ISSN 2395-1621

Review: Secure Multi-keyword Ranked Search over Encrypted Cloud Data

[#1]Siddheshwar S. Metkari, [#2]Dr. S.B. Sonkamble

[1]smetkari@gmail.com
[2]sonkamblesulochana@gmail.com

[1] Department of Computer Engineering, University of Pune,
JSPM's Rajarshi Shahu School of Engineering & Research, Narhe, Pune, Maharashtra, India.
[2]ME Coordinator,
JSPM's Rajarshi Shahu School of Engineering & Research, Narhe, Pune, Maharashtra, India.

ABSTRACT

In recent years, cloud computing gaining popularity due to its charming features such as scalability, availability, low cost service. Individual and organizations are using cloud service to store large amount of data on cloud storage. By outsourcing data on cloud users gets relief from storage maintenance. But in this case the control of data is going towards cloud service provider. To gain control over the data the traditional technique is encrypt data at client side and outsource the data. In consumer centric cloud computing user wants to find most relevant product or data. Keyword search on encrypted data is difficult. The search techniques which are used on plain text cannot be used over encrypted data. There are some papers based on single keyword search, fuzzy keyword search and multi-keyword search on encrypted data. But these techniques are not supporting synonym based search. The semantic search allows search over encrypted data supporting synonym queries. So this paper survey multi-keyword ranked search over encrypted data and proposed semantic multi-keyword ranked based approach over the encrypted data.

Keywords— Cloud- computing, Searchable encryption, Multi-keyword Search, TF-IDF

ARTICLE INFO

Article History

Received:21st December 2015

Received in revised form :

23th December 2015

Accepted:24st December, 2015

Published online :

28th December 2015

I. INTRODUCTION

In today's scenario, huge amount of data is generated therefore everyone is using cloud for storing their data on cloud storage. The cloud characteristics are on-demand self-service, location independent network access, ubiquitous network access and usage based pay. The applications such as Gmail, drop-box, Whatsapp, Facebook uses cloud computing environment such Amazon S3, Amazon EC2, eucalyptus and openstack [1]. But user lost his physical control on data. The control is in the service provider's hand. Therefore the data is not secured as well as there are many attacks by internal external attackers. Cryptography techniques are used to secure the data. But if you encrypt data the problem is how to search over encrypted data. Search encryption methods are proposed by the researchers which allows user to store encrypted data on cloud and execute keyword search on encrypted data. Our contribution to this paper is summarized in two aspects: multi-keyword ranked search to achieve more accurate search and synonym-based search to support synonym quires.

The remainder of this paper is organized as follows. Related work is discussed in Section II and Section III describes system in detail. Section IV discuss Design Goals and section V summary and conclusion.

II RELATED WORKS

To secure search on encrypted data searchable encryption scheme is used. Searchable encryption allows storing data in encrypted format and you can apply keyword search over encrypted data. There are different searchable encryptions schemes can be constructed using public key based cryptography.

Song et al [2] proposed the first symmetric searchable encryption. These techniques are single keyword Boolean search which gives course search results. The results are not accurate. Then the research has done on multi-keyword search, ranked search. By using multi-keyword ranked search user can query with multiple keywords and retrieve accurate search result. But all these search schemes does not allow synonym based queries.

Searching Techniques

a) **Searchable encryption**

Song el al [2] described new techniques for remote searching on encrypted data using an untrusted server and provided proofs of security for the resulting crypto systems.

Advantages:- these techniques are secure , hidden search, query isolation, simple and fast and the searching time of their scheme is linear to the size of the data collection.

Disadvantages:- In this technique user gets all the files containing all query keywords and user have to process every file and also the network traffic is also increased.

b) **Single Keyword Searchable Encryption**

Ning Caoy et al. [3] presented a Traditional single keyword searchable encryption schemes usually build an encrypted searchable index such that its content is hidden to the server unless it is given appropriate trapdoors generated via secret key(s). Our early work solves secure ranked keyword search which utilizes keyword frequency to rank results instead of returning undifferentiated results. However, it only supports single keyword search. Where anyone with public key can write to the data stored on server but only authorized users with private key can search. Public key solutions are usually very computationally expensive however.

c) **Ranked Keyword Search**

Cong Wang et al [4] discuss the major disadvantages of above mentioned techniques and give the better technique of ranked keyword search. This technique gives user most relevant document in the relevance order with query rather than burdensome sorting through every match in the content collection. This technique avoids unnecessary traffic by sending only most relevant documents in order. But they are only useful single keyword search

d) **Boolean Keyword Searchable Encryption**

Ning Caoy et al. [4] focused a Boolean Keyword Searchable Encryption given without capturing any relevance of the files in the search result. This technique has two drawbacks. First the search user does not have a pre-knowledge of the encrypted cloud data and user have to go through every retried file to find the one he need.

Second the search sends back all files which are only depend on presence or absent of query keywords. This increases unnecessary network traffic and consumes bandwidth.

e) **Fuzzy Keyword search**

B. Wang [5] in this paper enhances the system usability by returning exact matching results. If the exact match fails, it returns the closest match as the result. Edit distance is used to quantify the keyword similarity

Main modules in Fuzzy keyword search are

1. Wildcard-based technique

A wildcard is used to edit the operations at the same position. The edit distance can be calculated using substitution, deletion and insertion.

2. Gram-based technique

Here the fuzzy set is constructed based on grams. The gram of a string is a substring and can be used for effective approximate search. The order of the characters after the primitive operation is always kept the same before the operations.

f) Multi-keyword ranked Search

Cao et al[6] for the first time define and solve the problem of multi-keyword ranked search over encrypted cloud data, and establish a variety of privacy requirements.

In this scheme the documents and queries are represented in the form vectors of dictionary size. The elements in the vector are the normalized TF values. The "co-ordinate matching" is used to rank the documents. But in this scheme the importance of different keyword is not considered.

g) Secure multi-keyword search

Sun et al [7] describes secure multi-keyword search which support similarity search. In this scheme the searchable index tree is constructed by using vector space model.

The cosine similarity with TF-IDF is used to find the relevant score between document and query vectors query vectors and results are returned in rank order.

This algorithm search time is better than linear but the results in precision loss.

h) Secure kNN algorithm

W. K. Wang [8] focuses on query processing over encrypted cloud database. By using the k-nearest neighbor (kNN) query problem over encrypted database outsourced to a cloud: a user issues an encrypted query record to the cloud, and the cloud returns the k closest records to the user. To provide better security, they propose a secure kNN protocol that protects the confidentiality of the data, user's input query, and data access patterns. Also, we empirically analyze the efficiency of their protocols through various experiments.

III PROPOSED MODEL

As shown in the fig.1 the system consists of three different entities Data owner, Data user and Cloud server.

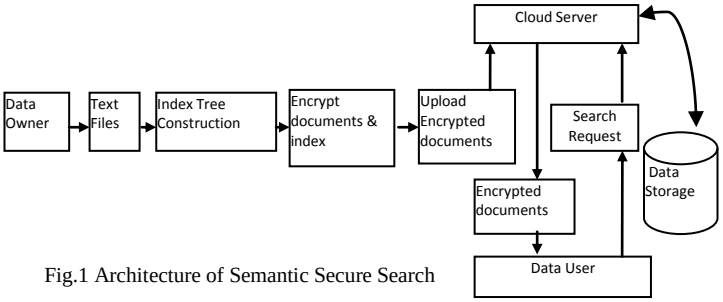

Fig.1 Architecture of Semantic Secure Search

Data Owner has a collection of documents F={f1,f2,…,fn}. He wants to store these documents on cloud storage in encrypted form keeping the capability of keyword search on encrypted data. The data owner first builds a secure searchable tree index from the document collection. Then he encrypts these documents and index tree. And after that he outsources the encrypted documents C and index tree on cloud server. The key information (including keyword IDF values) are securely distributed to the authorized users.

Data Users are authorized users to access the documents of owner. The user first generates a trapdoor TD with query keywords and secure key. User retries top k most relevant k encrypted documents from the cloud server. The retrieved documents are decrypted with the shared secrete key.

Cloud Server stores data owner's documents and searchable index tree. Upon receiving request form user, the cloud server searches over the index tree and finally returns the top k most relevant documents.

IV DESIGN GOALS

To enable semantic multi-keyword ranked search over encrypted cloud data under above model our system has following goals.

Multi-keyword Search This search is like Google search over encrypted cloud data. The search results are more accurate than the single keyword search.

Ranked Results To quickly identify most relevant results, the cloud server sort the results in relevant order.

Search Efficiency The special search tree index and efficient search algorithms achieve better search efficiency.

Privacy Preserving The underlying plain text documents and index tree is encrypted before outsourcing. The cloud server is unable to learn any information about the data.

V CONCLUSIONS

In this scheme, a multi-keyword ranked search scheme over encrypted cloud data is proposed, which meanwhile supports semantic search. We use the vectors consisting of TF values as indexes to documents. Taking security and privacy into consideration, we employ a secure splitting k-NN technique to encrypt the index and the queried vector, so that we can obtain the accurate ranked results and protect the confidence of the data well.

REFRENCES

[1] Zhihua Xia, Xinhui Wang, Xingming Sun, and Qian Wang, "A Secure and Dynamic Multi-keyword Ranked Search Scheme over Encrypted Cloud Data", IEEE Transactions on Parallel and Distributed Systems 2015.

[2] 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 on. IEEE, 2000, pp. 44– 55.

[3] Ning Caoy, Cong Wangz, Ming Liy, Kui Renz, and Wenjing Louy" Privacy-Preserving Multi-keyword Ranked Search over Encrypted Cloud Data" INFOCOM, 2011 Proceedings IEEE.

[4] Cong Wang†, Ning Cao‡, Jin Li†, Kui Ren†, and Wenjing Lou"Secure Ranked Keyword Search over Encrypted Cloud Data" Distributed Computing System,2010 IEE 30th international conference.

[5] B. Wang, S. Yu, W. Lou, and Y. T. Hou, "Privacy-preserving multi-keyword fuzzy search over encrypted data in the cloud," in IEEE
a. INFOCOM, 2014.

[6] N. Cao, C. Wang, M. Li, K. Ren, and W. Lou, "Privacy-preserving multi-keyword ranked search over encrypted cloud data," in IEEE INFOCOM, April 2011, pp. 829–837.

[7] W. Sun, B. Wang, N. Cao, M. Li, W. Lou, Y. T. Hou, and H. Li, "Privacy-preserving multi-keyword text search in the cloud supporting similarity-based ranking," in Proceedings of the 8th ACM SIGSAC symposium on Information, computer and communications security. ACM, 2013, pp. 71–82.

[8] W. K. Wong, D. W.-l. Cheung, B. Kao, and N. Mamoulis, "Secure knn computation on encrypted databases", in Proceedings of the 2009 ACM SIGMOD International Conference on Management of data. ACM, 2009, pp. 139152

[9] P. Golle, J. Staddon, and B. Waters, "Secure conjunctive keyword search over encrypted data," in Applied Cryptography and NetworkSecurity. Springer, 2004, pp. 31–45.

[10] Y. H. Hwang and P. J. Lee, "Public key encryption with conjunctive keyword search and its extension to a multi-user system," in Proceedings of the First international conference on Pairing-Based Cryptography. Springer-Verlag, 2007, pp. 2–22.

[11] L. Ballard, S. Kamara, and F. Monrose, "Achieving efficient conjunctive keyword searches over encrypted data," in Proceedings of the 7th international conference on Information and Communications Security. Springer-Verlag, 2005, pp. 414–426.

[12] C. Orencik, M. Kantarcioglu, and E. Savas, "A practical and secure multi-keyword search method over encrypted cloud data," in Cloud Computing (CLOUD), 2013 IEEE Sixth International Conference on. IEEE, 2013, pp. 390–397.

[13] W. Zhang, S. Xiao, Y. Lin, T. Zhou, and S. Zhou, "Secure ranked multi-keyword search for multiple data owners in cloud computing," in Dependable Systems and Networks (DSN), 2014 44th Annual IEEE/IFIP International Conference on. IEEE, 2014, pp. 276–286.

[14] S. Kamara, C. Papamanthou, and T. Roeder, "Dynamic searchable symmetric encryption," in Proceedings of the 2012 ACM conferenceon Computer and communications security. ACM, 2012, pp. 965–976.

[15] S. Kamara and C. Papamanthou, "Parallel and dynamic searchable symmetric encryption," in Financial Cryptography and DataSecurity. Springer, 2013, pp. 258–274.

[16] J. Katz, A. Sahai, and B. Waters, "Predicate encryption supporting disjunctions, polynomial equations, and inner products," in Advances in Cryptology–EUROCRYPT 2008. Springer, 2008, pp. 146–162.

[17] E. Shen, E. Shi, and B. Waters, "Predicate privacy in encryption systems," in Proceedings of the 6th Theory of Cryptography Conference on Theory of Cryptography. Springer-Verlag, 2009, pp. 457–473.

[18] A. Lewko, T. Okamoto, A. Sahai, K. Takashima, and B. Waters, "Fully secure functional encryption: attribute-based encryption and (hierarchical) inner product encryption," in Proceedings of the 29th Annual ternational conference on Theory and Applications of Cryptographic Techniques. Springer-Verlag, 2010, pp. 62–91.

