ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 9, September 2016

Security Analysis Using Probabilistic OPE Based Cloud Searching

Ashlesha Khatke[1], Neha Sharma[1], Sachin Kade[1], Sayyad Sofia[1], Sonali Patil[2]

B.E. Students, Department of Computer Engineering, JSPM's BSIOTR, Wagholi, Pune, India [1]

Assistant Professor, Department of Computer Engineering, JSPM's BSIOTR, Wagholi, Pune, India [2]

ABSTRACT: Cloud computing commercial enables pattern of data service outsourcing .To protect data privacy, sensitive cloud data has to be encrypted before outsourcing to the cloud, which is very challenging task. As the consequent to this, the search over encrypted data becomes a complex task. However, encryption on sensitive data is big task for transformation of data. Information healing become difficult in the encrypted format due to large outsourced files and traditional search pattern cannot display the cipher text directly. As, Ranked Search in encrypted cloud data ,order preserving encryption(OPE) is an productive way to encrypt relevance score of the inverted index but, in deterministic OPE the cipher texts report same dissemination as of relevance score. To overcome this, we proposed the probabilistic OPE for the function of searchable encryption which can raze the distribution of the plaintexts. In this project we also suggested the differential attack on probabilistic OPE by using variance of the ordered cipher texts .The results depicts that the cloud server satisfying evaluation of the distribution of the relevance score by a differential attack.

KEYWORDS: Cloud Computing, Cipher text, Probabilistic OPE, Searchable encryption.

I. INTRODUCTION

Cloud computing is a very efficient representative to organize enormous resources of computing ,and available users to enjoy pervasive ,available and on-demand network access to a shared pool of computing resources with great productivity and less economic overhead [1][2] . Cloud Computing becomes frequent, more and more sensitive information are being unified into the cloud, such as confidential emails, personal health records, company data, and government records, etc. The data owners and cloud server are not always available in the same trusted domain which would outsource unencrypted data at risk [4]; the cloud server may fissure data information to unauthorized access [3] or even be hacked. It means that sensitive data need to be encrypted before to outsourcing for data privacy and combating undesired accesses. But which lead to huge cost in term of data usability. Downloading all the data from the cloud and decrypt is not possible, because the information healing become challenging through encrypted domain as size of the outsourced file can be large and so conventional search cannot arrange the chipher text revival directly. To overcome this, Searchable Encryption (SE) [5] was proposed to make query in the encrypted domain through the authorized user. As there are many drawbacks of searchable encryption.

To overcome the drawback of Searchable Encryption, we are applying order preserving encryption which increases the efficiency of ranked search. This algorithm was proposed in 2004 to solve encrypted query problem in the database system. The order-preserving property means that if the plaintexts x1 < x2, then the corresponding ciphertexts E(x1) and E(x2) satisfy E(x1) < E(x2). Boldyreva et al defined the security of OPE and proposed a deterministic secure OPE theory. However, the security definition and the constructions of OPE in [3], are based on the assumption that OPE is a deterministic encryption scheme which means that a given plaintext will always be encrypted as a fixed ciphertext. Therefore, deterministic encryption displays the distribution of the plaintexts, so it cannot secure data privacy in most applications. For example, in privacy-preserving keywords search, OPE is used to encrypt relevance scores in the inverted index [4]. As noted by Wang et al. [3], when using a deterministic OPE, the resulting ciphertext gives precisely the same distribution as the relevance score, by which the server can specify the keywords [6] . Therefore, Wang et al. [3] improved the OPE in [7] and proposed a “One-to-Many OPE” in their secure keyword search scheme, where they tried to formulate a probabilistic encryption theory and cover the distribution of the plaintexts.

II. RELATED WORK

In [1] author used secure multi-keyword ranked search on the encrypted data which concurrently support dynamic update procedure .The vector space model and the mostly used TF x IDF model are united in index and query generation. The “Greedy Depth-first Search” algorithm is used to provide multi-keyword ranked search. The use of special tree-based index structure, the described theory can achieve sub-linear search time and deal with dynamic operation efficiently. In [2] author meets challenging conditions where the outsourced dataset can be shared by many authorized users and searched by many users. Attribute-based Encryption (ABE) here first attribute-based keyword search scheme with user revocation that enable file-level search authorization. This theory allows multiple owners to encrypt and outsource the data to cloud server individually. In [3] author defined and solves the problem of Secure ranked search on encrypted cloud data. The author proposed Order Preserving Encryption (OPE) technique to support search process ranked manner [5]. The relevance score and inverted index are secured with the Order Preserving Encryption (OPE). The distribution and index differences are utilized to estimate the search keyword in differential attacks. Someone introduces new schemes for confidentiality preserving rank-ordered search and regeneration over large document collections [6]. The described scheme not only assures document/query confidentiality against an unauthorized user, but also prevents an untrusted data center from learning information about the query and the document collection. In [7] we propose a security notion in the spirit of pseudorandom functions (PRFs) and related primitives asking that an OPE scheme look “as-random-as-possible” subject to the order-preserving constraint. We then design an efficient OPE scheme and prove its security under our notion based on pseudo randomness of an underlying block cipher. Here we aim at constructing efficient and programmable OPE framework for outsourced database and implementation detail including how to use our OPE framework in database application [8].

Author reports challenging open issues by stating and enforcing access policies based on data attributes and gives the data owner alternate task involved in fine-grained data access control to untrusted cloud server without exposing the data content [9]. This is a survey of the different security risks that pose threat to the cloud is presented [10, 11]. In this system, we explain and solve the interesting problem of privacy preserving multi keywords ranked search over encrypted cloud data, and create a set of strict privacy necessities for such a safe cloud data application system to be effected in real. We analyze the security of the OP E encryption scheme SE, m, n and give the upper bound on the probability for the adversary to recover the plain text encrypted by SE, m, n under chosen plain text attacks [12].

III. PROPOSED SYSTEM

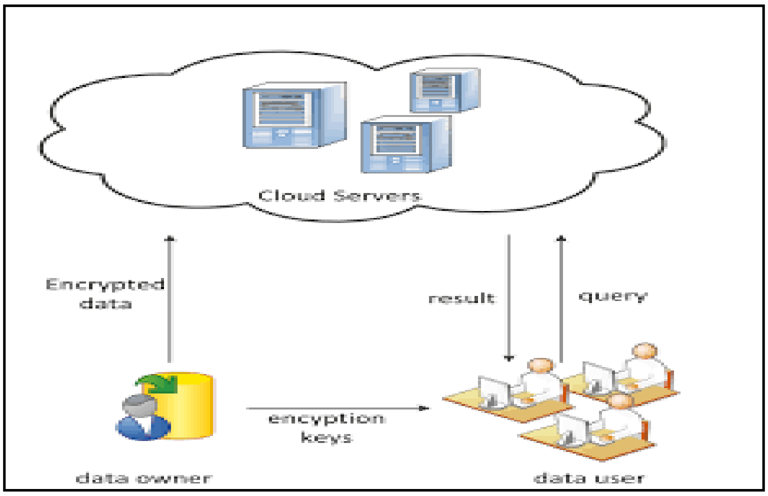


Fig. 1. Framework of retrieval over encrypted cloud data

To gain useful data retrieval of enormous documents, it is mandatory to perform relevance ranking on the results. Ranked search can also quietly reduce network traffic by sending back only the most required data. In ranked search, the ranking function plays an vital role in calculating the relevance between files and the given searching query. The most used relevance score is defined based on the model of TF x IDF, where term frequency (TF) is the number of times a term (keyword) arrives in a file and inverse document frequency (IDF) is the ratio of the total number of files to the number of files containing the term. There are many types of TF x IDF-based ranking functions, and in [3], the following one is used.

$$\text{Score}\ (w,F_d) = \frac{1}{|F_d|}.\left(1+\ln f_{d,w}\right).\ln\left(1+\frac{N_d}{f_w}\right) \qquad (1)$$

Where, w denotes the keyword and $f_{d,w}$ denotes the TF of term w in file F_d; N/fw denotes IDF where f_w is the number of files that contain term w and N_d is the total number of documents in the collection; and |Fd| is the number of indexed terms containing in file F_d, i.e., the length of F_d. For fast search, the keywords, IDs of files, and the relevance scores are usually organized as an index structure named "Inverted Index". An example on posting list of the Inverted Index is shown in TABLE 1. With a complete Inverted Index, the server can complete retrieval task by s comparing the relevance scores saved in the index which predicts the significant level of each file for a certain keyword.

Table 1. Example of Posting List of Inverted Index

Keyword		**W**		
File ID	F_1	F_2	…	$F_{f\,w}$
Relevance Score	8.6	6.1	……	7.3

A data owner can be an individual or a corporation, i.e., it is the entity that owns a collection of documents Dc = {D1, D2……… D_{Nd} } that it wants to share with authorized users. The keyword set is stated asW = {w_1,w_2…..w_{Nw}}. For security and privacy concerns, documents need to be encrypted into E = {E(D1);E(D2) ……E(D_{Nd})} before being uploaded to the cloud server. The plaintext index has to be encrypted into I to prevent information issuing. The encrypted form of the example of the posting list of the Inverted Index is shown in TABLE 2, in which the keyword w_i is secured by a Hash function hash (), and the relevance scores are encrypted by an encryption scheme E′().

Table 2. Example of Encrypted Posting List of The Inverted Index

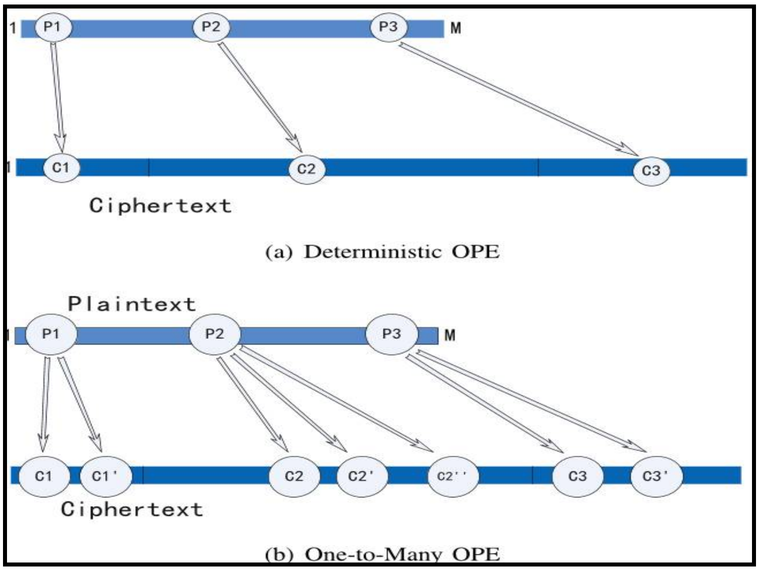
Keyword		**Hash(w)**		
File ID	F_1	F_2	…	$F_{f\,w}$
Relevance Score	E`(8.6)	E`(6.1)	……	E`(7.3)

We use Table 2. as an example to see how a cloud server make secure search based on an encrypted index. In the search procedure, a user first generates a search request in a secret form a trapdoor T (w). In this example, the trapdoor is just the hash values of the keyword of interest. Once the cloud server receives the trapdoor T (w), it checks it with the hash values of all keywords in the index I, then the required documents which are corresponding to keyword w are found. The server returns the matched file IDs: F_1, F_2, ... , F_{fw} to the user. Finally, the user can download all the encrypted documents based on the given IDs and decrypt them. A desirable system is supposed to return the documents in a ranked order by their relevance with queried keyword, but using traditional encryption theory will disorder relevance score.

IV. PROPOSED ALGORITHM

A. BINARY SEARCH:

The range R is divided into some non-overlapping interval buckets with contingent sizes. The contingent-sized bucket is determined by a binary search based on a random HGD sampler. In [3], procedure of binary search is described as Algorithm 1, where TapeGen() is a random coin generator After the binary search, a plaintext m is mapped into a bucket in the range R, and then the OPE algorithm assigns a fixed value in the bucket as the encrypted value of m. The encryption process of Algorithm 1 is illustrated in Fig. 2(a), which shows that a given plaintext mi will mapped to a fixed ciphertext ci associated to a bucket choose by the binary search method therefore it is a deterministic encryption.



(a) Deterministic OPE

(b) One-to-Many OPE

Fig.2. Comparison between Deterministic OPE and Probabilistic OPE

B. ONE-TO-MANY ORDER PRESERVING ENCRYPTION:

For a given plaintext m, i.e., a relevance score, the “One-to-Many OPE” first uses Algorithm1 to select a bucket for m, and then randomly chooses a value in the bucket as the ciphertext. The randomly choosing method in the bucket is seeded by the unique file IDs together with the plaintext m, and thus the same relevance score in the Inverted Index will be encrypted as different ciphertexts. The encryption process of “One-to- Many OPE” is described in Algorithm2 [3], which is also illustrated in Fig. 2(b).

V. CONCLUSION

User data security and privacy are promoted by the encrypted cloud storage services. One to many Order Preserving Encryption (OPE) is used to perform searching on encrypted data records. Differential attack handling mechanism is associated with the probabilistic OPE scheme. Semantic query based indexing and document retrieval scheme is used to increase the efficiency of search levels. The system provides query privacy in search process through encrypted cloud data services. Search duration is minimized in the semantic relationship based encrypted keyword search process. Accuracy is improved with relevance score and semantic query model. The system controls the keyword inferring attacks with change point modification and noise keyword insertion mechanism.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 9, September 2016

REFERENCES

1. A Secure and Dynamic Multi-keyword Ranked Search Scheme over Encrypted Cloud Data :Zhihua Xia, Member, IEEE, Xinhui Wang, Xingming Sun, Senior Member, IEEE, and Qian Wang, Member, IEEE, 2015
2. Protecting Your Right: Verifiable Attribute-based Keyword Search with Fine-grained Owner-enforced Search Authorization in the Cloud: Wenhai Sun, Student Member, IEEE, Shucheng Yu, Member, IEEE, Wenjing Lou, Senior Member, IEEE, Y. Thomas Hou, Fellow, IEEE, Hui Li, Member, IEEE
3. Enabling Secure and Efficient Ranked Keyword Search over Outsourced Cloud Data: Parallel and Distributed Systems, IEEE Transactions 23(8), pp. 1467-1479, 2012.
4. A. Boldyreva, N. Chenette and A. ONeill, “Order-preserving encryption revisited: improved security analysis and alternative solutions,” Advances in CryptologyCCRYPTO, 2011. Springer Berlin Heidelberg, pp. 578-595, 2011
5. Protecting Data and Query against Differential Attacks in Outsourced Cloud Search: International Journal On Engineering Technology and Sciences – IJETS™ ISSN(P): 2349-3968, ISSN (O): 2349-3976 Volume III, Issue VII, July- 2016
6. Swaminathan, Y. Mao and G.-M Su, “Confidentiality-preserving rank-ordered search,” Proceedings of the 2007 ACM workshop on Storage security and survivability. ACM, pp. 7-12, 2007.
7. A. Boldyreva, N. Chenette and Y. Lee , “Order-preserving symmetric encryption,” Advances in Cryptology-EUROCRYPT, 2009. Springer Berlin Heidelberg, pp. 224-241, 2009.
8. New order preserving encryption model for outsourced databases in cloud environment :Zheli Liu, Xiaofeng Chen, Jun Yang, Chunfu Jia, Ilsun You, Journal of Network and Computer Application Volume 59,January 2016
9. S. Yu, C. Wang and K. Ren, “Achieving secure, scalable, and fine grained data access control in cloud computing,” *INFOCOM, 2010 Proceedings IEEE*. IEEE, pp. 1-9, 2010.
10. A survey on security issues in service delivery models of cloud computing : S.Subashini ,V.Kavitha, Journal of Network and Computer Application Volume 34 ,January 2011.
11. N. Cao, C. Wang and M. Li, “Privacy-preserving multi-keyword ranked search over encrypted cloud data,” *INFOCOM, 2011 Proceedings IEEE*. IEEE, pp. 829-837, 2011.
12. L. Xiao, I.-L Yen, “Security analysis for order preserving encryption schemes,” *Proc. of 46th Annual Conference on Information Sciences and System*, pp. 1-6, 2012.

.