many meanings and a distinct structure.

II. RELATED WORK

Searchable Index Schemes

Song et al. provided the first structure for searching over encrypted data in 2000[3]. The system didn't have an index; instead, it encrypts each word in the file separately, thus the search had to go through the entire file. This method is easy and quick, however it requires sequential data scanning and isn't ideal for big amounts of data. When searching for a big number of papers, the method is too slow.

Goh [4] created per-file searchable index schemes in 2002. He created indexes for the data files using a bloom filter, which reduces the cost of searching based on the number of files.

Boneh et al., 2004 [5], present the first searchable encryption system based on the public key system, in which the server includes encrypted files and keywords. To search, the user builds a keyword trapdoor using its private key. The server looks for encrypted keywords that already exist and provides encrypted files that match. The trapdoor in Bonh's scheme can be learned, and then it will expose knowledge about the keyword [5]. The G.Duntao approach [6] attempted to overcome the memorised trapdoor problem. G.Duntao et al. proposed a temporary keyword search strategy over public key encryption based on Boneh's scheme. The G.Duntao scheme addresses the memorised trapdoor problem. The G.Duntao scheme separates time into a few time slides and creates a temporary trapdoor for each of them. In some cases, the trapdoor of a

| e-ISSN: 2319-8753, p-ISSN: 2320-6710| www.ijirset.com | Impact Factor: 7.512|

|| Volume 10, Issue 6, June 2021 ||

| DOI:10.15680/IJIRSET.2021.1006432|

keyword does not reveal anything about the trapdoor at all. To improve the efficiency of your search, An index searching strategy was introduced by Curtmola et al, 2006 [7]. It creates an index file in the Curtmola scheme. Each entry in the index file contains a trapdoor for each keyword, and the IDs for the related files contain the keyword. Q.Liu et al. [8] suggested secure and privacy-preserving keyword searching (SPKS) in 2011. In the SPKS scheme, a cloud service provider (CSP) can identify files that include query keywords and perform a partial decryption of these files before giving search results. The Q.Liu technique minimises the user's decryption overhead.

B.Long et al [9] presented a better public encryption system with keyword search in 2012. The scheme creates an index in the same way as the Boneh scheme [5] does. By splitting keywords into sets, the technique decreases the time cost of keyword searches. When a user searches for a term, the server scans the specific keywords in each set. When a user's search results do not match the input search, similarity search over outsourced cloud encrypted data remains a serious issue.

For each term in each text, C.Wang et al [10] proposed building a storage with a group of comparable keywords. The Wang technique shows the user all of the most comparable matched entries. The similarity of keywords is calculated by calculating the distance between them. In addition, C.Wang presents a symbol-based trie-traverse searching mechanism for achieving similarity search with constant search time complexity.

Monika Rokade and Yogesh Patil [11] proposed a system deep learning classification using nomaly detection from network dataset. The Recurrent Neural Network (RNN) has classification algorithm has used for detection and classifying the abnormal activities. The major benefit of system it can works on structured as well as unstructured imbalance dataset.

The MLIDS A Machine Learning Approach for Intrusion Detection for Real Time Network Dataset has proposed by Monika Rokade and Dr. Yogesh Patil in [12]. The numerous soft computing and machine learning classification algorithms have been used for detection the malicious activity from network dataset. The system depicts around 95% accuracy ok KDDCUP and NSLKDD dataset.

Monika D. Rokade and Yogesh Kumar Sharma [13] proposed a system to identification of Malicious Activity for Network Packet using Deep Learning. 6 standard dataset has sued for detection of malicious attacks with minimum three machine learning algorithms.

Sunil S. Khatal and Yogesh kumar Sharma [14] proposed a system Health Care Patient Monitoring using IoT and Machine Learning for detection of heart and chronic diseases of human body. The IoT environment has used for collection of real data while machine learning technique has used for classification those data, as it normal or abnormal.

Data Hiding In Audio-Video Using Anti Forensics Technique For Authentication has proposed by Sunil S.Khatal and Yogesh kumar Sharma [15]. This is a secure data hiding approach for hide the text data into video as well as image. Once sender hide data into specific objects while receivers does same operation for authentication. The major benefit of this system can eliminate zero day attacks in untrusted environments.

Sunil S.Khatal and Yogesh Kumar Sharma [16] proposed a system to analyzing the role of Heart Disease Prediction System using IoT and Machine Learning. This is the analytical based system to detection and prediction of heart disease from IoT dataset. This system can able to detect the disease and predict accordingly.

Fuzzy Keyword Search

The fuzzy keyword search strategy is based on pre-defined keywords, and the server should return files that include the term. The search results are returned using fuzzy keyword schemes based on the user's search input. If the user's search input perfectly matches one of the pre-set keywords, the system will automatically retrieve documents that contain that term. Otherwise, if the searching input contains errors as well as design abnormalities, the server will return the closest possible results based on pre-determined comparability semantics. The distance between two words is used to calculate the pre-determined comparability semantics for the erroneous searching phrase and the proper intended keyword. The amount of operations required to turn one word into the other is used to calculate the distance between two words. Substitution, deletion, and insertion are the three basic operations. To adjust the distance, two methods are used: straight forward and

| e-ISSN: 2319-8753, p-ISSN: 2320-6710| www.ijirset.com | Impact Factor: 7.512|

|| Volume 10, Issue 6, June 2021 ||

| DOI:10.15680/IJIRSET.2021.1006432|

wild-card based. Edit distance is calculated using wild card fuzzy sets in a wild-card based technique. The asterisk is used in Wild-Card search to denote one or more characters. Because all types of keywords must be included in order to compute the distance between two words, fuzzy search requires a vast amount of storage to keep data. This is one of the most significant disadvantages of fuzzy search. Wild-Card has the advantage of lowering the size of fuzzy keyword sets.

2.3 Conjunctive Keyword Search

The goal of a conjunctive keyword search (CKS) is to enable the user to ask a conjunctive keyword query to the server, so that the server can test which encrypted data contains the conjunctive keyword. In CKS system, data and keywords are firstly encrypted and then uploaded into the storage system, hence the server cannot see the data and keywords

2.4 Multi-keyword Ranked Search

Ranked search, over outsourced cloud data, facilitates searching settings of single keyword and multi-keyword search. This scheme supports only single keyword search. It uses the order preserving symmetric encryption, and gives each keyword a weight by. Cloud server rank relevant data files doesn't have knowledge about the specific keyword weight. Multi-keyword search over encrypted cloud data (MRSE) is firstly defined in 2011, when N.cao et al. made the first attempt to define and solve multi-keyword ranked query problem (kNN). In N.cao et al, data owner first defines a set of keywords and builds a dictionary containing them. The dictionary contains an index vector built for each keyword. If the file in dataset contains a keyword, the element is set to , otherwise, is set to . To execute multi keyword ranked query , firstly a trapdoor to equerry keywords set is sent to cloud server provider (CSP). CSP uses inner product between the trapdoors and to determine the similarity between them. Finally, the first results with the highest scores are returned to user. But in this scheme, position of each keyword in the dictionary is fixed, so it must rebuilt it when the number of keywords increased.

2.5 Semantic Search

Semantic search means retrieving of synonyms of query keywords, where this keywords are a predefined. In semantic search the keywords extracted from outsourced text documents are extended by its common synonyms, for examples, the synonym of the keyword “journey” is “travel” or “trip”, these keywords are totally different in spelling. Semantic search is used into knowing search engine Wikipedia, where there are three schemes was developed to improved search over it such as, Synonym-keyword search (SBKS): Wikipedia-based keyword search, and Wikipedia-based synonym keyword search.

III. COMPARISON

In this section we make a comparison between searchable encryption techniques according to this features:

1. Multiple keyword: Which search algorithms allow you to search through data for more than one term, reducing the amount of search results?

2. Amount of Scale Data: This refers to the type of search algorithm that should be used when searching a large amount of data. It's not a standard phrase, but it's often used to describe data that expands enormously in size over time.

3. Algorithmic complexity is concerned with how quickly or slow a particular algorithm runs, while time complexity is the computational complexity that specifies how long it takes to run an algorithm. The quickest algorithms have a high level of complexity.

4. Accuracy: This refers to the approximation ratio of the retrieved results from the desired search. Some searching algorithms achieve great accuracy for the intended target, while others do not, and others have a satisfactory accuracy. The proximity of the outcomes to the known or true results is referred to as accuracy.

5. Large storage requirements (Storage): Storage provides a straightforward means of maintaining, managing, backing up, and making data accessible to users. It is a key component of each search technique's desired requirement. Some algorithms are dependent on the quantity of files saved on the server; others require a huge amount of storage, while others require less.

| e-ISSN: 2319-8753, p-ISSN: 2320-6710| www.ijirset.com | Impact Factor: 7.512|

|| Volume 10, Issue 6, June 2021 ||

| DOI:10.15680/IJIRSET.2021.1006432|

Table.1. A comparison between searching techniques types

	Single keyword search	Ranked keyword search	Fuzzy keyword search	Multiple-keyword search	Semantic search
Multiple keyword.	Not support	support	support	support	support
Amount of Scale data.	Suitable	Suitable	Not suitable	Not suitable	Not suitable
Search complexity	O(N)	O(log N) where Mis domain score of keyword W	O(1) search oneach document	O(mn) Where mdocuments and nunique terms.	O(rn) Where r number of files and nnumber of distinct keywords.
Accuracy	Not accurate	High accuracy	Acceptable	Acceptable	Less accuracy
Storage	Less	Not large	Large	Less	Large

Table.1 summarised the comparison of the preceding methods using the metrics listed above. As can be seen, the above-mentioned searching algorithms have been offered for conducting searches over encrypted data stored in the cloud. Eachof these strategies has its own set of benefits and drawbacks. For example, the single keyword search technique improves search result relevance while excluding multiple keyword searches. Ranking keyword search strategies, on the other hand, have a high level of search accuracy for a wide range of applications. However, the most significant disadvantage is the extensive post-processing of encrypted information. Also, fuzzy keyword search approaches improve search effectiveness, so the distance can be adopted, but they don't solve the ranked search problem and demand a lot of storage.
Multiple-keyword search approaches, on the other hand, improve search accuracy. However, it is not appropriate for large-scale data. It is also more efficient and practical in terms of semantic search strategies, and it performs better in terms of search quality. However, index creation takes a long time.

IV. CONCLUSIONS

In this work, encryption strategies relating to search-based file retrieval from outsourced encrypted data were thoroughly examined. Many searchable strategies based on single keyword, multiple keyword search, and ranking, as well as fuzzy tolerance, have been investigated. The ultimate goal is to provide search semantics while maintaining privacy. The use of rank-based data retrieval has been considered, which allows quick search access and is thought to be the most efficient method for searching encrypted data because only related files are returned.

REFERENCES

[1] S. Zhang, X.Chen, and S.Wu: Analysis and Research of Cloud Computing instance, second International Conference on Future Network, China, IEEE, pp. 88-92, (2010).
[2] C.Gong , J.Liu , Q.Zhang , H.Chen , and Z.Gong :The Characteristics of Cloud Computing, 39th International Conference on Parallel Processing Workshops,pp. 275-279,(2010).
[3] S. Zhang, X.Chen, and S.Wu: Analysis and Research of Cloud Computing instance, second International Conference on Future Network, China, IEEE, pp. 88-92, (2010).
[4] C.Gong , J.Liu , Q.Zhang , H.Chen , and Z.Gong :The Characteristics of Cloud Computing, 39th International Conference on Parallel Processing Workshops,pp. 275-279,(2010).
[5] D.Song, D.Wagner, and A.Perrig :Practical Techniques for Searches on Encrypted Data, In Proc. Of IEEE Symposium on Security and Privacy ,pp. 44-55,(2000).
[6] E-J.Goh :Secure indexes, In Cryptology ePrint Archive on October 7th, pp. 1-18,(2003).
[7] D.Boneh, G.Di Crescenzo, R.Ostrovsky, and G.Persiano :Public Key Encryption with Keyword Search, In Proc. Of Euro Crypto'04, pp. 506-522,(2004).
[8] G.Duntao, H.Dawei, C.Haibin, and Y.Xiaoyuan :A new Public Key Encryption with Temporary Keyword Search, In International Conference on computer, Mechatronics, Control and Electronic Engineering(CMCE), (volume:4), pp.80-83,(2010).

| e-ISSN: 2319-8753, p-ISSN: 2320-6710| www.ijirset.com | Impact Factor: 7.512|

|| Volume 10, Issue 6, June 2021 ||

| DOI:10.15680/IJIRSET.2021.1006432|

[9] R.Curtmola, J.A.Garay, S.Kamara, and R.Ostrovsky :Searchable Symmetric Encryption: improved definitions and efficient constructions, In Proc. Of ACM CCS, pp. 79-88,(2006).
[10] Q.Liu, G.Wang, J.Wu :Secure and Privacy Preserving Keyword Searching for Cloud Storage Services , Journal of Network and Computer Applications, (volume:35),pp 927-933,(2011).
[11] Monika D.Rokade ,Dr.Yogesh kumar Sharma,"Deep and machine learning approaches for anomaly-based intrusion detection of imbalanced network traffic." IOSR Journal of Engineering (IOSR JEN), ISSN (e): 2250-3021, ISSN (p): 2278-8719
[12] Monika D.Rokade ,Dr.Yogesh kumar Sharma"MLIDS: A Machine Learning Approach for Intrusion Detection for Real Time Network Dataset", 2021 International Conference on Emerging Smart Computing and Informatics (ESCI), IEEE
[13] Monika D.Rokade, Dr. Yogesh Kumar Sharma. (2020). Identification of Malicious Activity for Network Packet using Deep Learning. *International Journal of Advanced Science and Technology*, *29*(9s), 2324 - 2331.
[14] Sunil S.Khatal ,Dr.Yogesh kumar Sharma, "Health Care Patient Monitoring using IoT and Machine Learning.", **IOSR Journal of Engineering (IOSR JEN),** ISSN (e): 2250-3021, ISSN (p): 2278-8719
[15] Sunil S.Khatal ,Dr.Yogesh kumar Sharma, "Data Hiding In Audio-Video Using Anti Forensics Technique ForAuthentication ", IJSRDV4I50349, Volume : 4, Issue : 5
[16] Sunil S.Khatal Dr. Yogesh Kumar Sharma. (2020). Analyzing the role of Heart Disease Prediction System using IoT and Machine Learning. *International Journal of Advanced Science and Technology*, *29*(9s), 2340 - 2346

INNO SPACE
SJIF Scientific Journal Impact Factor

Impact Factor:
7.512

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

निस्केयर
NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

9940 572 462 6381 907 438 ijirset@gmail.com

www.ijirset.com

Scan to save the contact details