ISSN 2395-695X (Print)
ISSN 2395-695X (Online)
International Journal of Advanced Research in Biology Engineering Science and Technology (IJARBEST)
Vol. 2, Special Issue 10, March 2016

Security Enhancement of Prospect Order Preserving Encryption (POPE) with Attack Resistance

M.Haritha[1], C.Thirumalai selvan[2], Dr.V.Venkatachalam3

1, 2, 3 Department of Computer Science & Engineering
1,2K.S.R College of Engineering, Tamilnadu, India
3 The Kavery Engineering College, Tamilnadu, India

ABSTRACT

Highly scalable services are provided to the users through the Internet. Cloud services are provided on user request basis. In cloud environment users' data are usually processed remotely in unknown machines that users do not own or operate. User data control is reduced on data sharing under remote machines. Sensitive data values are protected and shared in encrypted form. Data encryption is carried out before outsourced to a commercial public cloud. Traditional search patterns cannot be deployed to ciphertext retrieval directly. User privacy is ensured with encrypted data search methods in cloud data centers

Encrypted cloud storage is used to share user data with security and privacy. Ranked 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.Attack on one-to-many OPE is initiated by exploiting the differences of the ordered ciphertexts.

The Prospect OPE based scheme is enhanced with security measures to handle 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.

1. INTRODUCTION

ISSN 2395-695X (Print)
ISSN 2395-695X (Online)
International Journal of Advanced Research in Biology Engineering Science and Technology (IJARBEST)
Vol. 2, Special Issue 10, March 2016

One vision of 21st century computing is that users will access Internet services over lightweight portable devices rather than through some descendant of the traditional desktop PC. Because users won't have powerful machines, who will supply the computing power? The answer to this question lies with cloud computing. Cloud computing is a recent trend in IT that moves computing and data away from desktop and portable PCs into large data centers. It refers to applications delivered as services over the Internet as well as to the actual cloud infrastructure — namely, the hardware and systems software in data centers that provides these services.

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.

A cloud comprises processing, network and storage elements and cloud architecture consists of three abstract layers. Infrastructure is the lowest layer and is a means of delivering basic storage and compute capabilities as standardized services over the network. Servers, storage systems, switches, routers and other systems handle specific types of workloads, from batch processing to server or storage augmentation during peak loads. The middle platform layer provides higher abstractions and services to develop, test, deploy, host and maintain applications in the same integrated development environment. The application layer is the highest layer and features a complete application offered as a service.

In 1961, John McCarthy envisioned that "computation may someday be organized as a public utility." The cloud computing paradigm can be viewed as a big step toward this dream. To realize it fully, several significant problems and unexploited opportunities concerning the deployment, efficient operation and use of cloud computing infrastructures. Data is replicated across large geographic distances, where its availability and durability are paramount for cloud service providers. It's also stored at untrusted hosts, which creates enormous risks for data privacy [15]. Computing power in clouds must be elastic to face changing conditions. For instance, providers can allocate additional computational resources on the fly to handle increased demand. They should deploy novel data management approaches, such as analytical data management tasks, multitenant databases for SaaS, or hybrid designs among database management systems (DBMSs) and MapReduce-like systems so as to address data limitations and harness cloud computing platforms' capabilities.

2. 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. [9] 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. [11] 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. Further, Liu et al. [1] propose a ranked search scheme which adopts a mask matrix to achieve cost effectiveness. Yu et al. [5] propose a multi keyword retrieval scheme that can return the top-k relevant documents by leveraging the fully homomorphic encryption. [2], [3] adopt the attribute-based encryption technique to achieve search authority in SPE.

Although SPE can achieve above rich search functionalities, SPE are not efficient since SPE involves a good many asymmetric cryptography operations. This motivates the research on SSE mechanisms. The first SSE scheme is introduced by Song et al. [4], which builds the searchable encrypted index in a symmetric way but only supports single keyword. Curtmola et al. further improve the security definitions of SSE. Their work forms the basis of many subsequent works, such as [10], [13] and [6], by introducing the fundamental approach of using a keyword-related index, which enable the quickly search of documents that contain a given keyword. To meet the requirements of practical uses, conjunctive multi-keyword search is necessary. Moreover, to give the search user a better search experience, some proposals [12], [8] propose to enabled ranked results instead of returning undifferentiated results, by introducing the relevance score to the searchable encryption. To further improve the user experience, fuzzy keyword search over the encrypted data has also been developed in [7] and [14].

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. [10] adopt the inverted index *TSet*, which maps the keyword to the documents containing it, to achieve efficient multi-keyword search for large-scale datasets. The works is later extended the implementation on real-world datasets. The ranked result is not supported. Naveed et.al. [13] construct a blind storage system to achieve searchable encryption and conceal the access pattern of the search user. Only single-keyword search is supported.

ISSN 2395-695X (Print)
ISSN 2395-695X (Online)
International Journal of Advanced Research in Biology Engineering Science and Technology (IJARBEST)

Vol. 2, Special Issue 10, March 2016

3. PLAINTEXT AND CIPHERTEXT SEARCHING MODEL

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.

$$score(w, F_d) = \frac{1}{|F_d|} \cdot (1 + \ln f_{d,w}) \cdot ln\left(1 + \frac{N_d}{f_w}\right) \quad (1)$$

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.

ISSN 2395-695X (Print)
ISSN 2395-695X (Online)
International Journal of Advanced Research in Biology Engineering Science and Technology (IJARBEST)
Vol. 2, Special Issue 10, March 2016

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). 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 compares it with the hash values of all keywords in the index I, then the desired documents which are corresponding to keyword w are found. Next, 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 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.

4. SECURITY ANALYSIS ON CLOUD DATA SEARCH

Nowadays users connected to the Internet may store their data on cloud servers and let the servers manage or process their data. They can enjoy convenient and efficient service without paying too much money and energy, as one of the most attractive feature of cloud computing is its low cost. No matter how advantageous cloud computing may sound, large number of people still worry about the safety of this technology. If cloud servers get direct access to all these users' data, it may try to analyze the documents to get private information. The initial purpose of this action may be kind. The server wants to provide better service by digging into these data and then displaying customer-oriented advertisement, which could be convenient but also annoying. We consider sensitive data such as personal health records and secret chemical ingredients, the situation becomes even more serious. Theoretically, the server is not supposed to have access to sensitive data at all; therefore we should ensure the server has no access to leaking these data to an untrusted third party. Thus, sensitive data have to be encrypted before being outsourced to a commercial public cloud.

ISSN 2395-695X (Print)
ISSN 2395-695X (Online)
International Journal of Advanced Research in Biology Engineering Science and Technology (IJARBEST)
Vol. 2, Special Issue 10, March 2016

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. 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.

Deterministic encryption leaks the distribution of the plaintexts, so it cannot ensure data privacy in most applications. 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 prospect 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. So in this system propose attack on the One-to-Many OPE. Our experimental results when applying this attack to the secure keyword search scheme, the

cloud server can get an estimation of the distribution of the relevance scores and furthermore accurately reveal the encrypted keywords.

5. ISSUES ON CLOUD DATA SEARCH PROCESS

Encrypted cloud storage is used to share user data with security and privacy. Ranked 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 prospect OPE Scheme. Binary search algorithm is applied to perform document search on encrypted data environment. Attack on one-to-many OPE is initiated by exploiting the differences of the ordered ciphertexts. The following issues 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

6. SECURITY ENHANCEMENT OF PROSPECT OPE

The Prospect OPE based scheme is enhanced with security measures to handle 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 Prospect 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, Prospect 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 Prospect OPE module. Attack analysis module is used to discover the attacks. Index values are protected using index distribution security process. Query process is called to perform encrypted data search process.

6.1. Cloud Server

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.

6.2. Relevance Score Assignment

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.

6.3. Prospect OPE

Prospect 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.

6.4. Attack Analysis

Attack analysis is initiated to verify the index distribution levels. Attacks are discovered with distribution relationship values. Index subsets are analyzed in the attack analysis process. Attacks are discovered with query keyword intervals.

6.5. Index Distribution Security

Index distribution security process is used to control 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

6.6. Query Process

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. Semantic relationship based search scheme is integrated with the query process

7. 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. Attack handling mechanism is integrated with the prospect 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.

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] D. X. Song, D. Wagner and A. Perrig, "Practical techniques for searches on encrypted data," in *Proc. IEEE Symp. Secur. Privacy*, May 2000, pp. 44-55.

[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] D. Boneh and B. Waters, "Conjunctive, subset and range queries on encrypted data," in *Proc. TCC*, 2007, pp. 535-554.

[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.

[14] J. Li, Q. Wang, C. Wang, N. Cao, K. Ren and W. Lou, "Fuzzy keyword search over encrypted data in cloud computing," in *Proc. IEEE INFOCOM*, Mar. 2010, pp. 1-5.

[15] Hongwei Li, Dongxiao Liu, Yuanshun Dai, 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.