Volume 4, Issue 12, December 2014 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

The Privacy-Assured and Searchable Cloud Data Storage Services

Prof. Rutuja Nitin Benkar*, Prof. Pranali Ramchandra Ekatpure, Prof. Savita Shivaji Waghmode
Sahakar Maharshi Shankaraao Mohite Patil Institute of Technology and Research, Solapur University, Yashwantnagar, Akluj, Maharashtra, India

Abstract*— Cloud computing is envisioned as the next generation architecture of IT enterprises, providing convenient remote access to massively scalable data storage and application services. While this outsourced storage and computing paradigm can potentially bring great economical savings for data owners and users, its benefits may not be fully realized due to wide concerns of data owners that their private data may be involuntarily exposed or handled by cloud providers. Although end-to-end encryption techniques have been proposed as promising solutions for secure cloud data storage, a primary challenge toward building a full-fledged cloud data service remains: how to effectively support flexible data utilization services such as search over the data in a privacy-preserving manner. In this article, we identify the system requirements and challenges toward achieving privacy-assured searchable outsourced cloud data services, especially, how to design usable and practically efficient search schemes for encrypted cloud storage. We present a general methodology for this using searchable encryption technique, which allows encrypted data to be searched by users without leaking information about the data itself and users' queries. In particular, we discuss three desirable functionalities of usable search operations: supporting result ranking, similarity search, and search over structured data. For each of them, we describe approaches to design efficient privacy-assured searchable encryption schemes, which are based on several recent symmetric-key encryption primitives. We analyse their advantages and limitations, and outline the future challenges that need.to be solved to make such secure searchable cloud data service a reality.

Keywords*— Cloud, Cloud computing Security, searchable encryption technique, symmetric-key encryption, cryptographic cloud storage.

I. INTRODUCTION

This The cloud has long been envisioned as the next generation IT architecture, which promises to provide massively scalable data storage and application services to society at a reduced cost, primarily attributed to the centralized management of elastic resources. In this emerging computing platform, the cloud provider, application developers, and end users can all reap benefits. One of the most attractive cloud services nowadays is data storage, where end users outsource large volumes of data to cloud servers to enjoy virtually unlimited hardware/software resources and ubiquitous access, without investing a large amount of capital up front for their own data warehousing and maintenance. Despite the tremendous business and technical advantages of the cloud, the security and privacy concern has been one of the major hurdles preventing its widespread adoption. Especially for outsourced data services, the owners' exclusive control over their data is ultimately relinquished to the CSPs. For example, Google's recent privacy policy implies that they essentially own the right to arbitrarily handle the uploaded user data. As a result, from the data owners' point of view, whenever their outsourced data contain sensitive personal information, such as financial and medical records, and social network profiles, it can no longer be considered as private as before.

On the other hand, although in reality CSPs usually enforce data security through mechanisms like firewalls and virtualization, these measures do not fully guard against threats of unauthorized data access from insiders, outsiders, or other cloud tenants due to the non-bug-free deployment and low degree of transparency. A promising approach for owners to take back control of their data is to adopt end-to-end data encryption (i.e., cryptographic cloud storage), which has been investigated by numerous researchers recently. However, while data encryption guarantees data confidentiality, it also rules out many routine manipulations over the data necessary in the plaintext domain. One fundamental requirement is to be able to perform search operations that can sort out relevant information from huge amounts of data.

II. LITERATURE SURVEY

A. Toward Privacy-Assured and Searchable Cloud Data Storage Services

In this article, we identify the system requirements and challenges toward achieving privacy-assured searchable outsourced cloud data services, especially, how to design usable and practically efficient search schemes for encrypted cloud storage. We present a general methodology for this using searchable encryption technique, which allows encrypted data to be searched by users without leaking information about the data itself and users' queries. In particular, we discuss three desirable functionalities of usable search operations: supporting result ranking, similarity search, and search over

© 2014, IJARCSSE All Rights Reserved

structured data. For each of them, we describe approaches to design efficient privacy-assured searchable encryption schemes, which are based on several recent symmetric-key encryption primitives.

B. Privacy-Preserving Multi-Keyword Ranked Search over Encrypted Cloud Data

In this paper, for the first time, we define and solve the challenging problem of privacy-preserving multi-keyword ranked search over encrypted data in cloud computing (MRSE). We establish a set of strict privacy requirements for such a secure cloud data utilization system. Among various multi-keyword semantics, we choose the efficient similarity measure of "coordinate matching," i.e., as many matches as possible, to capture the relevance of data documents to the search query. We further use "inner product similarity" to quantitatively evaluate such similarity measure. We first propose a basic idea for the MRSE based on secure inner product computation, and then give two significantly improved MRSE schemes to achieve various stringent privacy requirements in two different threat models.

C. Towards Secure and Dependable Storage Services in Cloud Computing

We propose in this paper a flexible distributed storage integrity auditing mechanism, utilizing the Homomorphic token and distributed erasure-coded data. The proposed design allows users to audit the cloud storage with very lightweight communication and computation cost. The auditing result not only ensures strong cloud storage correctness guarantee, but also simultaneously achieves fast data error localization, i.e., the identification of misbehaving server. Considering the cloud data are dynamic in nature, the proposed design further supports secure and efficient dynamic operations on outsourced data, including block modification, deletion, and append. Analysis shows the proposed scheme is highly efficient and resilient against Byzantine failure, malicious data modification attack, and even server colluding attacks.

D. NCCloud: A Network-Coding-Based Storage System in a Cloud-of-Clouds

To provide fault tolerance for cloud storage, recent studies propose to stripe data across multiple cloud vendors. However, if a cloud suffers from a permanent failure and loses all its data, we need to repair the lost data with the help of the other surviving clouds to preserve data redundancy. We present a proxy-based storage system for fault-tolerant multiple-cloud storage called NCCloud, which achieves cost-effective repair for a permanent single-cloud failure. NCCloud is built on top of a network-coding-based storage scheme called the functional minimum-storage regenerating (FMSR) codes, which maintain the same fault tolerance and data Redundancy as in traditional erasure codes (e.g., RAID-6), but use less repair traffic and, hence, incur less monetary cost due to data transfer. We implement a proof-of-concept prototype of NCCloud and deploy it atop both local and commercial clouds.

E. From Single to Multi-Clouds Computing Privacy and Fault Tolerance

Security issues of data hosted in a Cloud Computing provider remain hidden seen excessive marketing that led to a totally unrealistic view of cloud computing security. Although Cloud Computing has not yet reached the level of maturity expected by its customers, and that the problems of confidentiality, integrity, reliability and consistency (CIRC) are still open, the researchers in this field have already considered a future cloud strategy which aims : a better QoS, reliability and high availability, it is the Multi-Clouds, Cloud of Clouds or Inter clouds. This paper will present the security limitations in the single Cloud and the usefulness of adopting rather Multi-Clouds strategy to reduce security risks, through the use of DepSky which is a virtual storage system that ensures better availability and high confidentiality of data.

III. PROBLEM DEFINITION

For outsourced data services, the owners' exclusive control over their data is ultimately relinquished to the CSPs. Although end-to-end encryption techniques have been proposed as promising solutions for secure cloud data storage, the sensitive data can no longer be considered as private. A primary challenge toward building a full-fledged cloud data service remains: We focus on how to enable privacy-assured search for cloud data services.

A. Privacy-Assured Searchable Cloud Storage Architecture

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 pay-per-user manner. There are one or more *data users* in the system, which may need to perform queries over the outsourced data in order to extract useful information. The owner's data are encrypted end-to-end using secret keys created by him/her, 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.

Within the scope of this article, we focus on how to enable privacy-assured 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 Drop box and wants to search

© 2014, IJARCSSE All Rights Reserved

for a particular photo afterward. We consider the cloud server to be semi-trusted. This means that most of the time it behaves properly and does not deviate from the protocol, but it will try to find out as much private information in the stored data as possible. This assumption considers data exposure threats from both insiders and outsiders, and is in line with the current technology trend and business model. For insider threats, there may be curious Data owner Data users Key or trapdoor Cloud service provider Encrypted query Search/match result Infrastructure Services Encrypted data employees inside the CSP who access user data for their own benefit.

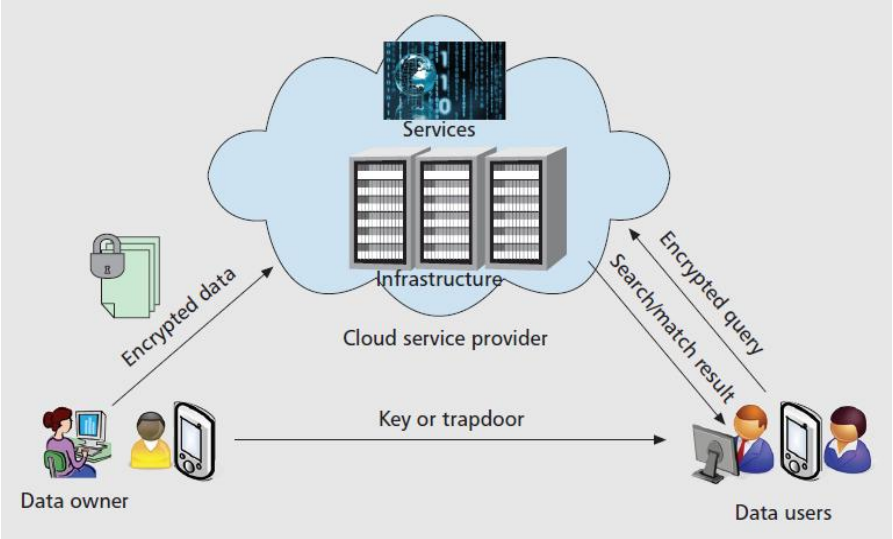

Fig.1 System architecture for searchable cloud data storage services

Thus, the CSP and data owners are not assumed to be in the same trust domain. In addition, some users may also try to access/utilize the data beyond their privileges, either individually or in collusion with each other.[1]

B. System Requirements

Next, we sketch a set of desirable system design requirements from both the functionality and privacy aspects.

1) 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.

2) Efficiency:

A privacy-assured data search scheme should have low computation, communication, and storage overheads. For such a scheme to be deployed in a large-scale 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).

C. Methodology and Building Blocks

Next, we introduce the methodology and briefly describe the main building blocks toward designing usable and efficient privacy-assured search schemes.

Methodology — we describe a top-down approach in Fig 2. Given 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. Then we can try to find a proper encryption scheme to encrypt the data while simultaneously allowing data operations. The second step is nontrivial. Although efficiency could be improved due to the adopted index structure, care

© 2014, IJARCSSE All Rights Reserved

needs to be taken to prevent leakage of private information to the cloud server, especially when the server possesses background information on the query statistics. As a result, sometimes the encryption primitive itself may need to be adapted to meet privacy requirements.

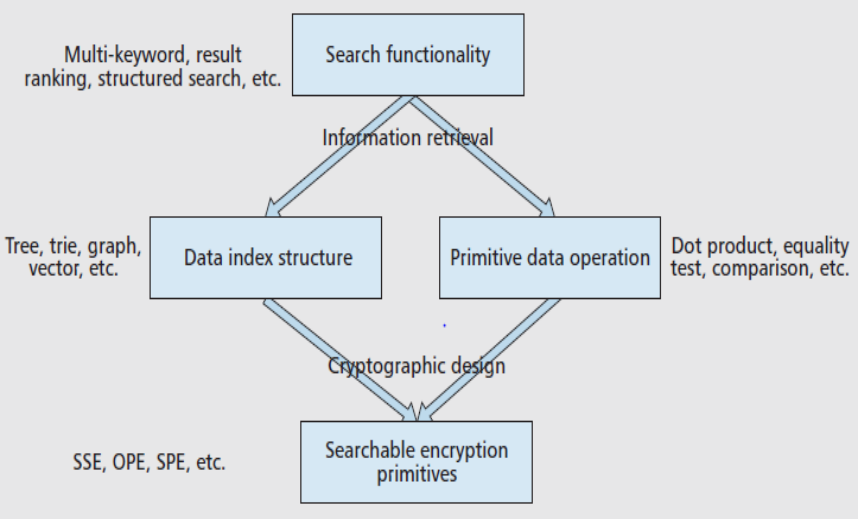

Fig. 2 Top-down methodology for designing privacy-assured search schemes[4]

1) Symmetric Searchable Encryption

Curtmola *et al.* proposed SSE, which is a deterministic symmetric key encryption scheme with security guarantees under rigorous definitions. The SSE scheme is based on the inverted index and uses pseudorandom functions/permutations, which makes the search quite efficient. Roughly speaking, the index consists of blinded keywords *fk* (*wi*) and lists of FIDs containing *wi*, where *f* () is a pseudorandom function and *k* is the secret key. The search trapdoor is also in the same form so that the server can perform matching. However, it only supports single-keyword exact query.

2) Order-Preserving Symmetric Encryption

In OPSE, the numerical ordering of plaintext is preserved after encryption. Boldyreva *et al.* 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*}.

D. Achieving Secure Ranked Search over Encrypted Data

An especially important functionality in plaintext IR is to support ranking mechanisms over search results according to user-specified relevance criteria. Usually, this is achieved by building an inverted index (index structure) and adopting a ranking function to compute the rank of each file relevant to a given search request (primitive data operations are keyword matching and sorting).Toward achieving secure ranked search in the encrypted domain, Zerr *et al.* observed that although the posting list elements (document IDs that contain each keyword in an inverted index) are encrypted, the term frequency distribution for each posting list can still help adversaries re-identify the keyword. Thus, they proposed to transform the relevance scores such that their distribution is uniform for each keyword. They show that this scheme satisfies a definition called *r*-confidentiality in a statistical sense. However, it requires much preprocessing and does not easily handle score dynamics, while the security level is weak.

To improve both the efficiency and privacy, Wang *et al.* proposed a ranked symmetric searchable encryption (RSSE) scheme that enables result ranking for single keyword query. To ensure privacy, a straightforward yet ideally secure RSSE scheme can be derived based on the existing SSE solution, but requires two rounds of interactions between the user and the cloud server, which incurs high communication overhead. To support secure multi-keyword ranked search over encrypted data (MRSE), Cao *et al.* proposed to adopt another similarity measure from the IR community, *coordinate matching*, which captures the relevance of documents to a query through the number of query keywords appearing in a document. Each document index and the query are described as a binary vector (index structure), respectively, such that the similarity is measured by the dot product of the two vectors (primitive data operation).

In the MRSE scheme, the data and index privacy are achieved since the encryption algorithm is secure in the KC model. In addition, under the KB model *query confidentiality* is achieved as well as trapdoors unlink ability. It introduces nearly constant search overhead with the increase of keywords; in contrast, in other multiple-keyword search schemes this is linear.

IV. SIGNIFICANCE

1. To maximize the use of computing power.
2. To give Accessible, Reliable and Virtual computing Technology to customers on-demand.
3. Allows encrypted data to be searched by users without leaking information about the data itself and users' queries.

© 2014, IJARCSSE All Rights Reserved

4. Improve search experience of the data search service.
5. supports secure and efficient dynamic operations on outsourced data.

V. CONCLUSION

In this paper, we identify the problem and challenges of enabling privacy-assured searchable cloud data storage services. Recent research advances in this field are surveyed, which suggest that achieving functionally rich, usable, and efficient search on encrypted data is possible without sacrificing privacy guarantee too much. The steady evolution of this field will need to bring expertise from the cryptography, database, and information retrieval communities. Future Challenges, There are many interesting research issues worth further investigation. The works mentioned above have a common characteristic: they relax the privacy guarantees (i.e., "as strong as possible") to achieve higher efficiency performance. While there are formal privacy definitions for searchable encryption that reveal the access pattern, for as-strong-as possible schemes, how to formally analyse the privacy level given various known background information remains an interesting and important open problem.

ACKNOWLEDGMENT

We thank immensely our management for extending their support in providing us infrastructure and allowing us to utilize them in the successful completion of our paper.

REFERENCES

[1] Ning Cao, Cong Wang, Ming Li, Kui Ren, and Wenjing Lou," Privacy-Preserving Multi-Keyword Ranked Search over Encrypted Cloud Data", IEEE Transactions On Parallel And Distributed Systems, Vol. 25, NO. 1, January 2014, pp 222-232.

[2] Cong Wang, Qian Wang,Kui Ren,Ning Cao,and Wenjing Lou," Towards Secure and Dependable Storage Services in Cloud Computing", 17th IEEE International Workshop on Quality of Service (IWQoS'09), 2013.

[3] Henry C.H. Chen, Yuchong Hu, Patrick P.C. Lee, and Yang "NCCloud: A Network-Coding-Based Storage System in a Cloud-of-Clouds", IEEE TRANSACTIONS ON COMPUTERS, VOL. 63, NO. 1, JANUARY 2014.

[4] Ming Li, Wenjing Lou and Y. Thomas Hou: "Toward Privacy-Assured and Searchable Cloud Data Storage Services", IEEE Network, Vol.13, July/August 2013, pp 56-62.

[5] Maha TEBAA, Said EL HAJJI "From Single to Multi-Cloud Computing Privacy and Fault Tolerance", IERI Procedia 10 (2014), pp 112 – 118.

© 2014, IJARCSSE All Rights Reserved