Towards Secure Top-k Multi keyword Retrieval over encrypted cloud data

Lekshmi k pillai, P.Saranya,

PG Student
Affiliated Anna University Chennai,Dept of computer science and enginnering
Gnanamani college of engineering
Namakkal .india
lekshmikttnpll@gmail.com

Asst Professor Dept of computer science and engineering
Gnanamani college of engineering
Namakkal .india
info@gce.org.in

Abstract: Cloud computing has emerging as a promising pattern for data outsourcing and high-quality data services. However, a concern of sensitive information on cloud potentially causes privacy problems. Data encryption protects data security to some extent, but at the cost of compromised efficiency. To observe that server-side ranking based on order-preserving encryption (OPE) inevitably leaks data privacy. To eliminate the leakage, traditional method uses a two-round searchable encryption (TRSE) scheme that supports top-k multikeyword retrieval. The TRSE method employs a vector space model and homomorphic encryption. The vector space model helps to provide sufficient search accuracy, and the homomorphic encryption enables users to involve in the ranking while the majority of computing work is done on the server side by operations only on cipher text. As a result, information leakage can be eliminated and data security is ensured but, user side Top-K retrieval process is based on user's click event, sometimes it is difficult to users. So we propose a method for Top-K retrieval. In this method user gets an interested/used link in the top.

Keywords: about four key words separated by commas**.**

1. Introduction

2. The main threat on data privacy roots in the cloud itself. When users outsource their private data onto the cloud, the cloud service providers are able to control and monitor the data and the communication between users and the cloud at will, lawfully or unlawfully. Instances such as the secret NSA program, working with AT&T and Verizon, which recorded over 10 million phone calls between American citizens, cause uncertainty among privacy advocates, and the greater powers it gives to telecommunication companies to monitor user activity. To ensure privacy, users usually encrypt the data before outsourcing it onto cloud, which brings great challenges to effective data utilization. However, even if the encrypted data utilization is possible, users still need to communicate with the cloud and allow the cloud operate on the encrypted data, which potentially causes leakage of sensitive information.

Existing System

The existing method introduced the concepts of similarity relevance and scheme robustness to formulate the privacy issue in searchable encryption schemes, and then solve insecurity problem by two-round searchable encryption (TRSE) scheme. Novel Technologies in the cryptography community and information retrieval community are employed including homomorphic encryption and vector space model. The majority of computing work is done on the cloud while the user takes part in ranking.

Disadvantages

- No secure for server side data

Proposed System

In existing method the retrieval result is based on user-click event or ranking based process for every search. We propose a method for Top-K retrieval, in this method user search query & keyword is same as already they searched query & keyword means the already selected content click will displayed in the top link and then all other ranking based links are display.

Advantages

- Third party can't understand server side data's .

Modules:

1. Cloud based data privacy
2. Protection of homomorphic encryption
3. K-Similarity relevance
4. Server side ranking model
5. Key Request & Key Verification
6. High security and practical efficiency

Modules Description

Cloud based Data Privacy

The main threat on data privacy roots in the cloud itself. When users outsource their private data onto the cloud, the cloud service providers are able to control and monitor the data and the communication between users and the cloud at will, lawfully or unlawfully. To ensure privacy, users usually encrypt the data before outsourcing it onto cloud, which brings great challenges to effective data utilization.

Protection of Homomorphic Encryption

To alleviate the computational burden on the user side, computing work should be at the server side, so we need an encryption scheme to guarantee the operability and security at the same time on server side. Homomorphic encryption allows specific types of computations to be carried out on the corresponding ciphertext. The result is the ciphertext of the result of the same operations performed on the plaintext. That is, homomorphic encryption allows computation of ciphertext without knowing anything about the plaintext to get the correct encrypted result. Fortunately, as a result of employing the vector space model to top-k retrieval, only addition and multiplication operations over integers are needed to compute the relevance scores from the encrypted searchable index.

K-Similarity Relevance and Server Side Ranking Model

A server-side ranking based on OPE violates the privacy of sensitive information, which is considered uncompromisable in the security-oriented third- party cloud computing scenario, i.e., security cannot be tradeoff for efficiency. To achieve data privacy, ranking has to be left to the user side. Thus, the user-side ranking schemes are

challenged by practical use. A more server-siding scheme might be a better solution to privacy issues.

We propose a new searchable encryption scheme, in which novel technologies in cryptography community and IR community are employed, including homomorphic encryption and the vector space model. In the proposed scheme, the data owner encrypts the searchable index with homomorphic encryption. When the cloud server receives a query consisting of multikeywords, it computes the scores from the encrypted index stored on cloud and then returns the encrypted scores of files to the data user. Next, the data user decrypts the scores and picks out the top-k highest scoring files' identifers to request to the cloud server. The retrieval takes a two-round communication between the cloud server and the data user. We, thus, name the scheme the TRSE scheme, in which ranking is done at the user side while scoring calculation is done at the server side.

Key Request and Key Verification

Key Request:

The user can view all the relevant result and send request to server for getting secret code of that selected result, after that server view user request and sent the secret code to target reader's email id. This key will be used when user like to view the whole content of the result

Key Verification

Key verification is the process of checking whether the readers are authenticated or not based on the secret key, if it is matched they view the file; otherwise they can't proceed further process.

High Security and Practical Efficiency

A TRSE scheme employing the fully homomorphic encryption, which fulfills the security requirements of multikeyword top-k retrieval over the encrypted cloud data. By security analysis, we show that the proposed scheme guarantees data privacy.

Literature Survey

1.Fully Homomorphic Encryption over the Integers

We describe a very simple "somewhat homomorphic" encryption scheme using only elementary modular arithmetic, and use Gentry's techniques to convert it into a fully homomorphic scheme. Compared to Gentry's construction, our somewhat homomorphic scheme merely uses addition and multiplication over the integers rather than working with ideal lattices over a polynomial ring. The main appeal of our approach is the conceptual simplicity. We reduce the security of our somewhat homomorphic scheme to finding an approximate integer gcd – i.e., given a list of integers that are near-multiples of a hidden integer, output that hidden integer. We investigate the hardness of this task, building on earlier work of Howgrave-Graham.

2.Privacy-Preserving Multi-keyword Ranked Search over Encrypted Cloud Data

The advent of cloud computing, data owners are motivated to outsource their complex data management systems from local sites to commercial public cloud for great flexibility and economic savings. But for protecting data privacy, sensitive data has to be encrypted before outsourcing, which obsoletes traditional data utilization based on plaintext keyword search. Thus, enabling an encrypted cloud data search service is of paramount importance. Considering the large number of data users and documents in cloud, it is crucial for the

search service to allow multi-keyword query and provide result similarity ranking to meet the effective data retrieval need. Related works on searchable encryption focus on single keyword search or Boolean keyword search, and rarely differentiate the search results. In this paper, for the first time, we define and solve the challenging problem of privacy-preserving multi-keyword ranked search over encrypted cloud data (MRSE), and establish a set of strict privacy requirements for such a secure cloud data utilization system to become a reality. Among various multi-keyword semantics, we choose the efficient principle of "coordinate matching", i.e., as many matches as possible, to capture the similarity between search query and data documents, and further use "inner product similarity" to quantitatively formalize such principle for similarity measurement. We first propose a basic MRSE scheme using secure inner product computation, and then significantly improve it to meet different privacy requirements in two levels of threat models. Thorough analysis investigating privacy and efficiency guarantees of proposed schemes is given, and experiments on the real-world dataset further show proposed schemes indeed introduce low overhead on computation and communication.

3.Practical Techniques for Searches on Encrypted Data□

It is desirable to 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.

In this paper, we describe our cryptographic schemes for the problem of searching on encrypted data and provide proofs of security for the resulting crypto systems. Our techniques have a number of crucial advantages. They are provably secure: they provide provable secrecy for encryption, in the sense that the untrusted server cannot learn anything about the plaintext when only given the ciphertext; they provide query isolation for searches, meaning that the untrusted server cannot learn anything more about the plaintext than the search result; they provide controlled searching, so that the untrusted server cannot search for an arbitrary word without the user's authorization; they also support hidden queries, so that the user may ask the untrusted server to search for a secret word without revealing the word to the server. The algorithms we present are simple, fast and introduce almost no space and communication overhead, and hence are practical to use today.

4.Secure Conjunctive Keyword Search Over Encrypted Data

We study the setting in which a user stores encrypted documents (e.g. e-mails) on an untrusted server. In order to retrieve documents satisfying a certain search criterion, the user gives the server a *capability* that allows the server to identify exactly those documents. Work in this area has largely focused on search criteria consisting of a single keyword. If the user is actually interested in documents containing each of several keywords (*conjunctive* keyword search) the user must either give the server capabilities for each of the keywords individually and rely on an intersection calculation (by either the server or the user) to determine the

correct set of documents, or alternatively, the user may store additional information on the server to facilitate such searches. Neither solution is desirable; the former enables the server to learn which documents match each individual keyword of the conjunctive search and the latter results in exponential storage if the user allows for searches on every set of keywords.

We define a security model for conjunctive keyword search over encrypted data and present the first schemes for conducting such searches securely. We propose first a scheme for which the communication cost is linear in the number of documents, but that cost can be incurred "offine" before the conjunctive query is asked. The security of this scheme relies on the Decisional Diffie-Hellman (DDH) assumption. We propose a second scheme whose communication cost is on the order of the number of keyword fields and whose security relies on a new hardness assumption.

5.Enabling Secure and Efficient Ranked Keyword Search over Outsourced Cloud Data

Cloud computing economically enables the paradigm of data service outsourcing. However, to protect data privacy, sensitive cloud data has to be encrypted before outsourced to the commercial public cloud, which makes effective data utilization service a very challenging task. Although traditional searchable encryption techniques allow users to securely search over encrypted data through keywords, they support only Boolean search and are not yet sufficient to meet the effective data utilization need that is inherently demanded by large number of users and huge amount of data files in cloud. In this paper, we define and solve the problem of secure ranked keyword search over encrypted cloud data. Ranked search greatly enhances system usability by enabling search result relevance ranking instead of sending undifferentiated results, and further ensures the file retrieval accuracy. Specifically, we explore the statistical measure approach, i.e. relevance score, from information retrieval to build a secure searchable index, and develop a one-to-many order-preserving mapping technique to properly protect those sensitive score information. The resulting design is able to facilitate efficient server-side ranking without losing keyword privacy. Thorough analysis shows that our proposed solution enjoys "as-strong-as-possible" security guarantee compared to previous searchable encryption schemes, while correctly realizing the goal of ranked keyword search. Extensive experimental results demonstrate the efficiency of the proposed solution.

6. Searchable Symmetric Encryption: Improved Definitions and Efficient Constructions

Searchable symmetric encryption (SSE) allows a party to outsource the storage of his data to another party in a private manner, while maintaining the ability to selectively search over it. This problem has been the focus of active research and several security definitions and constructions have been proposed. In this paper we begin by reviewing existing notions of security and propose new and stronger security definitions. We then present two constructions that we show secure under our new definitions. Interestingly, in addition to satisfying stronger security guarantees, our constructions are more efficient than all previous constructions.

7.Processing Private Queries over Untrusted Data Cloud through Privacy Homomorphism

Query processing that preserves both the data privacy of the owner and the query privacy of the client is a new research problem. It shows increasing importance as cloud computing drives more businesses to outsource their data and querying services. However, most existing studies, including those on data outsourcing, address the data privacy and query privacy separately and cannot be applied to this problem. In this paper, we propose a holistic and efficient solution that comprises a secure traversal framework and an encryption scheme based on privacy homomorphism. The framework is scalable to large datasets by leveraging an index-based approach. Based on this framework, we devise secure protocols for processing typical queries such as k-nearest-neighbor queries (kNN) on R-tree index. Moreover, several optimization techniques are presented to improve the efficiency of the query processing protocols. Our solution is verified by both theoretical analysis and performance study.

References

M. Armbrust, A. Fox, R. Griffith, A. Joseph, R. Katz, A. Konwinski, G. Lee, D. Patterson, A. Rabkin, and M. Zaharia, "A View of Cloud Computing," Comm. ACM, vol. 53, no. 4, pp. 50-58, 2010.

[2] M. Arrington, "Gmail Disaster: Reports of Mass Email Deletions," http://www.techcrunch.com/2006/12/28/gmail-disasterreportsof- mass-email-deletions/, Dec. 2006.

[3] Amazon.com, "Amazon s3 Availability Event: July 20, 2008," http://status.aws.amazon.com/s3-20080720.html, 2008.

[4] RAWA News, "Massive Information Leak Shakes Washington over Afghan War," http://www.rawa.org/temp/runews/2010/08/20/massive-information-leak-shakes-washington overafghan- war.html, 2010.

[5] AHN, "Romney Hits Obama for Security Information Leakage," http://gantdaily.com/2012/07/25/romney-hits-obama-forsecurity- information-leakage/, 2012.

[6] Cloud Security Alliance, "Top Threats to Cloud Computing," http://www.cloudsecurity alliance.org, 2010.

[7] C. Leslie, "NSA Has Massive Database of Americans' Phone Calls," http://usatoday30.usatoday.com/news/washington/2006-05-10/, 2013.

[8] R. Curtmola, J.A. Garay, S. Kamara, and R. Ostrovsky, "Searchable Symmetric Encryption: Improved Definitions and Efficient Constructions," Proc. ACM 13th Conf. Computer and Comm. Security (CCS), 2006.

[9] C. Wang, N. Cao, J. Li, K. Ren, and W. Lou, "Secure Ranked Keyword Search over Encrypted Cloud Data," Proc. IEEE 30th Int'l Conf. Distributed Computing Systems (ICDCS), 2010.

[10] S. Zerr, D. Olmedilla, W. Nejdl, and W. Siberski, "Zerber+r: Top-k Retrieval from a Confidential Index," Proc. 12th Int'l Conf. Extending Database Technology: Advances in Database Technology (EDBT), 2009.