ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com
International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 3, Special Issue 12, March 2016

SECRECY PRESERVING MULTI - KEYWORD RANKING BASED EXPLORATION ON ENCRYPTED CLOUD DATA

Nandhini.R [1], Matheswaran.V [2]

1. P.G. Student, Dept. of MCA, VSB Engineering College, Karur, Tamilnadu, India
2. Asst.Professor, Dept. of MCA, VSB Engineering College, Karur, Tamilnadu, India

Abstract **The advent of cloud computing, data owners are motivated to outsource their complex data management systems from local sites to commercial public cloud for great flexibility and economic savings. But for protecting data privacy, sensitive data has to be encrypted before outsourcing, which obsoletes traditional data utilization based on plaintext keyword search. Thus, enabling an encrypted cloud data search service is of paramount importance. Considering the large number of data users and documents in cloud, it is crucial for the search service to allow multi-keyword query and provide result similarity ranking to meet the effective data retrieval need. Related works on searchable encryption focus on single keyword search or Boolean keyword search, and rarely differentiate the search results. In this paper, for the first time, we define and solve the challenging problem of privacy-preserving multi-keyword ranked search over encrypted cloud data (MRSE), and establish a set of strict privacy requirements for such a secure cloud data utilization system to become a reality. In this paper, we propose the problem of Secured Multi-keyword search (SMS) over encrypted cloud data (ECD), and construct a group of privacy policies for such a secure cloud data utilization system. From number of multi-keyword semantics, we select the highly efficient rule of coordinate matching, i.e., as many matches as possible, to identify the similarity between search query and data , and for further matching we use inner data correspondence to quantitatively formalize such principle for similarity measurement. We first propose a basic Secured multi keyword ranked search scheme using secure inner product computation, and then improve it to meet different privacy requirements. The Ranked result provides top k retrieval results.**

Keywords — **Encryption, Inner product similarity, Multi-keyword search, ranking.**

I. INTRODUCTION

Due to the rapid expansion of data, the data owners tend to store their data into the cloud to release the burden of data storage and maintenance [1]. However, as the cloud customers and the cloud server are not in the same trusted domain, our outsourced data may be under the exposure to the risk. Thus, before sent to the cloud, the sensitive data needs to be encrypted to protect for data privacy and combat unsolicited accesses. Unfortunately, the traditional plaintext search methods cannot be directly applied to the encrypted cloud data any more. The traditional information retrieval (IR) has already provided multi-keyword ranked search for the data user. In the same way, the cloud server needs provide the data user with the similar function, while protecting data and search privacy. It is meaningful storing it into the cloud server only when data can be easily searched and utilized.

Cloud computing saves money that users spend on annual or monthly subscription. Due to advantage of cloud services, more and more sensitive information are being centralized into the cloud servers, such as emails, personal health records, private videos and photos, company finance data,

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com
International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 3, Special Issue 12, March 2016

government documents, etc. To protect data privacy, confidential data has to be encrypted before outsourcing, so as to provide end-to-end data confidentiality assurance in the cloud. Data encryption makes effective data utilization a very challenging task given that there could be a large amount of outsourced data files. Besides, in Cloud Computing, data owners may share their outsourced data with a large number of users, who might want to only retrieve certain specific data files they are interested in during a given session. One of the most popular ways to do so is through keyword-based search. This keyword search technique allows users to selectively retrieve files of interest and has been widely applied in plaintext search scenarios. Unfortunately, data encryption, which restricts user's ability to perform keyword search and further demands the protection of keyword privacy, makes the traditional plaintext search methods fail for encrypted cloud data. Ranked search greatly improves system usability by normal matching files in a ranked order regarding to certain relevance criteria (e.g., keyword frequency).

II. BACKGROUND AND RELATED WORK

Organizations, companies store more and more valuable information is on cloud to protect their data from virus, hacking. The benefits of the new computing model include but are not limited to: relief of the trouble for storage administration, data access, and avoidance of high expenditure on hardware mechanism, software, etc. Ranked search improves system usability by normal matching files in a ranked order regarding to certain relevance criteria (e.g., keyword frequency),As directly outsourcing relevance scores will drips a lot of sensitive information against the keyword privacy, We proposed asymmetric encryption with ranking result of queried data which will give only expected data. In the literature, searchable encryption techniques [2-4] are able to provide secure search over encrypted data for users. They build a searchable inverted index that stores a list of mapping from keywords to the corresponding set of files which contain this keyword. When data users input a keyword, a trapdoor is generated for this keyword and then submitted to the cloud server Some researchers study the problem on secure and ranked search over outsourced cloud data. Wang *et al.*, [5] propose a secure ranked keyword search scheme. Their solution combines inverted index with order-preserving symmetric encryption (OPSE). In terms of ranked search, the order of retrieved files is determined by numerical relevance scores, which can be calculated by $TF \times IDF$. The relevance score is encrypted by OPSE to ensure security. It enhances system usability and saves communication overhead. This solution only supports single keyword ranked search. Cao *et al.*, [6] propose a method that adopts similarity measure of "coordinate matching" to capture the relevance of files to the query. They use "inner product similarity" to measure the score of each file. This solution supports exact multi-keyword ranked search. It is practical, and the search is flexible. Sun *et al.*, [7] proposed a MDB-tree based scheme which supports ranked multi-keyword search. This scheme is very efficient, but the higher efficiency will lead to lower precision of the search results in this scheme. In addition, fuzzy keyword search [8-10] have been developed. These methods employ a spell-check mechanism, such as, search for "wireless" instead of "wireiess", or the data format may not be the same *e.g.*, "data-mining" versus "datamining. Chuah *et al.*, [8] propose a

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com
International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 3, Special Issue 12, March 2016

privacy-aware bed-tree method to support fuzzy multi-keyword search. This approach uses edit distance to build fuzzy keyword sets. Bloom filters are constructed for every keyword. Then, it constructs the index tree for all files where each leaf node a hash value of a keyword. Li *et al.*, [9] exploit edit distance to quantify keywords similarity and construct storage-efficient fuzzy keyword sets. Specially, the wildcard-based fuzzy set construction approach is designed to save storage overhead. Wang *et al.*, [10] employ wildcard-based fuzzy set to build a private trie-traverse searching index. In the searching phase, if the edit distance between retrieval keywords and ones from the fuzzy sets is less than a predetermined set value, it is considered similar and returns the corresponding files. These fuzzy search methods support tolerance of minor typos and format inconsistencies, but do not support semantic fuzzy search. Considering the existence of polysemy and synonymy [11], the model that supports multi-keyword ranked search and semantic search is more reasonable. Christo Ananth et al. [16] proposed a system in which the complex parallelism technique is used to involve the processing of Substitution Byte, Shift Row, Mix Column and Add Round Key. Using S- Box complex parallelism, the original text is converted into cipher text. From that, we have achieved a 96% energy efficiency in Complex Parallelism Encryption technique and recovering the delay 232 ns. The complex parallelism that merge with parallel mix column and the one task one processor techniques are used. In future, Complex Parallelism single loop technique is used for recovering the original message.

A. Existing system

Existing searchable encryption schemes allow a user to securely search over encrypted data through keywords without first decrypting it, these techniques support only conventional Boolean keyword search, without capturing any relevance of the files in the search result. When directly applied in large collaborative data outsourcing cloud environment, they go through following disadvantage.

Drawbacks of existing system

1. Single-keyword search without ranking

2. Boolean- keyword search without ranking

3. Single-keyword search with ranking

4. Do not get relevant data.

III. PROBLEM FORMULATION

A. Proposed system

For our system, we choose the principle of coordinate matching, to identify the similarity between search query and data documents.Specially,we use inner data correspondence, i.e., the number of query keywords appearing in a document, to evaluate the similarity of that document to the search query in coordinate matching principle. Each document is linked with a binary vector as a sub index where each bit represents whether corresponding keyword is contained in the document.[1] The search query is also described as a binary vector where each bit means whether corresponding keyword appears in this search request, so the similarity could be exactly measured by inner product of query vector with data vector. However, directly outsourcing data vector or query vector will violate index

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com
International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 3, Special Issue 12, March 2016

privacy or search privacy. To meet the challenge of supporting such multi-keyword semantic without privacy breaches, we propose a basic SMS scheme using secure inner product computation, which is adapted from a secure k-nearest neighbour (kNN) technique, and then improve it step by step to achieve various privacy requirements in two levels of threat models.

1) Showing the problem of Secured Multi-keyword search over encrypted cloud data

2) Propose two schemes following the principle of coordinate matching and inner product similarity.

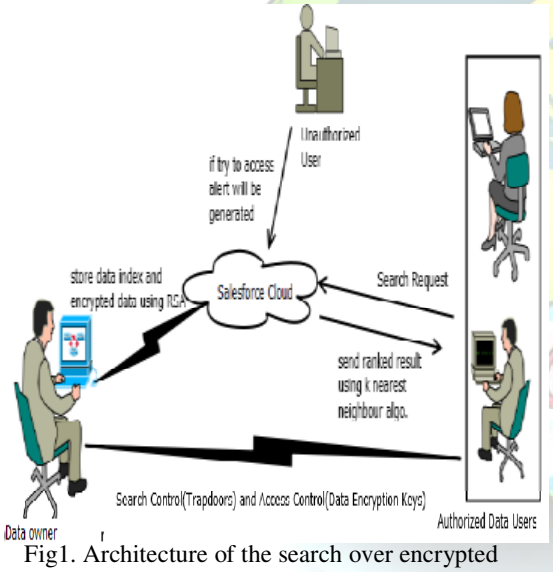

Fig1. Architecture of the search over encrypted cloud data

Considering three different entities, as illustrated in Fig1. Data owner, data user, and cloud server.

Data owner has a collection of data documents D = {$d1$, d 2 , ..., d m } .A set of distinct keywords W = {$w1$, $w2$, ..., wn } is extracted from the data collection D . The data owner will firstly construct an encrypted searchable index I from the data collection D . All files in D are encrypted and form a new file collection, C .Then, the data owner upload both the encrypted index I and the encrypted data collection$C$ to the cloud server.

Data user provides t keywords for the cloud server. A corresponding trapdoor w T through search control mechanisms is generated. In this paper, we assume that the authorization between the data owner and the data user is approximately done.

Cloud server received w T from the authorized user. Then, the cloud server calculates and returns to the corresponding set of encrypted documents. Moreover, to reduce the communication cost, the data user may send an optional number l along with the trapdoor$T$ so that the cloud server only sends back top- l files that are most relevant to the search query.

Data owner has a collection of data documents to be send to cloud server in the encrypted format. To activate the searching capability over encrypted data, data owner, before sending data, will first build an encrypted searchable manifestation (index), and then outsource both the index and the encrypted document collection to cloud server. To search the document, an authorized user require a corresponding trapdoor through search mechanisms, Upon receiving from data users, cloud server is responsible to search the index and return the corresponding set of encrypted documents. To improve document retrieval accuracy, search result should be ranked by cloud server according to some ranking criteria. Cloud server only sends back top-k documents that are most relevant to the search query. In Fig1.There is one another entity is shown i.e. Unauthorized User. If that Unauthorized user tries to access any data from clod then alert will be generated in the form of mail and message. The alert is given to the authorized person who is owner of that data.

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com
International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 3, Special Issue 12, March 2016

B. Design goals.

Data Owner Module

This module helps the owner to register those details and also include login details. This module helps the owner to upload his file with encryption using RSA algorithm. This ensures the files to be protected from unauthorized user.

Data User Module

This module includes the user registration login details. This module is used to help the client to search the file using the multiple key words concept and get the accurate result list based on the user query. The user is going to select the required file and register the user details and get activation code in mail email before enter the activation code. After user can download the Zip file and extract that file.

Encryption Module

This module is used to help the server to encrypt the document using RSA Algorithm and to convert the encrypted document to the Zip file with activation code and then activation code send to the user for download.

File Upload Module

This module helps the owner to upload his file with encryption using RSA algorithm. This ensures the files to be protected from unauthorized user.

Rank Search Module

These modules ensure the user to search the files that are searched frequently using rank search. This module allows the user to download the file using his secret key to decrypt the downloaded data. This module allows the Owner to view the uploaded files and downloaded files

File Download Module

This module allows the user to download the file using his secret key to decrypt the downloaded data.

View Uploaded and Downloaded File

This module allows the Owner to view the uploaded files and downloaded files

C. System Features

To activate ranked search for effective utilization of outsourced cloud data, our system design should simultaneously achieve security and performance guarantees as follows.

1. *Secured Multi-keyword Ranked Search*: To design search schemes which allow multi-keyword query and provide result similarity ranking for valuable data retrieval, instead of returning undifferentiated results.

2. *Privacy:* To prevent cloud server from learning additional information from dataset and index, and to meet privacy requirements.

3. *Effectiveness with high performance*: Above goals on functionality and privacy should be achieved with low communication and computation overhead.

4. Dynamic: The proposed scheme is designed to provide not only multi-keyword query and accurate result ranking, but also dynamic update on document collections.

5. Search Efficiency: The scheme aims to achieve sublinear search efficiency by exploring a special K-nearest neighbour efficient search algorithm.

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com
International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 3, Special Issue 12, March 2016

IV. ALGORITHMS USED

A .RSA Algorithm

This algorithm is used to encrypt n decrypt file contents. It is an asymmetric algorithm. The RSA algorithm involves three steps: key generation, encryption and decryption.

Key generation

RSA involves a **public key** and a **private key.** The public key can be known to everyone and is used for encrypting messages. Messages encrypted with the public key can only be decrypted using the private key. The keys for the RSA algorithm are generated the following way:

1. Choose two distinct prime numbers *a* and *b*.

2. Compute $n = ab$. n is used as the modulus for both the public and private keys

3. Compute $\varphi(n) = (a - 1)(b - 1)$, where φ is Euler's totient function.

4. Choose an integer e such that $1 < e < \varphi(n)$ and greatest common divisor of $(e, \varphi(n)) = 1$; i.e., e and $\varphi(n)$ are co prime. e is released as the public key exponent. having a short bit-length .

Encryption

Alice transmits her public key (n,e) to Bob and keeps the private key secret. Bob then wishes to send message **M** to Alice. He first turns **M** into an integer m, such that 0<m<n by using an agreed-upon reversible protocol known as a padding scheme. He then computes the cipher text C corresponding to

$$\mathbf{C = m^{e} \pmod n}$$

This can be done quickly using the method of exponentiation by squaring. Bob then transmits C to Alice. Note that at least nine values of m could yield a cipher text c equal to m, but this is very unlikely to occur in practice.

Decryption

Alice can recover M from C by using her private key exponent d via computing .

$$\mathbf{m = c^{d} \pmod n}$$

Given m, she can recover the original message **M** by reversing the padding scheme. (In practice, there are more efficient methods of calculating c^{d} using the pre computed values below.)

B. K-Nearest Neighbour

K-nearest neighbor search identifies the top *k* nearest neighbors to the query. This technique is commonly used in predictive analytics to estimate or classify a point based on the consensus of its neighbors. *K-nearest* neighbor graphs are graphs in which every point is connected to its *k* nearest neighbors. The basic idea of our new algorithm: The value of dmax is decreased keeping step with the ongoing exact evaluation of the object similarity distance for the candidates. At the end of the step by step refinement, dmax reaches the optimal query range Ed and prevents the method from producing more candidates than necessary thus fulfilling the r-optimality criterion.

Nearest Neighbor Search (q, k) // optimal algorithm

1. Initialize ranking = index.increm-ranking (F(q), df)
2. Initialize result = new sorted-list (key, object)
3. Initialize dmax = w
4. While o = ranking.getnext and d,(o, q) I d,,, do
5. If do@, s> s dmax then result.insert (d,(o, q) , o)
6. If result.length 2 k then dmax = result[k].key
7. Remove all entries from result where key > dmax
8. End while

Report all entries from result where key I dmax

V. EXPECTED RESULTS

1. Data Encryption and decryption Result

When RSA algorithm is applied on the data then we get encrypted data. and that encrypted data is

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com
International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 3, Special Issue 12, March 2016

store on the cloud. User can access the data after downloading and decrypting file. For encryption and decryption keys are provided.

2. Ranking Result

When any User request for the data then Ranking is done on requested data using k-nearest neighbor algorithm. For Ranking ―co-ordinate matching‖ principle is used. after ranking user gets the expected results of the query.

3. Alert System Results

If any unauthorized User tries to access or updating the data on cloud, then alert will be generated in the form of mail and messages. The alert intimates the authorized user.

VI. CONCLUSION AND FUTURE SCOPE

Thus we proposed the problem of multiple-keyword ranked search over encrypted cloud data, and construct a variety of security requirements. From various multi-keyword concepts, we choose the efficient principle of coordinate matching. We first propose secure inner data computation. Also we achieve effective ranking result using k-nearest neighbour technique. This system is currently work on single cloud, In future is will extended up to sky computing & Provide better security in multi-user systems.

References

[1] M. Armbrust, "A view of cloud computing", Communications of the ACM, vol. 53, no. 4, **(2010)**, pp. 50-58.

[2] D. Boneh, "Public key encryption with keyword search", Advances in Cryptology-Eurocrypt 2004, Springer, **(2004)**.

[3] R. Curtmola, "Searchable symmetric encryption: improved definitions and efficient constructions", Proceedings of the 13th ACM conference on Computer and communications security, ACM, **(2006)**.

[4] D. X. Song, D. Wagner and A. Perrig, "Practical techniques for searches on encrypted data. in Security and Privacy", 2000. S&P 2000, Proceedings 2000 IEEE Symposium, IEEE, **(2000)**.

[5] C. Wang, "Secure ranked keyword search over encrypted cloud data", Distributed Computing Systems (ICDCS), 2010 IEEE 30th International Conference, IEEE, **(2010)**.

[6] N. Cao, "Privacy-preserving multi-keyword ranked search over encrypted cloud data", INFOCOM, 2011 Proceedings IEEE, IEEE, **(2011)**.

[7] W. Sun, "Privacy-preserving multi-keyword text search in the cloud supporting similarity-based ranking", Proceedings of the 8th ACM SIGSAC symposium on Information, computer and communications security, ACM, **(2013)**.

[8] M. Chuah and W. Hu, "Privacy-aware bedtree based solution for fuzzy multi-keyword search over encrypted data", Distributed Computing Systems Workshops (ICDCSW), 2011 31st International Conference, IEEE, **(2011)**.

[9] S. Deshpande, "Fuzzy keyword search over encrypted data in cloud computing", World Journal of Science and Technology, vol. 2, no. 10, **(2013)**.

[10] C. Wang, "Achieving usable and privacy-assured similarity search over outsourced cloud data", INFOCOM, 2012 Proceedings IEEE, IEEE, **(2012)**.

[11] S. C. Deerwester, "Indexing by latent semantic analysis", JASIS, vol. 41, no. 6, **(1990)**, pp. 391-407.

[12] S. Zerr, "Zerber+ r: Top-k retrieval from a confidential index", Proceedings of the 12th International Conference on Extending Database Technology: Advances in Database Technology, ACM, **(2009)**.

[13] G. W. Furnas, "Information retrieval using a singular value decomposition model of latent

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com
International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 3, Special Issue 12, March 2016

semantic structure", Proceedings of the 11th annual international ACM SIGIR conference on Research and development in information retrieval, ACM, **(1988)**.

[14] W. K. Wong, "Secure kNN computation on encrypted databases", Proceedings of the 2009 ACM SIGMOD International Conference on Management of data, ACM, **(2009)**.

[15] C. Yang, "A Fast Privacy-Preserving Multi-keyword Search Scheme on Cloud Data", Cloud and Service Computing (CSC), 2012 International Conference, IEEE, **(2012)**.

[16] Christo Ananth, H.Anusuya Baby, "Encryption and Decryption in Complex Parallelism", International Journal of Advanced Research in Computer Engineering & Technology (IJARCET), Volume 3, Issue 3, March 2014,pp 790-795