Journal of Network and Computer Applications 64 (2016) 12–22

Contents lists available at ScienceDirect

Journal of Network and Computer Applications

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

An efficient and secure privacy-preserving approach for outsourced data of resource constrained mobile devices in cloud computing

Syam Kumar Pasupuleti [a,*], Subramanian Ramalingam [b], Rajkumar Buyya [c]

[a] Institute for Development and Research in Banking Technology (IDRBT), Hyderabad, India
[b] Department of Computer Science, Pondicherry University, Puducherry, India
[c] Cloud Computing and Distributed Systems (CLOUDS) Lab, Department of Computing and Information Systems, The University of Melbourne, Australia

ARTICLE INFO

Article history:
Received 1 January 2015
Received in revised form
16 October 2015
Accepted 8 November 2015
Available online 9 February 2016

Keywords:
Cloud computing
Privacy-preserving
Outsourced data
Probabilistic public-key encryption
Ranked keyword search
Mobile devices

ABSTRACT

Outsourcing of data into cloud has become an effective trend in modern day computing due to its ability to provide low-cost, pay-as-you-go IT services. Although cloud based services offer many advantages, privacy of the outsourced data is a big concern. To mitigate this concern, it is desirable to outsource sensitive data in an encrypted form but cost of encryption process would increase the heavy computational overhead on thin clients such as resource-constrained mobile devices. Recently, several keyword searchable encryption schemes have been described in the literature. However, these schemes are not effective for resource-constrained mobile devices, because the adopted encryption system should not only support keyword search over the encrypted data but also offer high performance. In this paper, we propose an efficient and secure privacy-preserving approach for outsourced data of resource-constrained mobile devices in the cloud computing. Our approach employs probabilistic public key encryption algorithm for encrypting the data and invoke ranked keyword search over the encrypted data to retrieve the files from the cloud. We aim to achieve an efficient system for data encryption without sacrificing the privacy of data. Further, our ranked keyword search greatly improves the system usability by enabling ranking based on relevance score for search result, sends top most relevant files instead of sending all files back, and ensures the file retrieval accuracy. As a result, data privacy ensures and computation, communication overheads in reduction. Thorough security and performance analysis, we prove that our approach is semantically secure and efficient.

1. Introduction

Cloud computing is emerging computing model where the data owners are outsourcing their data into the cloud storage. By outsourcing the data files into the cloud, it gives many benefits to the large enterprises as well as individual users because they can dynamically increase their storage space as and when required without buying any storage devices (Armbrust et al., 2009). They are: (1) the users can access the remotely stored data at anytime, from anywhere and gives permission to authorized users to share the data. (2) The users can be relieved from the burden of storage management at locally, (3) Avoidance of capital expenditure on hardware and software costs etc. To date, there are a number of cloud storage services: Amazon simple storage Space (S3), Rack space, Google, Microsoft, etc. (Jaeger et al. 2009).

Besides, all of these advantages of outsourced data in Cloud, there are also some significant issues. One of the major issues is the privacy of outsourced data in cloud (Jaeger and Schiffman, 2010) i.e., the sensitive information such as e-mail, health records, and government data may leak to unauthorized users (Slocum, 2009; Krebs, 2009) or even be hacked (Cloud Security Alliance, 2009). Since, the cloud is an open platform; it can be subjected to attacks from both malicious insiders and outsiders (Haclgiimfi et al., 2002). The Cloud service providers (CSPs) usually provide data security through mechanisms like firewalls and virtualization. However, these mechanisms do not protect users' privacy from the CSP itself due to remote cloud storage servers are untrusted.

A natural approach to preserve the privacy of sensitive data is to encrypt data before outsourcing it into the cloud and retrieves the data back through keyword based search over encrypted data. Although encryption provides protection from illegal accesses, it significantly increases the computation overhead on the data owners especially when they having resource-constrained mobile devices and large size of data files.

Further, the authorized users want to retrieve the certain files from cloud, need to communicate with the CSPs and allow him to

* Corresponding author.
E-mail addresses: psyamkumar@idrbt.ac.in (S.K. Pasupuleti), rsmanian.csc@pondiuni.edu.in (S. Ramalingam), rbuyya@unimelb.edu.au (R. Buyya).

http://dx.doi.org/10.1016/j.jnca.2015.11.023
1084-8045/

operate over the encrypted data. To meet effective data retrieval, it is preferred to get the most relevant files instead of getting all files i.e., The files should be ranked and only highest relevant files are send back to the users, which is highly desirable in the "pay-as-you-use" cloud model. However, it is challenging task that retrieves the data back in a secure and efficient manner without being able to extract useful information from the cloud.

Therefore, the efficient and secure mechanisms are needed to protect the privacy of sensitive data in a cloud environment. Moreover, the importance and necessity of privacy preserving of data search techniques are even more pronouncing in the cloud applications. For example, large companies that are operating on the public clouds like Google or Amazon may access the sensitive data, search patterns, hiding the query and retrieved data has great importance in ensuring the privacy of that using cloud services.

Recently, several keyword search based encryption schemes have been proposed to ensure the privacy of outsourced data (Song et al., 2000; Goh, 2003; Chang and Mitzenmacher, 2005; Curtmola et al., 2006; Li et al., 2010; Kuzu et al., 2012; Lu, 2012; Orencik and Savas, 2014; Wang et al., 2012; Cao et al., 2014; Boneh et al., 2004; Bellare et al., 2007; Attrapadung and Li Bert, 2010; Katz et al., 2008; Ogata and Kurosawa, 2004; Shi et al., 2007; Waters et al., 2004; Bao et al., 2008; Boldyreva et al., 2009; Liu et al., 2012; Yu et al., 2013). In all these schemes, the data owner first encrypts the data before outsourcing it and later retrieves them through keyword search or ranked keyword search. The schemes (Song et al., 2000; Goh, 2003; Chang and Mitzenmacher, 2005; Curtmola et al., 2006; Li et al., 2010; Kuzu et al., 2012; Lu, 2012; Orencik and Savas, 2014; Wang et al., 2012; Cao et al., 2014) and (Boneh et al., 2004; Bellare et al., 2007; Attrapadung and Li Bert, 2010; Katz et al., 2008; Ogata and Kurosawa, 2004; Shi et al., 2007; Waters et al., 2004; Bao et al., 2008; Boldyreva et al., 2009) proposed under symmetric-key cryptography and public-key cryptography respectively. However, such encryption schemes use too much CPU time and memory power of client during the encryption and decryption process. That is the thin client has only limited bandwidth, CPU power, and memory, therefore, the traditional encryption schemes cannot work well in cloud environment.

To avoid above problems, Liu et al. (2012) proposed a secure and privacy-preserving keyword search over the encrypted data for cloud storage applications using Elliptic Curve Cryptography (ECC) over Fp. However, this scheme supports only Boolean keyword search i.e., either a keyword exists in a file or not, without considering the difference of relevance with the queried keyword of these files in the result. To improve the efficiency without sacrificing privacy, Yu et al. (2013) proposed a Two Round Searchable Encryption (TRSE) scheme that supports ranked multi-keyword search over encrypted data for file retrieval. It employs the vector space model and homomorphic encryption as a result, the information leakage can be eliminated and data security is ensured. However, the computation and communication costs of this scheme are quite large, since every search term in a query requires several homomorphic encryption operations on the data owner side. Further, it uses two-round communication process to retrieve the files back which resulting the unnecessary communication overhead.

In this paper, we propose an efficient and secure privacy-preserving approach to avoid all above problems while preserving the privacy and integrity of outsourced data in the cloud. In our scheme, the data owner first builds the index for file collection, encrypts both index and data files, and stores them in the cloud. Later, to retrieve the stored files from the cloud server, the authorized user generate trapdoor for keywords and sends to the server. Upon receiving the trapdoor, the cloud server search for a list of matched file entries and their corresponding encrypted relevance scores. Then matched files should be sent back to the user in a ranked sequence based on the relevance scores. By decrypting it, the user gets the original files back. Further, our approach verifies the integrity of data in cloud. This approach utilizes the probabilistic public key encryption technique (Witten et al., 1999) and ranked keyword search (Cao et al., 2014; Yu et al., 2013). It greatly reduces the processing overhead of data owners while encrypting the files, index and it is most suitable for resource-constrained mobile devices (thin clients) in Cloud computing and ranked keyword search process reduce the communication overhead during the file retrieval. Through the security and performance analysis, we prove that our scheme is semantically secure and efficient.

The **key contributions** of our work can be summarized as follows:

1. We propose an efficient and secure privacy-preserving approach; it uses probabilistic public key encryption technique to reduce computational overhead on owners while encryption and decryption process without leaking any information about the plaintext.
2. Our approach uses ranked keyword search on encrypted data to retrieve the files back. It enables the cloud server to determine whether a given file contains certain keywords and associated relevance score without knowing of any information about both the keywords and the files. It greatly reduces the communication overhead during the file retrieval process. It also verifies the integrity of data stored in cloud
3. Through analysis on security demonstrates that propose scheme can be proved semantically secure under different attacks. Furthermore, the performance analysis and experiential results show that our scheme is efficient and it outperforms compared with existing schemes.

The rest of the paper is organized as follows: In Section 2, we briefly discuss about existing schemes. In Section 3, we explain system architecture used in our scheme. In Section 4, we describe our proposed approach. In Sections 5 and 6, we analyze security, and performance analysis respectively. In Section 7, we identify limitations and areas of improvements. Finally, we conclude our paper with future directions in Section 8.

2. Related work

The searchable encryption schemes have been widely investigated as a cryptographic primitive with a focus on security definition formalizations and efficiency improvements (Song et al., 2000; Goh, 2003; Chang and Mitzenmacher, 2005; Curtmola et al., 2006; Li et al., 2010; Kuzu et al., 2012; Lu, 2012; Orencik and Savas, 2014; Wang et al., 2012; Cao et al., 2014; Boneh et al., 2004; Bellare et al., 2007; Attrapadung and Li Bert, 2010; Katz et al., 2008; Ogata and Kurosawa, 2004; Shi et al., 2007; Waters et al., 2004; Bao et al., 2008; Boldyreva et al., 2009; Liu et al., 2012; Yu et al., 2013). These searchable encryption schemes can be divided into two types: symmetric key encryption (Song et al., 2000; Goh, 2003; Chang and Mitzenmacher, 2005; Curtmola et al., 2006; Li et al., 2010; Kuzu et al., 2012; Lu, 2012; Orencik and Savas, 2014; Wang et al., 2012; Cao et al., 2014) and public key encryption (Boneh et al., 2004; Bellare et al., 2007; Attrapadung and Li Bert, 2010; Katz et al., 2008; Ogata and Kurosawa, 2004; Shi et al., 2007; Waters et al., 2004; Bao et al., 2008; Boldyreva et al., 2009).