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. 4, Issue 12, December 2016

A Survey on Secure and Dynamic Multi-Keyword Ranked Search Scheme over Encrypted Cloud Data

Ghansham R. Rathod, Prof. Sonali Patil

M. E. Student, Dept. of Computer, JSPM's BSIOTR, Wagholi, Pune, Maharashtra, India

Asst. Professor, Dept. of Computer, JSPM's BSIOTR, Wagholi, Pune, Maharashtra, India

ABSTRACT: Due to the increasing popularity of cloud computing, more and more data owners are motivated to outsource their data to cloud servers for great convenience and reduced cost in data management. However, sensitive data should be encrypted before outsourcing for privacy requirements, which obsoletes data utilization like keyword - based document retrieval. In this paper, we present a secure multi - keyword ranked search scheme over encrypted cloud data, which simultaneously supports dynamic update operations like deletion and insertion of documents. Specifically, the vector space model and the widely - used TF _ IDF model are combined in the index construction and query generation. We construct a special tree - based index structure and propose a "Greedy Depth - first Sea rch" algorithm to provide efficient multi - keyword ranked search. The secure K NN algorithm is utilized to encrypt the index and query vectors, and meanwhile ensure accurate relevance score calculation between encrypted index and query vectors. In order to resist statistical attacks, phantom terms are added to the index vector for blinding search results. Due to the use of our special tree - based index structure, the proposed scheme can achieve sub - linear search time and deal with the deletion and insertion of documents flexibly. Extensive experiments are conducted to demonstrate the efficiency of the proposed scheme.

KEYWORDS:multi-keyword ranked search, Searchable encryption,and cloud computing, dynamic update

I. INTRODUCTION

Currently we are in an information-explosion era where constantly purchasing new hardware, software and training IT professional is becoming a nightmare for almost every IT person. Coincidentally, we are witnessing an enterprise IT architecture which shifted to a centralized, more powerful computing paradigm known as Cloud Computing, in which enterprise's or personage's databases and applications are moved to the servers in the large data centers (i.e. the cloud) managed by the third-party cloud service providers (CSPs)a in the Internet. Cloud computing has been recognized as the most momentous turning point in the development of information technology during the past decade. People are attracted by the benefits it offers, such as personal and flexible access [1][2], on-demand computing resources configuration, considerable capital expenditure savings, etc. Therefore, many companies, organizations, and individual users have adopted the cloud platform to improve their business operations, research, or everyday needs. With the remunerative option of pay- as-you-use, general and private data are outsourced by many individual users and organizations to third party CSPs. A data owner can outsource their data to the cloud and either he can query on that outsourced data or can authenticate a client to perform query.

Cloud computing is a conversational phrase used to express a variety of dissimilar types of computing idea s that occupy large number of computers that are connected through a real - time communication network i.e Intern et . In science, cloud computing is the capability to run a program on many linked computers at the same time. The fame of the term can be recognized to its use in advertising to sell hosted services in the sense of application service provisioning that run client server software on a remote location. Cloud computing relies on sharing of resources to attain consistency and financial system alike to a utility (like the e electricity grid) over a network. The cloud also centre s on maximize the effectiveness of t he shared resources. Cloud resources are typically not only shared by

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. 4, Issue 12, December 2016

multiple users but as well as dynamically re-allocated as per demand. This can perform for assigning resources to users in dissimilar time zones. For example, a cloud computing service which serves American usersduring American business timings with a specific application (e.g. email) while the same resources are getting reallocated and serve Indian users during Indian business timings with another application (e.g. web server).

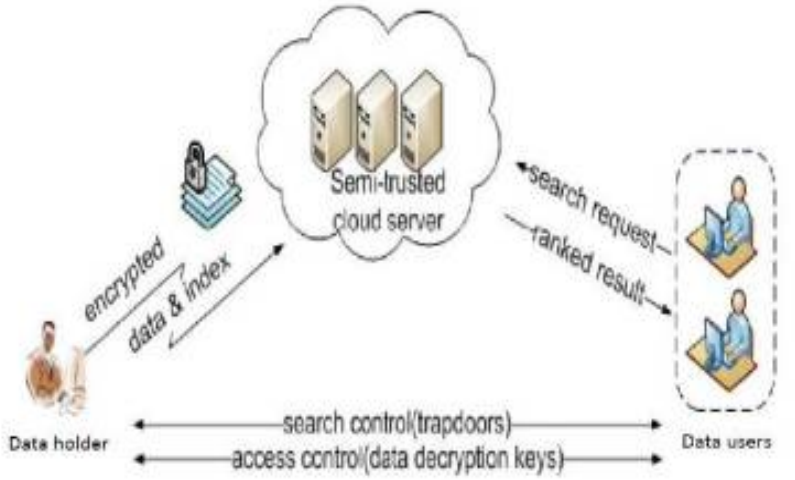

Figure 1: Architecture of the search over encrypted cloud data

This mechanism must take full advantage of the use of computing powers thus decreasing environmental damage as well, since less power, air conditioning an d so on, is necessary for the same functions. The expression "moving to cloud" also explains to an organization moving away from a traditional CAPEX model i.e buy the devoted hardware and decrease in value it over a period of time to the OPEX model i.e use a shared cloud infra structure and pay as you use it. Proponents maintain that cloud computing Permit Corporation to avoid direct infrastructure costs, and focus on projects that distinguish their businesses as an alternative of infrastructure. Proponents also maintain s that cloud computing permit scheme s to get their applications should run faster, with better manageability and less maintenance, and enable IT to more quickly adjust resources to meet random and changeable business demand.

The secure kNN algorithm is utilized to encrypt the index and query vectors, and meanwhile ensure accurate relevance score calculation between encrypted index and query vectors. To resist different attacks in different threat models, we construct two secure search schemes: the basic dynamic multi-keyword ranked search (BDMRS) scheme in the known ciphertext model, and the enhanced dynamic multi-keyword ranked search (EDMRS) scheme in the known background model. Our contributions are summarized as follows: 1) We design a searchable encryption scheme that supports both the accurate multi-keyword ranked search and flexible dynamic operation on document collection. 2) Due to the special structure of our tree-based index, the search complexity of the proposed scheme is fundamentally kept to logarithmic. And in practice, the proposed scheme can achieve higher search efficiency by executing our “Greedy Depth-first Search” algorithm. Moreover, parallel search can be flexibly performed to further reduce the time cost of search process.

II. RELATED WORK

1. Security challenges for the publiccloud [1] From This Paper I Referred-

Cloud computing represents today's most exciting computing paradigm shift in information technology [1]. However, security and privacy are perceived as primary obstacles to its wide adoption. Here, the author’s outline several critical securities challenges and motivate further investigation of security solutions for a trustworthy public cloud environment.

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. 4, Issue 12, December 2016

2. Cryptographic cloud storage, [2] From This Paper I Referred-

We consider the problem of building a secure cloud storage service on top of a public cloud infrastructure where the service provider is not completely trusted by the customer. Author [2][3] describes, at a high level, several architectures that combine recent and non-standard cryptographic primitives in order to achieve our goal. We survey the benefits such an architecture would provide to both customers and service providers and give an overview of recent advances in cryptography motivated specifically by cloud storage.

3. Software protection and simulationon oblivious rams[3] From This Paper I Referred-

Software protection is one of the most important issues concerning computer practice [3]. There exist many heuristics and ad-hoc methods for protection, but the problem as a whole has not received the theoretical treatment it deserves. In this paper author provide theoretical treatment of software protection. author reduce the problem of software protection to the problem of efficient simulation on oblivious RAM. A machine is oblivious if the sequence in which it accesses memory locations is equivalent for any two inputs with the same running time. For example, an oblivious Turing Machine is one for which the movement of the heads on the tapes is identical for each computation.

4.Publickey encryption with keyword search," [4] From This Paper I Referred-

We study the problem of searching on data that is encrypted using a public key system [4]. Consider user Bob who sends email to user Alice encrypted under Alice's public key. An email gateway wants to test whether the email contains the keyword "urgent" so that it could route the email accordingly. Alice, on the other hand does not wish to give the gateway the ability to decrypt all her messages. author define and construct a mechanism that enables Alice to provide a key to the gateway that enables the gateway to test whether the word "urgent" is a keyword in the email without learning anything else about the email. We refer to this mechanism as Public Key Encryption with keyword Search. As another example, consider a mail server that stores various messages publicly encrypted for Alice by others. Using our mechanism Alice can send the mail server a key that will enable the server to identify all messages containing some specific keyword, but learn nothing else. Authordefine the concept of public key encryption with keyword search and give several constructions.

5. Public key encryption that allows pir queries.[5]From This Paper I Referred

Consider the following problem: Alice wishes to maintain her email using a storage provider Bob (such as a Yahoo! or hotmail e-mail account) [5]. This storage-provider should provide for Alice the ability to collect, retrieve, search and delete emails but, at the same time, should learn neither the content of messages sent from the senders to Alice (with Bob as an intermediary), nor the search criteria used by Alice. A trivial solution is that messages will be sent to Bob in encrypted form and Alice, whenever she wants to search for some message, will ask Bob to send her a copy of the entire database of encrypted emails. This however is highly inefficient. We will be interested in solutions that are communication-efficient and, at the same time, respect the privacy of Alice. In this paper, we show how to create a publickey encryption scheme for Alice that allows PIR searching over encrypted documents. Our solution provides a theoretical solution to an open problem posed by Boneh, DiCrescenzo, Ostrovsky and Persiano on "Public-key Encryption with Keyword Search", providing the first scheme that does not reveal any partial information regarding user's search (including the access pattern) in the public-key setting and with non-trivially small communication complexity

III. PROPOSED ALGORITHM

A. *KEYWORD-NNE:*

In previous work, BKC algorithm drops its performance when the number of query keywords is increases. To solve this problem, here developed a more efficient keyword nearest neighbour expansion (keyword-NNE) which uses the different strategy. In this algorithm, one query is considered as a principal query keyword. Those objects are associated with principal query keyword are considered as principal objects. Keyword-NNE computes local best solution for each principal object. BKC algorithm returns the lbkc with having highest evaluation. For each of the principal object, its lbkc can be simply selects few closest and highly rated objects by the viewer/customer. Compared with the k-means clustering, the keyword covers significantly reduced. These keyword covers a further processe in

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. 4, Issue 12, December 2016

keyword-NNE-algorithm that will be optimal, and each keyword candidate covers processed generates very less new candidate keyword are covers.

- Algorithm to provide efficient multi-keyword ranked search.
- The secure kNN algorithm is utilized to encrypt the index and query vectors.
- Propose a "Greedy Depth-first Search" algorithm based on this index tree.
- Algorithm achieves better-than-linear search efficiency but results in precision loss.
- The LSH algorithm is suitable for similar search but cannot provide exact ranking.
- I′s ; ci} ← GenUpdateInfo (SK; Ts; i; up type)) This algorithm generates the update information {I′s ; ci} which will be sent to the cloud server.

a. DESIGN GOALS:

To enable secure, efficient, accurate and dynamic multi data under the above models, our system has the following Dynamic: The proposed scheme is designed to provide not only multi - keyword query and accurate result ranking, but also dynamic update on document collections. Search Efficiency: The scheme aims to achieve sublinear search efficiency by exploring a special tree - based index and an efficient search algorithm.

A.Privacy - preserving: The scheme is designed to prevent the cloud server from le arning additional information about the document collection, the index tree, and the query. The specific privacy requirements are summarized as follows,

B. Index Confidentiality and Query Confidentiality: The underlying plaintext information, including keywords in the index and query, TF values of keywords stored in the index, and IDF values of query keywords, should be protected from cloud server;

C. Trapdoor Unlinkability: The cloud server should not be able to determine whether two encrypted queries (trapdoors) are generated from the same search request;

D. Keyword Privacy: The cloud server could not identify the specific keyword in query, index or document collection by analyzing the statistical information like term frequency. Note that our proposed scheme is not designed to protect access pattern, i.e., the sequence of returned documents.

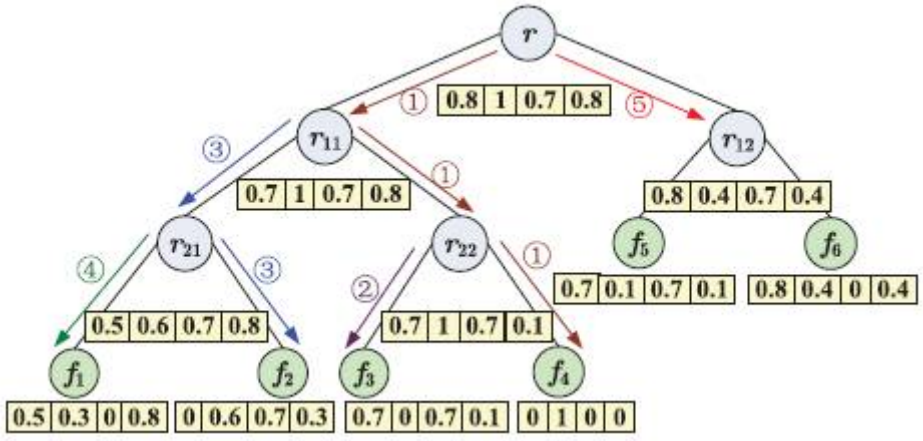

Figure 2. An example of the tree-based index with the document collection

We construct a special keyword balanced binary tree as the index, and propose a "Greedy Depth-first Search" algorithm to obtain better efficiency than linear search.

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. 4, Issue 12, December 2016

IV. SYSTEM ARCHITECTURE

A Secure and Dynamic Multi-keyword Ranked Search Scheme over Encrypted Cloud Data We construct a special tree-based index structure and propose a "Greedy Depth-first Search" algorithm to provide efficient multi-keyword ranked search. The proposed scheme can achieve sub-linear search time and deal with the deletion and insertion of documents flexibly.

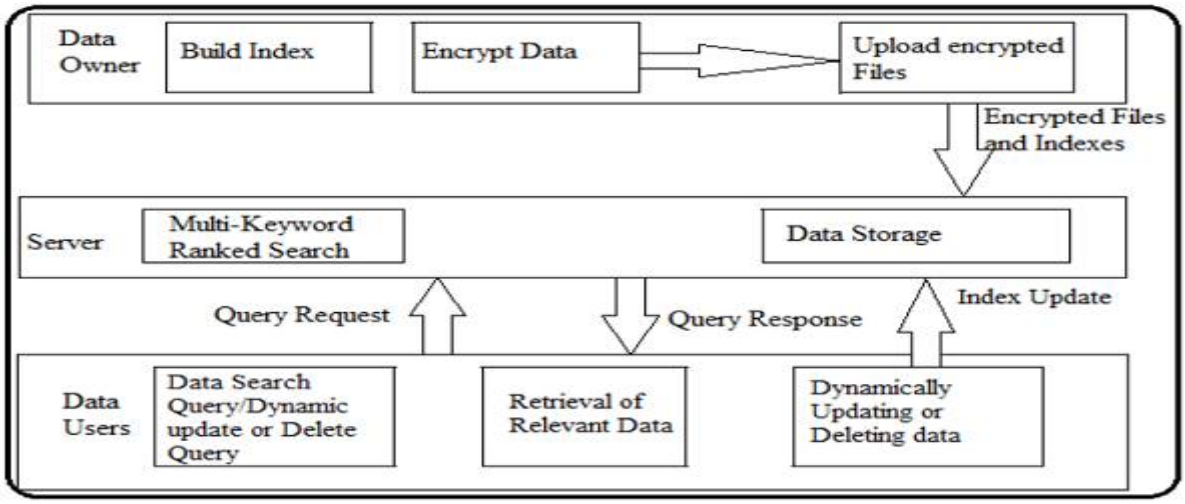

Figure 3: Block Diagram of proposed system

Extensive experiments are conducted to demonstrate the efficiency of the proposed scheme.

1. Abundant works have been proposed under different threat models to achieve various search functionality,
2. Recently, some dynamic schemes have been proposed to support inserting and deleting operations on document collection.
3. This paper proposes a secure tree-based search scheme over the encrypted cloud data, which supports multi keyword ranked search and dynamic operation on the document collection.

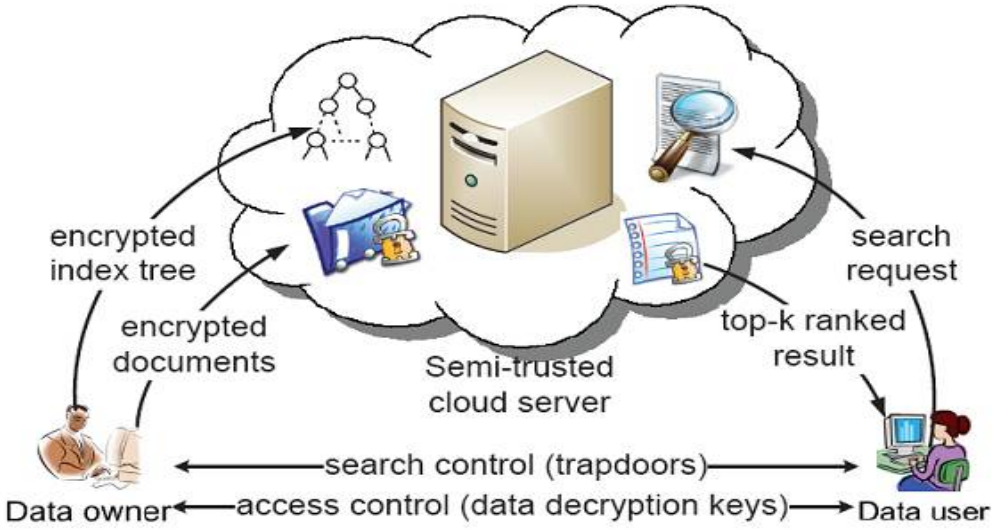

Figure 3. The architecture of ranked search over encrypted cloud data

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. 4, Issue 12, December 2016

Despite of the various advantages of cloud services, outsourcing sensitive information such as e-mails, personal health records, company finance data, government documents, etc.

V. CONCLUSION

In this survey paper, we have different kind of searching techniques for the encrypted data over cloud. A systematic study on the privacy and data utilization issues is covered here for various searching techniques. Some of the important issues to be handled by the searching technique for providing the data utilization and security are keyword privacy, Data privacy, Fine-grained Search, Scalability, Efficiency, Index privacy, Query Privacy, Result ranking, Index confidentiality, Query confidentiality, Query Unlink ability, semantic security and Trapdoor Unlink ability. The limitations for all the searching techniques mentioned in this paper are discussed as well. From the above survey, we can say that security can be provided by the Public-Key Encryption and data security cam be provided by some different methods like fuzzy keyword search or can provide binary balanced tree as an Index

REFERENCES

[1] K. Ren, C. Wang, and Q. Wang, "Security challenges for the public cloud," IEEE Internet Comput., vol. 16, no. 1, pp. 69–73, Jan-Feb. 2012.
[2] S. Kamara and K. Lauter, "Cryptographic cloud storage," in Proc. Financ. Cryptography Data Secur., 2010, pp. 136–149.
[3] C. Gentry, "A fully homomorphic encryption scheme," Ph.D. dissertation, Stanford Univ., Stanford, CA, USA, 2009.
[4] O. Goldreich and R. Ostrovsky, "Software protection and simulation on oblivious rams," J. ACM, vol. 43, no. 3, pp. 431–473, 1996.
[5] D. Boneh, G. Di Crescenzo, R. Ostrovsky, and G. Persiano, "Public key encryption with keyword search," in Proc. Adv. Cryptol.-Eurocrypt, 2004, pp. 506–522.
[6] D. Boneh, E. Kushilevitz, R. Ostrovsky, and W. E. Skeith III, "Public key encryption that allows pir queries," in Proc. Adv. Cryptol., 2007, pp. 50–67.
[7] D. X. Song, D. Wagner, and A. Perrig, "Practical techniques for searches on encrypted data," in Proc. IEEE Symp. Secur. Privacy, 2000, pp. 44–55.
[8] E.-J. Goh, "Secure indexes," IACR Cryptol. ePrint Archive, vol. 2003, p. 216, 2003.
[9] Y.-C. Chang and M. Mitzenmacher, "Privacy preserving keyword searches on remote encrypted data," in Proc. 3rd Int. Conf. Appl. Cryptography Netw. Secur., 2005, pp. 442–455.
[10] R. Curtmola, J. Garay, S. Kamara, and R. Ostrovsky, "Searchable symmetric encryption: Improved definitions and efficient constructions," in Proc. 13th ACM Conf. Comput. Commun. Secur., 2006, pp. 79–88.
[11] J. Li, Q. Wang, C. Wang, N. Cao, K. Ren, and W. Lou, "Fuzzy keyword search over encrypted data in cloud computing," in IEEE Proc. INFOCOM, 2010, pp. 1–5.
[12] M. Kuzu, M. S. Islam, and M. Kantarcioglu, "Efficient similarity search over encrypted data," in Proc. IEEE 28th Int. Conf. Data Eng., 2012, pp. 1156–1167.
[13] C. Wang, K. Ren, S. Yu, and K. M. R. Urs, "Achieving usable and privacy-assured similarity search over outsourced cloud data," in Proc. IEEE INFOCOM, 2012, pp. 451–459.