www.ijprems.com
editor@ijprems.com

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)

Vol. 03, Issue 04, April 2023, pp : 22-27

e-ISSN : 2583-1062

Impact Factor : 5.725

AN EFFICIENT MULTI-KEYWORD RANKED SEARCH OVER ENCRYPTED DATA USING EDGE COMPUTING

B. Arunthamizharasi[1], Dr. T. Kirubadevi[2], Dr. S. Geetha[3]

[1]Master of Technology, Department of Computer Science and Engineering, Dr. M.G.R Educational and Research Institute, Chennai 600 095, Tamilnadu, India.

[2]Assistant Professor, Department of Computer Science and Engineering, Dr. M.G.R Educational and Research Institute, Chennai 600 095, Tamilnadu, India.

[3]Professor, Department of Computer Science and Engineering, Dr. M.G.R Educational and Research Institute, Chennai 600 095, Tamilnadu, India.

ABSTRACT

Cloud computing encourages data owners to shift their sophisticated data management systems to commercial public clouds for greater flexibility and cost savings. For privacy, sensitive data must be encrypted before outsourcing, making plaintext keyword search obsolete. Hence, encrypted cloud data search is needed. Due to the enormous quantity of cloud data users and documents, the search service must enable multi-keyword queries and result similarity ranking. Searchable encryption studies usually focus on Boolean keyword searches or single-term searches without distinguishing search results. This research defines and solves the tough challenge of privacy-preserving multi-keyword ranking ontology keyword mapping and search over encrypted cloud data for the first time (BKCM). For a secure cloud data consumption system, we set strict privacy standards. "Boolean keyword coordinate matching," or as many matches as feasible, captures the similarity between the search query and data documents. "Inner product similarity" quantifies this similarity principle. We start with a safe inner product computing-based BKCM system. This technology is greatly expanded to suit privacy needs in two threat situations. Tests on real-world datasets reveal that the recommended solutions do not considerably raise computation or transmission costs.

Keywords: Searchable encryption, multi-keyword ranked search, Boolean keyword coordinate matching.

1. INTRODUCTION

1.1 The Overall Concern for Privacy Preservation

As the volume of data increases at a rate that makes secrecy increasingly difficult, this issue has assumed greater significance. The internet had only recently emerged at the time of the competition. It has matured into something massive and global by now. Very big databases that recorded massive amounts of transactional data were made possible by rapid advances in networking, storage, and computing technologies, as the authors noted at the time. We prioritise your security when it comes to data mining techniques. This is a place where one's privacy could potentially be compromised. According to the report, there is a close relationship between data mining and data warehousing. Most tools function by collecting data in one location and then applying an algorithm to it. Yet, issues with data privacy may make it difficult to construct a centralised data repository. Many persons might have copies of the same data, but no copy could be transferred to another location. Data mining algorithms generate knowledge, and the outcomes of data mining rarely compromise privacy because they often reveal abstract concepts rather than specifics. But privacy advocates have reason to be concerned, as pooling data for pattern mining simplifies illegal activities. The issue is in the implementation of data mining, rather than the concept itself.

1.2 Sharing Private Data Through Cloud Datamining

Due to the ever-increasing data volume, more and more people are storing their information in the cloud. Our outsourced data may be at risk, though, because cloud customers and cloud servers are not in the same trusted domain. In order to prevent unauthorised access and maintain the confidentiality of personal information, data encryption is a prerequisite for storing data on the cloud. Cloud-based encrypted data cannot be directly searched using common plaintext search techniques. Conventional IR provided data consumers with features like multi-keyword ranking, keyword mapping, and search. The cloud server must provide the same job while safeguarding the confidentiality of the user's data and search queries. The value of storing information in the cloud depends on how easily retrievable it is. According to the study's findings, searchable encryption methods can provide a secure means for users to access and browse encrypted material.

1.3 Motivation

Cloud security is important for both business and personal users. Everyone wants to know that their information is safe and secure. Businesses have legal obligations to keep client data secure, with certain sectors having more stringent rules about data storage. To prevent unauthorized access to our data we need to provide some security mechanism to our data. Now days the Third-party cloud service providers are increasing very fast rate uploading or using their services may lead to misuse of our data (e.g., Balance sheet, Employee details). To provide security to such important documents and data is our motivation behind this paper.

1.4 Problem Definition and Objectives

A general approach to protect the data confidentiality is to encrypt the data before outsourcing. Searchable encryption schemes enable the client to store the encrypted data to the cloud and execute keyword search over cipher text domain. So far, abundant works have been proposed under different threat models to achieve various search functionality, such as single keyword search, similarity search, multi-keyword Boolean search, ranked search, multi-keyword ranked search, etc. Among them, multi-keyword ranked search achieves more and more attention for its practical applicability. We define and solve the problem of effective yet secure ranked keyword search over encrypted cloud data while also encrypting keyword to be searched for providing more security.

1.5 Scope and Limitations

- The proposed system can be used to enhance the security mechanism in FSS (File Sharing Systems).
- The keyword search functionality is performed over encrypted cloud data without leaking any information about the search keywords by encrypting the Keywords.
- We can improve the security mechanism by different dynamic encryption algorithms.
- The files won’t be shared to anyone unless and until the user (Data owner) grants the access using security verification key.
- We can improve by encrypting multiple formats of file enhancing Multiple file sharing system.
- There are also, however, limitations when it comes to this technology like large files have problem to be encrypted.

2. LITERATURE REVIEW

2.1 Keyword Searches on Remote Encrypted Data

Protecting data confidentiality while maintaining the anonymity of search terms A remote file server is where user would like to store encrypted files. In the future, User might want to decrypt files with specific keywords. To read old emails from Yahoo or another big provider's server from a mobile device while on the go, for instance, a user may encrypt them and upload them to the cloud. handles this issue within well-defined security constraints. The methods are effective without the usage of a public-key cryptosystem. The process is unaffected by the data encryption method used for remote access. They, too, are making progress. User can accomplish this by contributing fresh data that cannot be indexed by prior searches but can be indexed by subsequent ones. What's most important to remember is that information can be kept on remote servers and accessed from anywhere through mobile device, laptop, or other device.

2.2 Storage in the Cryptographic Cloud

The public cloud has many advantages, but there are also significant security and privacy concerns. Concerns about data security and privacy likely to be the biggest barrier to widespread use of cloud storage and cloud computing. Provides a brief overview of the benefits of using a cryptographic storage service, including helping customers and cloud providers stay in compliance with the law. Furthermore, a cryptographic storage service can be used as the foundation for other cloud services, such as secure backups, archives, health record systems, safe data transfer, and e-discovery.

2.3 Effective and Safe Multi-Keyword Search on Cloud Data

One problem is that customers who aren't familiar with the encrypted cloud data will need to process every file they download in order to find the ones that are most relevant to them. Yet in the current pay-as-you-go cloud model, it's undesirable to use more network bandwidth to get all files that match the query keyword. In this research, we identify and solve the issue of how to efficiently and securely conduct keyword-ranked searches on encrypted cloud data. Ranked ontology keyword mapping and search enhances system usability and moves us closer to the widespread adoption of privacy-preserving data storage services in Cloud Computing by delivering matched files in ranked order depending on provided relevant criteria (such as keyword frequency). Privacy-preserving multi-keyword ranking ontology keyword mapping and search over encrypted cloud data is defined and solved for the first time in this study (BKCM). Further, it lays forth a set of strict privacy rules that must be completed before a safe cloud data use system can be put into place. The proposed ranking system efficiently returns documents that are highly relevant to the search criteria supplied. The proposed ranking approach is used by our system to improve the security of data stored in the cloud. Security and Privacy in the Cloud Computing Environment: Privacy is a major issue in cloud computing and must be considered at every level of development to ensure user confidence and regulatory compliance.

2.4 Easy Fuzzy Keyword Search Over Encrypted Data in Cloud Computing

The primary goal of this research was to formally address the issue of efficiently doing fuzzy keyword searches over encrypted cloud data while protecting the privacy of the keywords being used. Our proposed system (the BKCM scheme) does a multi-keyword raking search, but this core concept is still employed. To solve the problem of dependability while providing nearly optimal overall performance, the author proposes creating a secure cloud storage service. Safe, scalable, and granular data access in the cloud: The challenge of achieving data privacy, scalability, and granularity in access control systems remains unanswered. Assigning most of the calculation tasks involved with fine-grained information access control to untrusted cloud servers without revealing the hidden information contents is one way the company addresses this challenging open issue, while characterising and maintaining access strategies based on information ascribes is another. The authors of this study propose a system of public audits to ensure the privacy of data stored in the cloud. To prevent the TPA from discovering sensitive information about the cloud server, it employs random masking and the homomorphic linear authenticator. This eliminates the time-consuming and perhaps expensive auditing burden that cloud users previously faced.

2.5 Semantic Multi-Keyword Ranked Ontology Keyword Mapping and Search Over Encrypted Cloud Data in an Efficient and Privacy-Protecting Manner

In light of the many advantages of cloud computing, this study addresses the rise in the number of data owners who are entrusting the cloud with their most private information. Cloud storage facilitates the storage of vast amounts of data, making it imperative to provide information users with keyword-based search services. Nevertheless, because sensitive data is usually encrypted before being delivered to a cloud server to protect privacy, plaintext search technologies are made useless. In this research, we provide a semantic multi-keyword ranking ontology keyword mapping and search technique for encrypted cloud data that meets a wide range of privacy requirements. To get things rolling, we'll employ a technique called "Latent Semantic Analysis" to discover hidden meaning in texts and words. Latent semantic analysis uses the unspoken higher-order structure ("semantic structure") between terms and texts to map the latter to a lower-dimensional vector space. As a result, the associations between words are recorded mechanically. Second, we implement a safe search method called "k-nearest neighbour" (k-NN). It's possible that the proposed system will not only return fileIs that contain the query keyword, but also files that contain phrases that are latently semantically linked to the query keyword. This is the experiment's conclusive finding. The preceding section provides a comprehensive demonstration of the proposed system with the exception of the KeyGen technique. Our approach uses Gauss-Jordan to generate the inverse matrix. Key generation time is proportional to the matrix size. The SVD algorithm also needs more time to run than the proposed solution. All of our other proposed algorithms, including index building, trapdoor generation, and query, are compatible with the original BKCM in terms of runtime.

3. MODELLING AND ANALYSIS

SYSTEM MODELS

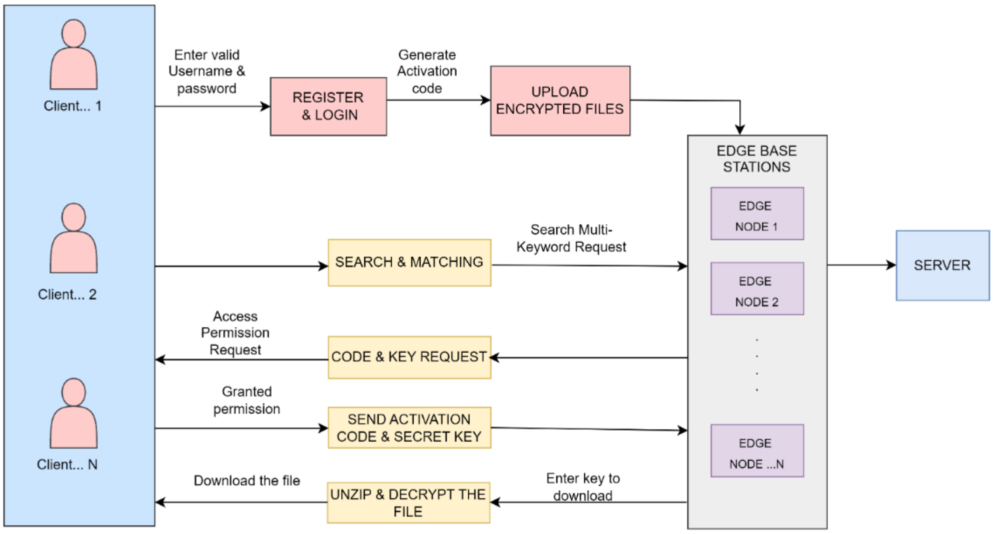

Figure 1. System Architecture

Symmetric searchable encryption (SSE) is an important area of research for searchable encryption. It has a low computational cost, a simplified algorithm, a fast speed, and is closer to the real-world application scene. The following is a classic SSE system:

- First, users pull keywords from local files to build the index. Then, they use the private key to encrypt the index and files, and then they upload them to the cloud server.
- People who have permission to ask questions use the private key to make a trapdoor for the keyword that needs to be asked about, which is then sent to the server. The trapdoor can't tell anything about what the keyword is. The easiest way to set a trap is to encrypt the keyword with the key.
- Once the server gets the trapdoor, it runs the retrieval algorithm. The inverted index structure is used by the retrieval algorithm to find the encrypted file name that matches the key word for the trapdoor. At the end, the user's file name is used to find the encrypted file. The server can only tell if a key word is in the trapdoor of an encrypted file.
- Users use the key to decrypt the encrypted file that the server sends back in order to get the results of their queries. In the process described above, the user can find out the query authority of the ciphertext by sharing the key and other means. The user's key should be kept hidden from third parties. The server doesn't know either the user's key or what's in the file. So, the user's files that are encrypted are safe.

3.1 Edge Cloud Configuration Module

This component enhances the approaches that permit multi-keyword queries and give result similarity rating for superior data retrieval, as opposed to just returning the results without any differentiation. Keeping data secure by stopping the edge cloud server from gaining knowledge from the indexed data. Efficiency: Little communication and processing overhead is required to meet the aforementioned functionality and privacy standards.

3.2 Coordinated Cipher Text Matching

Coordinate matching is a measure of intermediate similarity that uses the frequency with which query terms appear in a document to determine how relevant the document is to the query. Boolean searches work best when the user specifies the precise subset of the dataset that needs to be retrieved. The most pertinent publications can be retrieved in a prioritised list, and users can do so by selecting from a list of terms that suggests the user's concern.

3.3 Data Security

Before sending data to the cloud, the data's owner can protect it with encryption using standard symmetric key cryptography. If the cloud server can infer any connection between keywords and encrypted documents from the index, the index's privacy has been breached. Hence, a searchable index needs to be built to stop the cloud server from launching an association attack.

3.4 Keyword Privacy

Users prefer that their searches remain private, even from the cloud service they're using. Protecting the Keywords that will lead them to the correct trapdoor is of utmost importance. The trapdoor could be built using cryptography to ensure the safety of the search terms.

DESIGN GOALS

The proposed multi-keyword ranked search scheme should achieve the following design goals:

Multi-Keyword Ranked Search Over Encrypted Data for Multi-Data-Owner

The scheme should be able to use encrypted multi-keywords queries to calculate ranked search results over encrypted documents stored by different data owners.

Scalability and Efficiency

The scheme should be able to search a large number of encrypted documents in short time to quickly respond to user’s queries. In addition, the size of the trapdoors should be acceptable to reduce the communication overhead.

Index and Trapdoor Confidentiality

The cloud server should not be able to learn any useful information about the stored documents or trapdoors. In addition, analysing the statistical information of the documents from the same domain should not help the server to identify any specific keyword in the query or the index.

Trapdoors Unlikability

Given two trapdoors, the server should not ascertain whether the two trapdoors have the same set of keywords.

PERFORMANCE ANALYSIS

Precision

Precision is known as positive predictive value. It is defined as the number of correct results divided by the number of the retrieved result. Precision returns the exactness (accurate) of the result. If the threshold value changes the precision may increase or decrease accordingly.

$\text{Precision}=\sum TP / (\sum TP + \sum FP)$

where,

- TP = True positive, total number positive reviews which is positive

- FP = False positive, total number negative reviews which is positive

Recall

Recall is known as true positive value or sensitivity. It is defined as the number of correct results divided by the number of the relevant result. Recall returns the completeness of the result but doesn't depend on any threshold value. The recall value is calculated after finding the new rating, to check whether the user had given the positive or negative ratings correctly.

Recall = $\sum TP / \sum P$

Where,

- TP = True positive, total number of true positive conditions ranked.
- P = Total number of correctly ranked.

Accuracy

The Accuracy or recognition rate measures the proportion of true results to the total number of the instances in the dataset. It specifies how many instances correctly predicted. This affects the system performance. Misclassification measures the error rate (ie) the proportion of false results to the total number of the instances for the given dataset. It specifies how many instances falsely predicted. This affects the system performance. Accuracy= $(\sum TP+ \sum TN)/(\sum P + \sum N)$

Where,

- TP=True positive, total number of true positive reviews is classified as positive.
- TN= True negative, total number of true negative reviews is classified as negative.
- P = Total number of correctly ranked.

Table 1. Calculated Precision, Recall and Accuracy values

No. of Documents	50	100	150	200	250
Recall	0.9	0.98	0.95	0.8	0.85
Precision	0.8	0.75	0.9	0.6	0.8
Accuracy	0.97	0.96	0.95	0.89	0.88

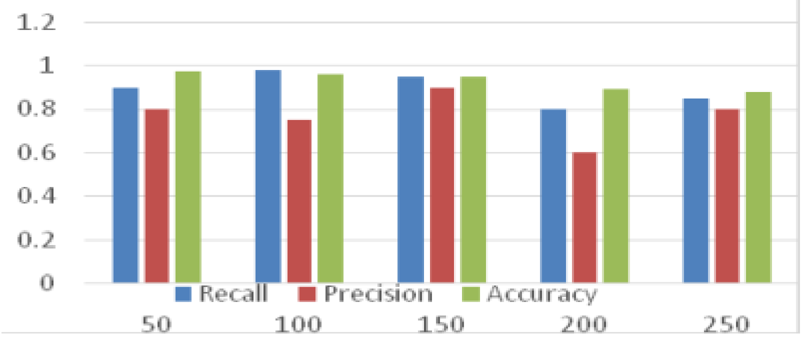

Figure 2. Comparison of Precision, Recall and Accuracy

4. CONCLUSION AND FUTURE WORK

In this work, we create a number of privacy restrictions and describe the first solution to the problem of multi-keyword ranking ontology keyword mapping and search over encrypted cloud data. To effectively capture the similarity between query keywords and outsourced documents, we select the principle of "coordinate matching," i.e., as many matches as possible, from among various multi-keyword semantics, and we use "inner product similarity" to quantitatively formalise such a principle for similarity measurement. We first describe a simple BKCM approach based on safe inner product computation to answer the challenge of providing multi-keyword semantic without compromising privacy, and then we significantly improve it to meet privacy criteria in two threat situations. Tests on real-world datasets show that our proposed approaches impose little cost on both computation and communication, and we conduct a thorough analysis of the privacy and efficiency guarantees provided by the suggested schemes. In order to create a more robust threat model, we plan to look into supporting alternative multi-keyword semantics (such as weighted query) over encrypted data, verifying the integrity of the search result's rank order, and providing privacy guarantees.

5. REFERENCES

[1] Jintao Ling, Huixian Gu, Haijiang Wang, and Lei Zhang, “An Efficient Ciphertext Index Retrieval Scheme Based on Edge Computing Framework”, Natural Science Foundation of Zhejiang Province under Grant LQ20F020010, 2021.

[2] S. Kamara and K. Lauter, “Cryptographic cloud storage,” in RLCPS, January 2010, LNCS. Springer, Heidelberg.

[3] D. Song, D. Wagner, and A. Perrig, “Practical techniques for searches on encrypted data,” in Proc. of S&P, 2000.

[4] Y.-C. Chang and M. Mitzenmacher, “Privacy preserving keyword searches on remote encrypted data,” in Proc. of ACNS, 2005.

[5] R. Curtmola, J. A. Garay, S. Kamara, and R. Ostrovsky, “Searchable symmetric encryption: improved definitions and efficient constructions,” in Proc. of ACM CCS, 2006.

[6] D. Boneh, G. D. Crescenzo, R. Ostrovsky, and G. Persiano, “Public key encryption with keyword search,” in Proc. of EUROCRYPT, 2004.

[7] M. Bellare, A. Boldyreva, and A. ONeill, “Deterministic and efficiently searchable encryption,” in Proc. of CRYPTO, 2007.