© 2017 IJSRST | Volume 3 | Issue 6 | Print ISSN: 2395-6011 | Online ISSN: 2395-602X
Themed Section: Science and Technology

Data Identification and Control Mechanism Using Distinguishing, Attacks In Cloud

S. Meena[1], Dr. N. Kowsalya[2]
[1]M.Phil Full time Research Scholar, PG& Research Department of Computer Science
[2]Assistant Professor, Department of Computer Applications,Vivekananda College of Arts & Sciences for Women (Autonomous), Namakkal, Tamil Nadu, India

ABSTRACT

Cloud environment provides encrypted data management facility to the shared data. Security and privacy are guaranteed with encrypted storage support model. The data management and security requirements are handled by the cloud server only. Encrypted cloud storage methods are adapted to secure the data shared under the clouds. All the outsourced operations are carried out on the encrypted data only. The data and query comparison operations are performed using the encrypted data search mechanism. The Order Preserving Encryption (OPE) technique is adapted to support search process ranked manner. The relevance score and inverted index are protected with the Order Preserving Encryption (OPE). Security and privacy are guaranteed with encrypted storage support model. The Order Preserving Encryption (OPE) technique is adapted to support search process ranked manner. The relevance score and inverted index are protected with the Order Preserving Encryption (OPE). The distribution of encrypted data values are unchanged in the deterministic OPE mechanism. The index distribution is managed to support search operation in One-to-many OPE. One to many OPE is also denoted as probabilistic OPE Scheme. The outsourced data search on encrypted data model is carried out with the binary search algorithm. The distribution and index differences are utilized to estimate the search keyword in differential attacks.

Keywords: Outsourced Data Search, Data Centers, Order Preserving Encryption and Differential Attacks

I. INTRODUCTION

The key driving forces behind cloud computing is the ubiquity of broadband and wireless networking, falling storage costs and progressive improvements in Internet computing software. Cloud-service clients will be able to add more capacity at peak demand, reduce costs, experiment with new services and remove unneeded capacity, whereas service providers will increase utilization via multiplexing and allow for larger investments in software and hardware. Currently, the main technical underpinnings of cloud computing infrastructures and services include virtualization, service-oriented software, grid computing technologies, management of large facilities and power efficiency. Consumers purchase such services in the form of infrastructure-as-a-service (IaaS), platform-as-a-service (PaaS), or software-as-a-service (SaaS) and sell value-added services to users. Within the cloud, the laws of probability give service providers great leverage through statistical multiplexing of varying workloads and easier management a single software installation can cover many users' needs.

Two different architectural models are considered for clouds. The first one is designed to scale out by providing additional computing instances on demand. Clouds can use these instances to supply services in the form of SaaS and PaaS. The second architectural model is designed to provide data and compute-intensive applications via scaling capacity. In most cases, clouds provide on-demand computing instances or capacities with a "pay-as-you-go" economic model. The cloud infrastructure can support any computing model compatible with loosely coupled CPU clusters. Organizations can provide hardware for clouds internally, or a third party can provide it externally. A cloud might be restricted to a single organization or group, available to the general public over the Internet, or shared by multiple groups or organizations.

II. Related Work

Searchable encryption is a promising technique that provides the search service over the encrypted cloud data. It can mainly be classified into two types: Searchable Public-key Encryption (SPE) and Searchable Symmetric Encryption (SSE). Boneh et al. [1] first propose the concept of SPE, which supports single keyword search over the encrypted cloud data. The work is later extended to support the conjunctive, subset and range search queries on encrypted data. Zhang et al. [2] propose an efficient public key searchable encryption scheme with conjunctive-subset search. The above proposals require that the search results match all the keywords at the same time and cannot return results in a specific order.

Cao et al. propose a privacy-preserving multi-keyword search scheme that supports ranked results by adopting secure *k*-nearest neighbors (kNN) technique in searchable encryption. The proposal can achieve rich functionalities such as multi-keyword and ranked results, but requires the computation of relevance scores for all documents contained in the database. This operation incurs huge computation overload to the cloud server and is therefore not suitable for large-scale datasets. Cash et al. [3] adopt the inverted index *TSet*, which maps the keyword to the documents containing it, to achieve efficient multi-keyword search for large-scale datasets.

III. Document search in clouds

In practice, to realize effective data retrieval on large amount of documents, it is necessary to perform relevance ranking on the results. Ranked search can also significantly reduce network traffic by sending back only the most relevant data. In ranked search, the ranking function plays an important role in calculating the relevance between files and the given searching query. The most popular relevance score is defined based on the model of T F×I DF, where term frequency (TF) is the number of times a term (keyword) appears 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 variations of T F × I DF-based ranking functions, the following one is adopted.

$$score(w, F_d) = \frac{1}{|F_d|}.(1 + \ln f_d, w).ln\left(1 + \frac{N_d}{f_w}\right) (1)$$

Herein, w denotes the keyword and fd,w denotes the TF of term w in file F_d ; N/ f_w 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 $|F_d|$ is the number of indexed terms containing in file F_d , i.e., the length of F_d .To realize 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. With a complete Inverted Index, the server can complete retrieval task by simply comparing the relevance scores stored in the index which represent the importance level of each file for a certain keyword.

Due to the special background of cloud computing, unlike traditional plaintext information retrieval, there are usually three entities in cloud data retrieval: data owner, remote cloud server and users. A data owner can be an individual or a corporation, i.e., it is the entity that owns a collection of documents $D_c = \{D_1, D_2 \ldots D_{Nd}\}$ that it wants to share with trusted users. The keyword set is marked as $W = \{w_1, w_2 \ldots W_{Nw}\}$. For security and privacy concerns, documents have to be encrypted into $\xi = \{E(D_1), E(D_2) \ldots E(D_{Nd})\}$ before being uploaded to the cloud server. Additionally, the plaintext index has to be encrypted into I to prevent information leakage.

The encrypted form of the example of the posting list of the Inverted Index in which the keyword w_i is protected by a Hash function hash() and the relevance scores are encrypted by a encryption scheme E'(). An example to see how a cloud server conducts a 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).

A desirable system is supposed to return the documents in a ranked order by their relevance with the queried keyword, but using traditional encryption schemes will disorder relevance scores. Order Preserving Encryption (OPE) is applied to encrypt the relevance scores, which enables the server to quickly perform ranked search without knowing the plain relevance scores.

Encryption on sensitive data presents obstacles to the processing of the data. Information retrieval becomes difficult in the encrypted domain because the amount of outsourced files can be very large and traditional search patterns can not be deployed to ciphertext retrieval directly. Users need to download all the data, decrypt it all and then search keywords like plaintext retrieval. To overcome this, Searchable Encryption (SE) was proposed to make query in the encrypted domain possible while still preserving users' privacy [4] . There are several problems in searchable encryption: fuzzy search, ranked search, multi keyword search and so on. Song *et al.* first proposed a search scheme only supporting single Boolean keyword search. After that plenty of searchable encryption methods arose to improve efficiency and reduce communication overhead. Applying order preserving encryption (OPE) is one practical way of supporting fast ranked search. This algorithm was first proposed in 2004 to solve encrypted query problems in database systems. OPE is a symmetric cryptosystem, therefore it is also called order-preserving symmetric encryption (OPSE). The order-preserving property means that if the plaintexts $x_1 < x_2$, then the corresponding ciphertexts $E(x_1)$ and $E(x_2)$ satisfy $E(x_1) < E(x_2)$. Boldyreva *et al.* initiated the cryptographic study of OPE schemes defined the security of OPE and proposed a provably secure OPE scheme. The security definition and the constructions of OPE and 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.

For instance, in privacy preserving keywords search, OPE is used to encrypt relevance scores in the inverted index. As noted by Wang *et al.*, using a deterministic OPE, the resulting ciphertext shares exactly the same distribution as the relevance score, by which the server can specify the keywords. Wang *et al.* improved the OPE and proposed a "One-to-Many OPE" in their secure keyword search scheme, where they tried to construct a probabilistic encryption scheme and conceal the distribution of the plaintexts. We discover that the One-to-Many OPE cannot ensure the expected security. In fact, although the ciphertexts of One-to-Many OPE conceals the distribution of the plaintexts, an adversary may estimate the distribution from the differences of the ciphertexts [9]. So in this system propose a differential attack on the One-to-Many OPE.

IV. Problem Statement

Encrypted cloud storage is used to share user data with security and privacy. Raked search in encrypted cloud data process is carried out using Order Preserving Encryption (OPE) technique. Order Preserving Encryption (OPE) is applied to encrypt relevance scores of the inverted index. In deterministic OPE the ciphertexts reveals the distribution of relevance scores. One-to-many OPE is employed to flatten the distribution of the plaintexts in applications of searchable encryption. One to many OPE is also referred as probabilistic OPE Scheme. Binary search algorithm is applied to perform document search on encrypted data environment. Differential attack on one-to-many OPE is initiated by exploiting the differences of the ordered ciphertexts. The following problems are identified from the current cloud data search methods.

- Keyword inferring process is not controlled
- Change point analysis based relevance score distribution estimation is not handled
- Background knowledge based attacks are not controlled
- Semantic query model is not provided

V. Attack detection and control in differential methods

The Probabilistic OPE based scheme is enhanced with security measures to handle differential attacks. Term subset reassignment mechanism is integrated with the One to many OPE scheme to control change point based activities. Inverted index is protected with noise document entries to secure relevance score values. Document search and indexing operations are improved with semantic analysis methods. Attack analysis and protection operations are integrated with the Probabilistic OPE technique. Conceptual relationship based search scheme is adapted in the encrypted data search process. Query process is carried out with privacy preserved manner. The system is divided into six major modules. They are Cloud Server, Relevance Score Assignment, Probabilistic OPE, Attack Analysis, Index Distribution Security and Query Process. Cloud server

manages the encrypted data values. Relevance score assignment module is designed to update weight values. Data encryption is carried out under the Probabilistic OPE module. Attack analysis module is used to discover the differential attacks. Index values are protected using index distribution security process. Query process is called to perform encrypted data search process.

Encrypted data values are maintained under the cloud server application. Data encryption and upload operations are initiated by the data owner. Data owner and user details are managed under the cloud server. Data owner provides the key value for the users. The relevance score is assigned for the plain text values. The system integrates the relevance score with weight values. Term weights are estimated using statistical analysis. Concept relationship analysis mechanism is applied to estimate the semantic weights. Probabilistic Order Preserving Encryption (POPE) is employed to encrypt the relevance scores with index values. Inverter index is used to arrange the relevance score values. Weight values are also integrated with the index process. Random values are used to reassign the distribution intervals. Attack analysis is initiated to verify the index distribution levels. Differential attacks are discovered with distribution relationship values. Index subsets are analyzed in the attack analysis process. Differential attacks are discovered with query keyword intervals.

Index distribution security process is used to control differential attacks. Noise document entries are inserted to protect the relevance score and index values. Change point activities are controlled with term subset reassignment technique. The index distribution security is also applied to protect the semantic index values. The query process is initiated to search on encrypted data values. User privacy is ensured with query keyword encryption process. Query results are ranked with relevance score and weight values.

VI. CONCLUSION

User data security and privacy are supported by the encrypted cloud storage services. One to many Order Preserving Encryption (OPE) is applied to perform document search on encrypted data collection. Differential attack handling mechanism is integrated with the probabilistic OPE scheme. Semantic query based indexing and document retrieval scheme is adapted to improve the search levels. The system provides query privacy in search process under encrypted cloud data services. Search duration is reduced 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.

VII. REFERENCES

[1]. Q. Liu, C. C. Tan, J. Wu and G. Wang, "Efficient information retrieval for ranked queries in cost-effective cloud environments," in Proc. IEEE INFOCOM, Mar. 2012, pp. 2581-2585.

[2]. W. Sun, S. Yu, W. Lou, Y. T. Hou and H. Li, "Protecting your right: Attribute-based keyword search with Fine-grained owner-enforced search authorization in the cloud," in Proc. IEEE INFOCOM, Apr./May 2014,pp. 226-234.

[3]. Q. Zheng, S. Xu and G. Ateniese, "VABKS: Verifiable attribute based keyword search over outsourced encrypted data," in Proc. IEEE INFOCOM, Apr. 2014, pp. 522-530.

[4]. Hongwei Li, Dongxiao Liu, Tom H. Luan and Xuemin (Sherman) Shen, "Enabling Efficient Multi-Keyword Ranked Search Over Encrypted Mobile Cloud Data Through Blind Storage", IEEE Transactions On Emerging Topics In Computing, 6 March, 2015.

[5]. J. Yu, P. Lu, Y. Zhu, G. Xue and M. Li, "Toward secure multikeyword top-k retrieval over encrypted cloud data," IEEE Trans. Dependable Secure Comput., vol. 10, no. 4, pp. 239-250, Jul./Aug. 2013.

[6]. D. Cash et al., "Dynamic searchable encryption in very-large databases: Data structures and implementation," in Proc. NDSS, Feb. 2014.

[7]. B.Wang, S.Yu,W. Lou and Y. T. Hou, "Privacy-preserving multi-keyword fuzzy search over encrypted data in the cloud," in Proc. IEEE INFOCOM, Apr./May 2014, pp. 2112-2120.

[8]. C. Wang, N. Cao, J. Li, K. Ren and W. Lou, "Secure ranked keyword search over encrypted cloud data," in Proc. IEEE 30th Int. Conf. Distrib. Comput. Syst. (ICDCS), Jun. 2010, pp. 253-262.

[9]. Amelie Chi Zhou, Bingsheng He and Cheng Liu, "Monetary Cost Optimizations for Hosting Workflow-as-a-Service in IaaS Clouds", IEEE Transactions on Cloud Computing, Jan-Mar 2016.
[10]. D. Cash, S. Jarecki, C. Jutla, H. Krawczyk, M.-C. Rou and M. Steiner, "Highly-scalable searchable symmetric encryption with support for Boolean queries," in Proc. CRYPTO, 2013, pp. 353-373.
[11]. B. Zhang and F. Zhang, "An efficient public key encryption with conjunctive subset keywords search," J. Netw. Comput. Appl., vol. 34, no. 1, pp. 262-267, Jan. 2011.
[12]. C. Wang, N. Cao, K. Ren and W. Lou, "Enabling secure and efficient ranked keyword search over outsourced cloud data," IEEE Trans. Parallel Distrib. Syst., vol. 23, no. 8, pp. 1467-1479, Aug. 2012.
[13]. M. Naveed, M. Prabhakaran and C. A. Gunter, "Dynamic searchable encryption via blind storage," in Proc. IEEE Symp. Secur. Privacy, May 2014, pp. 639-654.