IJESRT

INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES & RESEARCH TECHNOLOGY

INFORMATION SECURITY AND SECURE SEARCH OVER ENCRYPTED DATA IN CLOUD STORAGE SERVICES

Mr. A Mustagees Shaikh [*], Prof. Nitin B. Raut

Student Pursuing M.TECH. in Computer Science & Engineering,GNIET,Nagpur, India

Prof. at Department of Computer Science & Engineering, GNIET,Nagpur, India.

DOI: 10.5281/zenodo.49753

ABSTRACT

Cloud computing is most widely used as the next generation architecture of IT enterprises, that provide convenient remote access to data storage and application services. This cloud storage can potentially bring great economical savings for data owners and users, but due to wide concerns of data owners that their private data may be exposed or handled by cloud providers. Hence end-to-end encryption techniques and fuzzy fingerprint technique have been used as solutions for secure cloud data storage. In this project we use searchable encryption techniques, which allows encrypted data to be searched by users without leaking information about the data itself and user's queries. We build a secure searchable index, and develop a one to many order preserving mapping technique to protect those sensitive score information. The resulting design is able to facilitate efficient server side ranking without losing keyword privacy. Hence to avoid loose and loss of data, we use privacy preserving data-leak detection (DLD) solution to solve the issue where sensitive data digests is used in detection. The advantage of this method is that it enables the data owner to safely delegate the detection operation to a semi honest provider without revealing the sensitive data to the provider.In this project, we identify the challenges towards achieving privacy in searchable outsourced cloud data services and we use the DLD solution which helps to detect the leak data. This technique helps us to save securely our sensitive data in cloud storage and retrieve this data while we required without leaking our private data through Information Security and Secure Search over Encrypted Data in Cloud Storage Services.

KEYWORDS: Cloud Storage, Data Owner, Data User, Data leak, Network Security, Privacy, Collection Intersection.

INTRODUCTION

In this computing platform, cloud computing protects data privacy, sensitive cloud data has to be encrypted before being outsourced to the commercial public cloud. Traditional searchable encryption techniques allow users to securely search over encrypted data through keywords, they support only Boolean search, which is not sufficient to meet the effective data utilization need that is inherently demanded by large number of users. In this paper, we solve the problem of secure ranked keyword search over encrypted cloud data. Specially, we explore the statistical measure approach, i.e. relevance score, from information retrieval to build a secure searchable index, and also protect those sensitive score information. A privacy approach for owners to take back control of their data is to adopt end-to-end data encryption. The fact that data owners and cloud server are no longer in the same trusted domain may put the outsourced unencrypted data at risk the cloud server may leak data information to unauthorized entities [1] or even be hacked [2]. Therefore, to build a full-fledged cloud data service, it is highly desirable to enable privacy assured search over encrypted data, which ideally does not leak any sensitive user information to the cloud, such as business secrets or private personal activities. Detecting and preventing data leaks requires a set of complementary solutions, which may include data-leak detection [3], [4], data confinement, stealthy malware detection and policy enforcement [5].We propose a data-leak detection solution which can be outsourced and be deployed in a semi honest detection environment. We design, implement, and evaluate our *fuzzy fingerprint* technique that enhances data privacy during data-leak detection operations. Our approach is based on a fast and practical one-way computation on the sensitive data. It enables the data owner to securely delegate the content-inspection task to DLD providers without exposing the sensitive data. Hence, in our detection procedure, the data owner computes a special set of digests or fingerprints from the sensitive data and then discloses only a small amount of them to the DLD provider. The DLD provider computes fingerprints from network traffic and identifies potential leaks in them. To prevent the DLD provider from gathering exact knowledge about the sensitive data, the collection of potential leaks is composed of real leaks and noises. It is the data owner, who post processes the potential leaks sent back by the DLD provider and determines whether there is any real data leak.

MATERIALS AND METHODS

Related Techniques:

First, we briefly discuss and compare several existing techniques, and their relevance to the privacy-assured cloud-based search problem.

- *Secure multiparty computation (SMC):* In SMC, each party *Pi* possess some private input *xi*, and every party computes some (public) function *f*(*x*1, …, *xn*) without revealing *xi* to others, except what can be derived from the input and output.
- *Private information retrieval (PIR):* PIR involves two parties: a client and a server. In asymmetric PIR, the server hosts a public database D, while the client retrieves a record *i* from D without revealing *i* to the server. In symmetric PIR (a.k.a. oblivious transfer), the non-retrieved records should also be withheld from the client, which can be regarded as a special case of SMC.
- *Searchable encryption (SE):* SE also involves a client and a server, where the latter stores an encrypted database ~ D, and the former possesses a private query *Q* that wants to obtain the query result *Q*(D) without revealing both *Q* and plaintext D to the server.
- *Order-Preserving Symmetric Encryption (OPSE)***:** In OPSE [7], the numerical ordering of plaintext is preserved after encryption that provide the first cryptographic construction of OPSE that is provably secure under the security framework of pseudorandom function or pseudorandom permutation. It can be regarded as a function *g* (◊) from a domain D = {1…*M*} to a range R = {1…*N*}.

Methodology:

We describe a top-down approach in which the search functionality in the plaintext domain, one can decompose it into a certain data index structure and primitive data operations using relevant information retrieval (IR) principles.

Model and Overview:

We begin by describing a general cloud data storage service architecture involving three (types of) entities (Fig. 1). The *data owner* (or data contributor) is one or multiple entities who generate and encrypt data, and upload them to the cloud server. The owner can be either an organization or an individual. The *cloud server* belonging to a CSP possesses significant storage and computation resources, and provides them to end users in a payperuser manner. There are one or more *data users* in the system, who may need to perform queries over the outsourced data in order to extract useful information. In addition, in order to enable public auditing, a third party auditor can be employed, which is discussed in [6] and is outside the scope of this article. The owner's data are encrypted end-to-end using secret keys created by him/herself, and a searchable index is usually created and encrypted along with the outsourced data. To allow data access and search by users, the data owner usually generates and distributes search tokens (or trapdoors), which are encrypted queries to users, either actively or upon users' requests. When a user wants to gain file access or initiate a query, he/she submits a corresponding token to the server, who then returns a matching set of documents in an encrypted format. In some situations, the data user and data owner can be the same physical entity.

First we describe the two most important players in our abstract model: the organization (i.e., data owner) and the data-leak detection (DLD) provider.

- Organization owns the sensitive data and authorizes the DLD provider to inspect the network traffic from the organizational networks for anomalies, namely inadvertent data leak. However, the organization does not want to directly reveal the sensitive data to the provider.
- DLD provider inspects the network traffic for potential data leaks. The inspection can be performed offline without causing any real-time delay in routing the packets. However, the DLD provider may attempt to gain knowledge about the sensitive data.

Privacy-preserving keyword search [8] or fuzzy keyword search [9] provide string matching approaches in semi-honest environments, but keywords usually do not cover enough sensitive data segments for data-leak detection **pFigure**:

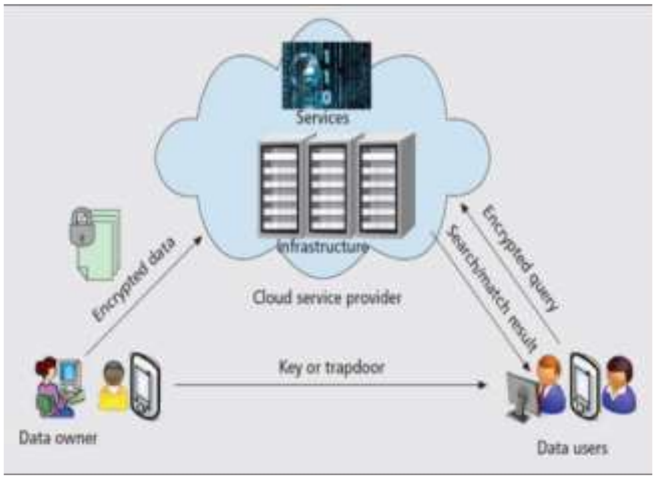

Fig. 1: System architecture for searchable cloud data storage services.

Within the scope of this article, we focus on how to enable privacyassured search for cloud data services. The above system architecture captures a wide range of searchable cloud data storage applications. In some scenarios, the data owner and user can be the same person; for example, Alice uploads her personal albums to Dropbox and wants to search for a particular photo afterward. Or if we consider a corporate data owner, a company may outsource its business records to a cloud server to enjoy the lowcost storage. At the same time, an employee in the auditing department may need to search the business database for records containing sensitive activities. Alternatively, the data owner may be an individual while the user can be a company. For instance, consider a pervasive healthcare application where each patient uploads her health monitoring data periodically to a third party medical service.

SYSTEM REQUIREMENTS:

Functional Properties:For data search, perhaps the most important property is *usability*, which is the basis for attracting customers. The current Google search is a great example of what is necessary in plaintext domain search. The following is an (incomplete, but typical) list of them:

- *Multi-keyword search*: The search condition should support Boolean expressions consisting of combinations of multiple keywords, including conjunctive normal form (CNF) and disjunctive normal form (DNF).
- *Result ranking*: The ranked search function greatly enhances the relevance of returned search results and reduces communication overhead, which is highly desirable for building usable cloud data services.
- *Error tolerance*: To accommodate various typos, representation inconsistencies, and so forth, search schemes should have a fuzzy nature. This means a search needs to also return relevant results for keywords within a certain edit distance from the input query.
- *Handle structured data*: A large portion of today's online data is represented using rich structures beyond simple text form, such as social network graphs. Without being able to utilize those structured data, the economic potential of cloud services will not be fully realized. We note that in the encrypted domain, it is very difficult for the above properties to be simultaneously achieved. We describe how the stateoftheart schemes achieve some combination of them.
- *Privacy Assurance* :In a searchable cloud storage service, both the owner's outsourced data and users' queries over those data may contain sensitive information and need protection against an adversary. More specifically, the system should meet the following privacy requirements:
- *Data and index confidentiality*: Without the secret key *K*, no one, including the cloud server, should be able to learn sensitive information from the owner's private data. Similarly, they should not be able to deduce sensitive information underlying the data index, because the index is often closely related to the data itself.
- *Query confidentiality*: Users' most important concern is to hide the search criteria on which they are evaluating the data (e.g., their query keywords). These should not be derivable from the search trapdoor and

data/index sent to the cloud server, even when the server possesses some additional background information such as keyword distribution. A higherlevel requirement is *query unlinkability*, that is, the cloud server shall not learn whether two queries have the same criteria. Note that this intrinsically requires the trapdoor to be nondeterministic.

- *Efficiency*: A privacyassured data search scheme should have low computation, communication, and storage overheads. For such a scheme to be deployed in a largescale cloud storage system with economic practicality, we argue that the search process should be completed within both constant communication round and computation time (independent of the database size). In general, the privacy guarantee conflicts with efficiency and functionality goals.

RESULTS AND DISCUSSION

We begin by describing a general cloud data storage service architecture involving three entities. The data owner (or data contributor) is one or multiple entities who generate and encrypt data, and upload them to the cloud server. The owner can be either an organization or an individual. The owner’s data are encrypted end-to-end using secret keys created by him/herself, and a searchable index is usually created and encrypted along with the outsourced data. To allow data access and search by users, the data owner usually generates and distributes search tokens (or trapdoors), which are encrypted queries to users, either actively or upon users’ requests. When a user wants to gain file access or initiate a query, he/she submits a corresponding token to the server, who then returns a matching set of documents in an encrypted format. In some situations, the data user and data owner can be the same physical entity.In our detection procedure, the data owner computes a special set of digests or fingerprints from the sensitive data and then discloses only a small amount of them to the DLD provider. The DLD provider computes fingerprints from network traffic and identifies potential leaks in them. To prevent the DLD provider from gathering exact knowledge about the sensitive data, the collection of potential leaks is composed of real leaks and noises. It is the data owner, who post-processes the potential leaks sent back by the DLD provider and determines whether there is any real data leak.

CONCLUSION

We identify the problem and challenges of enabling privacy assured searchable cloud data storage services, which suggest that achieving functionally rich, usable, and efficient search on encrypted data is possible without sacrificing privacy guarantee too much as well as we proposed fuzzy fingerprint, a privacy-preserving data-leak detection model and present its realization. Using special digests, the exposure of the sensitive data is kept to a minimum during the detection.

REFERENCES

[1] M. Chase and S. Kamara, ―Structured Encryption and Controlled Disclosure,Advances in Cryptology-ASIACRYPT 2010, 2010, pp. 577–94.

[2] R. Curtmola et al., ―Searchable Symmetric Encryption: Improved Definitions and Efficient Constructions,‖ Proc. ACM CCS ’06, 2006.

[3] Identity Finder. *Discover Sensitive Data Prevent Breaches DLP Data Loss Prevention*. [Online]. Available: http://www.identityfinder.com/, accessed Oct. 2014.

[4] K. Borders and A. Prakash, “Quantifying information leaks in outbound web traffic,” in *Proc. 30th IEEE Symp. Secur. Privacy*, May 2009,pp. 129–140.

[5] G. Karjoth and M. Schunter, “A privacy policy model for enterprises,”in *Proc. 15th IEEE Comput. Secur. Found. Workshop*, Jun. 2002,pp. 271–281.

[6] C. Wang et al., ―Toward Publicly Auditable Secure Cloud Data Storage Services,‖ IEEE Network, vol. 24, no. 4, July– Aug. 2010, pp. 19–24.

[7] A. Boldyreva et al., ―Order-Preserving Symmetric Encryption,‖ Proc. Eurocrypt ’09, LNCS, vol. 5479, Springer, 2009.

[8] S. Ananthi, M. Sadish Sendil, and S. Karthik, “Privacy preserving keyword search over encrypted cloud data,” in *Advances in Computing and Communications* (Communications in Computer and Information Science), vol. 190. Berlin, Germany: Springer-Verlag, 2011,pp. 480–487.

[9] J. Li, Q. Wang, C. Wang, N. Cao, K. Ren, and W. Lou, “Fuzzy keyword search over encrypted data in cloud computing,” in *Proc. 29th IEEE Conf. Comput. Commun.*, Mar. 2010, pp. 1–5.