ISSN 2395-1621

Secure Search of Cloud Data by Encrypted Relevance Scores

[#1]Leena S. Shewale

[1]leenashewale123@gmail.com

[#1]Department of Computer Engineering, JSPM's ICOER Wagholi, Pune, Savitribai Phule Pune University, Pune, India

ABSTRACT

The translation of data into a secret code. Encryption is the most effective way to achieve data security. To read an encrypted file, you must have access to a secret key or password that enables you to decrypt it. Unencrypted data is called plain text ; encrypted data is referred to as cipher text. Order-preserving encryption allows encrypting data, while still enabling efficient range queries on the encrypted data. This makes its performance and functionality very suitable for data outsourcing in cloud computing scenarios, but the security of order-preserving is still debatable. We present a scheme that achieves a strictly stronger notion of security than any other scheme so far. The basic idea is to randomize the cipher texts to hide the frequency of plaintexts. Still, the client storage size remains small, As a result, one can more securely outsource large data sets, since we can also show that our security increases with larger data sets. We proposed a differential attack on One-to-Many OPE by exploiting the differences of the ordered cipher texts. The experimental results show that the cloud server can get a good estimate of the distribution of relevance scores by a differential attack. Furthermore, when having some background information on the outsourced documents, the cloud server can accurately infer the encrypted keywords by using the estimated distributions.

Keywords— Encryption, Data Security, Cloud Computing, Order Preserving Encryption

ARTICLE INFO

Article History

Received :4th January 2016

Received in revised form :

5th January 2016

Accepted : 6th January,2016

Published online :

6th January, 2016

I. INTRODUCTION

In recent era, cloud services are used by many users as well as industries. Cloud provides large amount of space to store data as well as share data so that it can be available any time over network when user requires. Cloud provides such services in low cost. Compared to traditional technologies, the cloud has many specific features, such as its large scale and the fact that resources belonging to cloud providers are completely distributed, heterogeneous and totally virtualized [1]. Users can store as well as share pictures, videos or any file over cloud so that it can be accessed on demand.The data stored over cloud has security issues; it is vulnerable to security threats. User can store any sensitive information over cloud. If cloud server get direct access to all these users' data, it may try to analyse 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. Besides, when we consider sensitive data such as personal health records and secret chemical ingredients, the situation becomes even more serious [2].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 [3].However, 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 cipher text 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) [4]Applying order preserving encryption (OPE) [5] 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 $x1 < x2$, then the corresponding ciphertexts $E(x1)$ and $E(x2)$ satisfy $E(x1) < E(x2)$. Boldyreva et al. initiated the cryptographic study of OPE schemes [6], [7], in which they defined the security of OPE and proposed a provably secure OPE scheme. However, the security definition and the constructions of OPE in [6], [7] 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. However, 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 [8].However, we discover that the One-to-Many OPE [8] 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 paper, we propose a differential attack on the One-to-Many OPE. Our experimental results show that, when applying this attack to the secure keyword search scheme of [8], the cloud server can get an estimation of the distribution of the relevance scores, and furthermore accurately reveal the encrypted keywords.

II. LITERATURE SURVEY

T. Grance and P. Mell al has focused on issues related to cloud. paper mainly focuses on the issues related to Privacy in cloud computing. Privacy is defined as a fundamental human right related to the collection, use, disclosure, storage and destruction of personal data (Personally Identifiable Information-PII). The American Institute of Certified Public Accountants (AICPA) and Canadian Institute of Charted Accountants (CICA) define that it is the right and obligation of individuals and organizations with respect to the collection, use, retention, and disclosure of personal information. Privacy is the protection of appropriate use of personal information of cloud user. [1]

D. Catalano et al has proposed the system which identify and fills some gaps with regard to consistency (the extent to which false positives are produced) for public-key encryption with key word search (PEKS). We define computational and statistical relaxations of the existing notion of perfect consistency, show that the scheme of is computationally consistent, and provide a new scheme that is statistically consistent. We also provide a transform of an anonymous IBE scheme to a secure PEKS scheme that, unlike the previous one, guarantees consistency.

Dawn Song et al have presented the system which store data on data storage servers such as mail servers and file servers in encrypted form to reduce security and privacy risks. But this usually implies that one has to sacrifice functionality for security. For example, if a client wishes to retrieve only documents containing certain words, it was not previously known how to let the data storage server perform the search and answer the query without loss of data confidentiality. [9]

S.Buyrukbilen et al [10], introduce the first method that provides ranked results from multi-keyword searches on public-key encrypted data. By avoiding a linear scan of the documents and by parallelizing the computations to the possible extent, this method reduces the computational complexity of public key cryptosystem.

Wenhai Sun et al have proposed a MRSE scheme that works on similarity based ranking. Here search index is created on the basis of term frequency and vector space. Search index is used for multi keyword search and ranking the search result. Search efficiency is improved by applying tree structure on index.

Ruturaj Desai, et al has proposes a new scheme to solve the problem of multi keyword search over encrypted data using trusted third party in cloud computing. User will encrypt their data locally. Before encrypting data, the index will be created. Trusted third party will use all these indexes to search data similar to the search query of user. Using these search results, cloud server will send encrypted document to the user.

III. PROBLEM STATEMENT AND PROPOSED SYSTEM

A. Problem Statement

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. Large number of people still worries about the safety of this technology. If cloud server get direct access to all these users' data, it may try to analyses the documents to get private information. The initial purpose of this action may be kind. Traditional searchable encryption schemes allow a user to securely search over encrypted data through keywords without first decrypting it, these techniques support only conventional Boolean keyword search without capturing any relevance of the files in the search result.

B. Proposed System

In ranked search of encrypted cloud data, probabilistic OPE is needed to preserve the order of relevance scores and conceal their distributions at the same time. One-to-Many OPE is a scheme designed for such a purpose. However, in this paper, we demonstrate that the cloud server can estimate the distribution of relevance scores by change point analysis on the differences of ciphertexts of One-to-Many OPE. Furthermore, the cloud server may identify what the encrypted keywords are by using the estimated distributions and some background knowledge. On the other hand, some methods can be used to resist the proposed attack. One is to improve the One-to-Many OPE itself. For instance, we can divide plaintexts having the same value into several sets and divide the corresponding bucket into several sub-buckets. By mapping each plaintext set into one sub- bucket, some new change points will appear in the differential attack, which will cover up the original distribution of plaintexts. Another possible method is to add noise into the inverted

index by adding some dummy documents IDs and keywords, and forging corresponding relevance scores. In our future work, we will elaborate these ideas to design secure methods of probabilistic OPE and schemes for search in encrypted data.

IV. ARCHITECTURE OF PROPOSED SYSTEM MODEL

The main task of this system is to provide fine grained access in authorization time period and self destruction of data after expiration of access time. Framework of retrieval over encrypted cloud data Fig 1.

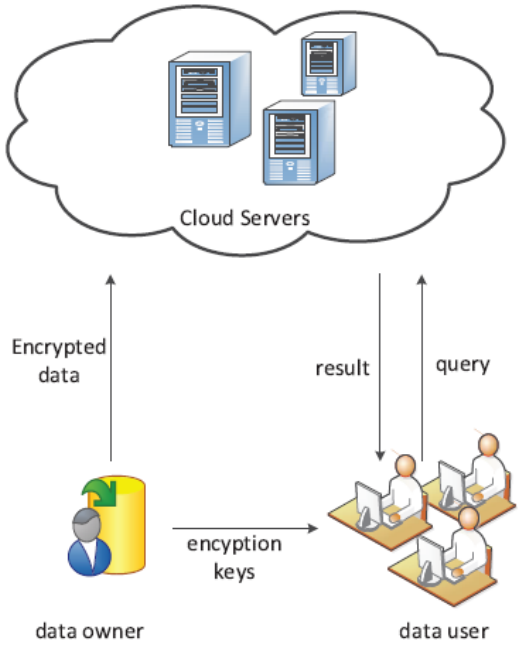

Figure1. Framework of retrieval over encrypted cloud data

1) *Data Owner*: A data owner can be an individual or a corporation, i.e., it is the entity that owns a collection of documents *Dc = fD1;D2 : : :DNd g* that it wants to share with trusted users. The keyword set is marked as *W = fw1;w2 : : :wNw g*.
2) *Cloud Servers*: It is the place of hardware and software resources where a pool of data files and different applications can store. 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.
3) *Data Users*: 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.

V. MATHEMATICAL MODEL

Let us consider a set S
Where, S= {U, R, SER, D, Encryption}
Here, S: System which includes:
U: Set of Users Where U= {U1, U2, U3 …, Un}
SER: Server.
R: Set of Request,
Where R= {R1, R2, R3,…….,Rn}.
Owner Upload data,
Data Encryption(data)
User request for data
R1 query (keyword) Ranking
Decryption (data) download

VI. IMPLEMENTATION STRATEGY AND EXPERIMENTAL SETUP

To implement this system One To Many OPE scheme is used in this paper. One To Many OPE is implemented by using four algorithms: Setup, Encrypt Multi-keyword and Decrypt.
•Setup: This algorithm takes security parameter and attribute as input and outputs the system public parameters and master keys. As this algorithm is run by Authority, so after getting output, Authority provides parameter publically by keeping master keys secrete.
•Encryption: Before encryption all keywords from documents are extracted and used to build index tree.
Encryption process uses master key and data owners secure private key to create encryption key at runtime. Encryption key will be used to encrypt the document. This algorithm generates cipher text of sensitive document and uploaded to cloud server.
•Multi-keyword search: Data user builds search query of multiple keywords. This search query is encrypted with master key and sent to cloud server for document retrieval.
Decrypt: Input to this algorithm is ciphertext (generated in Encrypt step) and private key (generated in KeyGen step). When all the attribute in access tree satisfies time instant then this algorithm decrypts the ciphertext and generate plain text. This plain text is nothing but the original message.

By following this algorithm a proposed system can be generated.

VII. CONCLUSION

In ranked search of encrypted cloud data, probabilistic OPE is needed to preserve the order of relevance scores and conceal their distributions at the same time. One-to-Many OPE is a scheme designed for such a purpose. However, in this paper, we demonstrate that the cloud server can estimate the distribution of relevance scores by change point analysis on the differences of ciphertexts of One-to-Many OPE. Furthermore, the cloud server may identify what the encrypted keywords are by using the estimated distributions

and some background knowledge.On the other hand, some methods can be used to resist the proposed attack. One is to improve the One-to-Many OPE itself. For instance, we can divide plaintexts having the same value into several sets and divide the corresponding bucket into several sub-buckets. By mapping each plaintext set into one sub-bucket, some new change points will appear in the differential attack, which will cover up the original distribution of plaintexts. Another possible method is to add noise into the inverted index by adding some dummy documents IDs and keywords, and forging corresponding relevance scores.

VIII. REFRENCES

[1] P. Mell and T. Grance, "The NIST definition of cloud computing," NIST special publication, 800(145): 7, 2011.

[2] S. Subashini and V. Kavitha, "A survey on security issues in service delivery models of cloud computing," Journal of Network and Computer
Applications, 34(1): 1-11, 2011.

[3] B. Krebs, "Payment processor breach may be largest ever,"http://voices.washingtonpost.com/securityfix/2009/01/ payment processor breach may b. html, 2009.

[4] M. Abdalla, M. Bellare and D. Catalano, "Searchable encryption revisited: Consistency properties, relation to anonymous IBE, and extensions," Advances in Cryptology–CRYPTO, 2005. Springer Berlin Heidelberg, pp. 205-222, 2005.

[5] R. Agrawal, J. Kiernan and R. Srikant, "Order preserving encryption for numeric data," Proceedings of the 2004 ACM SIGMOD international conference on Management of data. ACM, pp. 563-574, 2004.

[6]] A. Boldyreva, N. Chenette and Y. Lee , "Order-preserving symmetric encryption," Advances in Cryptology-EUROCRYPT, 2009. Springer Berlin Heidelberg, pp. 224-241, 2009.

[7] A. Boldyreva, N. Chenette and A. ONeill, "Order-preserving encryption revisited: improved security analysis and alternative solutions," Advances in Cryptology CCRYPTO, 2011. Springer Berlin Heidelberg, pp. 578-595, 2011.

[8] C. Wang, N. Cao and K. Ren, "Enabling secure and efficient ranked keyword search over outsourced cloud data," Parallel and Distributed Systems, IEEE Transactions 23(8), pp. 1467-1479, 2012.

[9] D. Song, D. Wagner and A. Perrig, "Practical techniques for searches on encrypted data," Security and Privacy, 2000. Proceedings. 2000 IEEE Symposium on. IEEE, pp. 44-55, 2000.

[10] S. Buyrukbilen "Privacy-Preserving Ranked Search on Public-Key Encrypted Data", Security and Privacy, 2013.

