Computer Science Review ▮ (▮▮▮▮) ▮▮▮–▮▮▮

Contents lists available at ScienceDirect

Computer Science Review

journal homepage: www.elsevier.com/locate/cosrev

A survey on design and implementation of protected searchable data in the cloud

Rafael Dowsley [a], Antonis Michalas [b,*], Matthias Nagel [c,*], Nicolae Paladi [d,*]

[a] Cryptography and Security Research Group, Department of Computer Science, Aarhus University, Aarhus, Denmark
[b] Cyber Security Group, Department of Computer Science, University of Westminster, London, UK
[c] Institute of Theoretical Informatics, Karlsruhe Institute of Technology, Karlsruhe, Germany
[d] Security Lab, SICS Swedish ICT, Kista, Sweden

ARTICLE INFO

Article history:
Received 29 November 2016
Received in revised form 9 August 2017
Accepted 14 August 2017
Available online xxxx

Keywords:
Searchable encryption
Security
Cloud computing
Cloud storage

ABSTRACT

While cloud computing has exploded in popularity in recent years thanks to the potential efficiency and cost savings of outsourcing the storage and management of data and applications, a number of vulnerabilities that led to multiple attacks have deterred many potential users.

As a result, experts in the field argued that new mechanisms are needed in order to create trusted and secure cloud services. Such mechanisms would eradicate the suspicion of users towards cloud computing by providing the necessary security guarantees. Searchable Encryption is among the most promising solutions—one that has the potential to help offer truly secure and privacy-preserving cloud services. We start this paper by surveying the most important searchable encryption schemes and their relevance to cloud computing. In light of this analysis we demonstrate the inefficiencies of the existing schemes and expand our analysis by discussing certain confidentiality and privacy issues. Further, we examine how to integrate such a scheme with a popular cloud platform. Finally, we have chosen – based on the findings of our analysis – an existing scheme and implemented it to review its practical maturity for deployment in real systems. The survey of the field, together with the analysis and with the extensive experimental results provides a comprehensive review of the theoretical and practical aspects of searchable encryption.

Contents

* Corresponding authors.
E-mail addresses: rafael@cs.au.dk (R. Dowsley), a.michalas@westminster.ac.uk (A. Michalas), matthias.nagel@kit.edu (M. Nagel), nicolae@sics.se (N. Paladi).

http://dx.doi.org/10.1016/j.cosrev.2017.08.001
1574-0137/

1. Introduction

In recent years we have witnessed an astonishing increase in the offer of cloud computing solutions. Leveraging savings through large scale optimizations and reduction of wasted resources (inactive computer time, unused hardware space, etc.), this business model offers clear economic advantages. Along with the continuous increase in the amount of data, this provides a strong incentive for both organizations and private users to opt for storing their data with cloud service providers (CSPs). However, this trend raises a security issue, since many clients want to keep their files confidential. The solution may be to encrypt the files before sending them to the CSP, but there are two seemingly contradictory goals that an encryption scheme should achieve in order to be useful in this scenario. On one hand, the encryption must satisfy a strong notion of security in order to keep the data hidden from the CSP. On the other hand, the scheme should allow the clients to continue performing their operations efficiently, i.e. with time and computational costs comparable to the ones for locally stored files. Searching often a quintessential requirement for many clients. It is therefore essential to develop and employ encryption schemes that allow for efficient searching of the data stored in the cloud; if the clients have to download the entire data set and perform the search locally, then the scheme is completely impractical.

Searchable Encryption (SE) is an enhanced encryption technique that allows encryption while enabling search for keywords in the encrypted data (as it would be possible in the plaintext). Its quintessential application is cloud storage. Using searchable encryption should enable a CSP – with the help of a search token sent by the client – to locally perform some operations and then send the relevant data to the client. The relevant data should on one hand contain the matching documents (i.e. the documents that contain the searched keyword), while on the other hand its size should be not much larger than that of the matching documents (i.e. the server cannot simply transfer a large part of the database to the client on every query). Of course the CSP should not learn the keyword that is being searched or the search query, otherwise he is learning partial information about the documents.

Searchable encryption clearly displays trade-offs between efficiency, functionality and security. From an efficiency point of view, it is desirable to reduce as much as possible the number of operations performed by the server during a search. It is also highly important to make these operations parallelizable and increase their locality (in order to improve I/O performance), in order to improve the search time. From the functionality point of view, one important parameter is the query expressiveness. An SE scheme should support as powerful queries as possible, thus increasing the usefulness of the scheme to the clients. Other important parameters are whether a single or multiple clients should be able to write data to the cloud and whether a single or multiple clients should be able to read the data. Additionally, schemes for practical applications should be dynamic, i.e. they should allow database updates without additional leakage. From a security point of view, it is essential to reduce the leakage caused by all operations as much as possible.

Depending on the requirements of the desired scheme, it possible to use either public-key cryptography or symmetric-key cryptography. However, often searchable public-key encryption schemes with good security guarantees do not scale well because they have search time which is linear in the number of documents.

Symmetric searchable encryption was introduced by Song et al. [1], who presented a scheme that allowed linear search time (in the number of documents) by the server. Unfortunately their scheme does not achieve a strong notion of security: it has no security guarantees related to the leakage that can be caused by the use of the search tokens that are given to the server in order to allow the search to be performed on the server side. Goh [2] introduced the approach of using secure indexes in order to achieve linear search time with stronger security guarantees. Unfortunately the search time of this approach is inherently linear in the number of files. Curtmola et al. [3] presented the first secure scheme with sub-linear search time using an inverted index approach (uses the keywords as index) and also introduced a strong security model for searchable encryption which became the standard security notion for searchable encryption in the last several years. The inverted index approach is quite efficient and is in fact optimal for the number of operations that the server has to perform during a search, which led to it being used in many subsequent works (e.g., [4–6]). One limitation of this method is that it is inherently sequential, preventing it from taking advantage of parallelism to improve performance. Another issue is that it is not well-suited for dynamic databases, which is the case of most applications. Recent works made progress in the direction of dynamic [5,7–10] and parallel [7,8,11] schemes.

Symmetric searchable encryption perfectly fits the scenario of a single user writing to/reading from the database. However, there is a generic construction that combines a single writer/reader scheme with broadcast encryption in order to obtain a scheme that supports multiple readers [3]. One additional issue in this case is revocation: a revoked user should not be able to perform searches after the revocation has occurred.

In terms of query expressiveness, most symmetric searchable encryption schemes focus on single equality queries. Some recent works [11,12] demonstrated that it is possible to extend data structures for single keyword symmetric searchable encryption in order to deal with more complex queries, such as conjunctive queries for keyword combinations and general Boolean queries.

Public-key searchable encryption was introduced by Boneh et al. [13]. It allows multiple clients to encrypt data into the database, which can be decrypted by the data owner that has the secret-key. Other solutions allow conjunctive, subset and range queries [14]. The efficiency of these schemes is limited by the cost of public-key operations. Another problem of the proposed

Please cite this article in press as: R. Dowsley, et al., A survey on design and implementation of protected searchable data in the cloud, Computer Science Review (2017), http://dx.doi.org/10.1016/j.cosrev.2017.08.001.