Efficient Conjunctive Keyword Search on Encrypted Data Storage System*

Jin Wook Byun, Dong Hoon Lee, and Jongin Lim

Center for Information Security Technologies (CIST),
Korea University, Anam Dong, Sungbuk Gu, Seoul, Korea
{byunstar, donghlee, jilim}@korea.ac.kr

Abstract. We study *conjunctive* keyword search scheme allowing for remote search of data containing each of several keywords on encrypted data storage system. A data supplier first uploads encrypted data on a storage system, and then a user of the storage system searches data containing keywords over encrypted data hence insider (such as an administrator of the storage system) and outsider attackers do not learn anything else about the data. Recently, Golle *et al.* first suggested conjunctive keyword search scheme, but the communication and storage costs linearly depend on the number of stored data in the database, hence it is not really suitable for a large scale database.

In this paper, we propose an efficient conjunctive keyword search scheme over encrypted data in aspects of communication and storage costs. Concretely, we reduce the storage cost of a user and the communication cost between a user and a data supplier to the constant amounts. We formally define security model for a conjunctive keyword search scheme and prove that the proposed scheme is secure under the decisional bilinear Diffie-Hellman (DBDH) assumption in the random oracle model.

Keywords: Conjunctive keyword search over encrypted data, database security and privacy.

1 Introduction

As the amount of information to be stored and managed on the Internet rapidly increases, protecting data in a database from outsider/insider attackers has been hot issues in a secure database management system. The most simple solution to prevent theft and misuse of data from outsider/insider attackers is that a user of storage system simply encrypts personal data with his own private key, and stores the encrypted results on the storage system. The user should also manage his encryption key securely without revealing it to the outsider/insider attackers. However, secure encryption makes data look random, and unreadable to anyone

* This research was supported by the MIC(Ministry of Information and Communication), Korea, under the ITRC(Information Technology Research Center) support program supervised by the IITA(Institute of Information Technology Assessment).

A.S. Atzeni and A. Lioy (Eds.): EuroPKI 2006, LNCS 4043, pp. 184–196, 2006.

other than the users holding the encryption keys, hence the server is unable to determine which encrypted data contain specific keywords. And then, how can original documents be efficiently searched including the user specific keywords over the encrypted documents? Intuitively, one may think a trivial search process that the user downloads all documents and decrypt them with his secret key, and searches documents containing specific keywords on the user's machine. As one can easily see, this process is very inefficient and would impose massive burdens on the user side as stored documents rapidly increase. To resolve this problem, there has been much research on efficient and secure keyword search over the encrypted documents based on the various scenarios [1, 3, 4, 5, 7, 8, 10, 11, 12, 13].

1.1 Related Works and Our Contributions

In this paper, we consider a *conjunctive keyword search* [7, 12] which finds data containing each of several keywords by asking one query. One may argue that a conjunctive keyword search scheme can be built from the multiple executions of any single keyword search scheme. In this case, however, the server should find all data containing each keyword by using the single keyword search, check the intersection set of all data, then return the results to the user. This approach requires high computation cost and redundancy to the server due to duplicated comparisons and search.

A conjunctive keyword search scheme over encrypted data consists of three entities: a data supplier, a storage system such as a database, and a user of storage system. A data supplier uploads encrypted data on a storage system, and then a user of the storage system searches data containing keywords. Let's suppose an untrusted web-based personal storage (PS) system in which a user of personal storage system oneself may store encrypted data over the server and search data containing appropriate keywords on the encrypted data. Many schemes [13, 6, 8, 7] have been suggested in this setting by using only symmetric cryptography such as block, stream cipher, and Bloom filter. Song *et al.* suggested an efficient and provably secure keyword search scheme by using stream and block cipher [13]. In [8], Goh suggested a secure search scheme using a Bloom filter [2]. Very recently, Chang and Mitzenmacher also suggested a more practical keyword search protocol in terms of communication and storage overheads. However, these schemes are not appropriate for fully conjunctive keyword search. As pointed out in [6, 8], the design of conjunctive keyword search scheme using only symmetric cryptography still remains as a challenging open problem. Recently, to provide a conjunctive keyword search in this setting, Golle *et al.* applied public key cryptography to the keyword search scheme, and first proposed two secure conjunctive keyword search protocols over encrypted data.

However, the one of schemes is very inefficient in aspect of communication and storage costs, as analyzed in Table 1. That is, the costs linearly depend on the number of stored data, and the scheme requires huge communication and computation costs in case of a large scale database. For example, MS SQL Server™ 2005 Edition has at most 1,048,516 TBytes size [9], and if we suppose that one record requires about 10 MBytes then the server has at most 104,851,600,000