Practical Multi keyword Ranked Search with Access Control over Encrypted Cloud Data

ABSTRACT

With the explosive growth of data volume in the cloud computing environment, data owners are increasingly inclined to store their data on the cloud. Although data outsourcing reduces computation and storage costs for them, it inevitably brings new security and privacy concerns, as the data owners lose direct control of sensitive data. Meanwhile, most of the existing ranked keyword search schemes mainly focus on enriching search efficiency or functionality, but lack of providing efficient access control and formal security analysis simultaneously. To address these limitations, in this paper we propose an efficient and privacy-preserving **M**ulti-keyword **R**anked **S**earch scheme with **F**ine-grained access control (MRSF). MRSF can realize highly accurate ciphertext retrieval by combining coordinate matching with Term Frequency-Inverse Document Frequency (TF-IDF) and improving the secure kNN method. Besides, it can effectively refine users’ search privileges by utilizing the polynomial-based access strategy. Formal security analysis shows that MRSF is secure in terms of confidentiality of outsourced data and the privacy of index and tokens. Extensive experiments further show that, compared with existing schemes, MRSF achieves higher search accuracy and more functionalities efficiently.

**EXISTING SYSTEM**

The pioneering work proposed by Boneh et al. in [41] is the first public-key encryption scheme that supports single keyword search. This work is extended in [15], [42], [43], supporting more operations over encrypted data such as conjunctive keyword search, range query, etc. However, the ASE schemes are less efficient than SSE schemes due to the complex encryption procedures. In [39], Yu et al. proposed a two-round searchable encryption (TRSE) scheme that supports ranked multi-keyword search. In TRSE, homomorphic encryption is leveraged to encrypt index and query generated by a vector space model. Although TRSE guar antees high security, it takes two rounds of communications between the data user and the cloud server to complete one search process. In the work of Cheng et al., a public-key cryptosystem based kNN scheme is proposed [44]. Different from the former secure kNN methods based on symmetric encryption, the proposed scheme leverages the distributed two trapdoors public-key cryptosystem (DT-PKC), which enables secure k-NN query with multiple keys.

The concept of SSE is first proposed by Song et al. in [4], but this scheme lacks support for keywords relevance calculation and multi-keyword search. Another SSE based ranked search over encrypted data is proposed by Wang et al. in [45], leveraging OPSE (order-preserving symmetric encryption). In order to conduct kNN search over encrypted dataset, Wong et al. first proposed the asymmetric scalar product- preserving encryption (ASPE) in [24], which is viewed as the original secure kNN scheme. Since then, ASPE has been thoroughly studied in many works [16], [38], [40]. However, most ranked keyword search schemes based on ASPE are vulnerable to level-3 attack [24], where the adversary is able to gain a certain amount of plain text ciphertext pairs. In a word, secure kNN computation is anSE method with high usability but relatively low security.

Functionality extension. Proposed schemes in literature support at least one search functionality. As mentioned before, [10]-[15] focus on boolean keyword search, while [1], [16], [38], [40] focus on multi-keyword search. Schemes that focus on geometric search include [22], [46], [47]. Relevance scores in keyword/textual search are replaced by the distance between coordinate points, and the data structure is

often designed specially in those schemes. Other schemes support mixed search objects, for example, [48] proposed a scheme that returns top-k location points with keywords matching the queried keywords. In this paper, we mainly focus on keyword search over encrypted data. There are variant functionality extensions towards keyword search schemes over encrypted data.

Towards semantic-aware keyword search over encrypted data, many works provide a solution. Guan et al. [49] proposed a multi-keyword ranked search scheme with a semantic extension (CLRSE) that can be applied to a cross-lingual dataset. The data user is allowed to set the language preference before launching the search. CLRSE

adopts a two-cloud system model and Paillier cryptosystem to achieve a higher security level. To better extract search intents of the data user, Dai et al. [50] proposed a scheme that enables semantic-aware keyword search over encrypted data. They adopt the secure kNN method as the encryption algorithm. This scheme utilizes a natural language processing model to extract features from both the document

and search query, such that the document-query similarity is transformed into feature vectors similarity. Lang et al. [51] adopted the same natural language processing model in their semantic-based compound secure keyword search scheme, they also utilize Locality-Sensitive Hashing (LSH) to eliminate unnecessary privacy leakage.

Disadvantages

1) The system was not implemented MRSF which leads less security on outsourced data.

2) The system is less security due to lack of Term Frequency-Inverse Document Frequency (TF-IDF) and there is no block verification.

**PROPOSED SYSTEM**

The proposed system designed several extensive experiments to explore the relationship between randomness and search accuracy. Inspired by [27], we give an **optimal security notion** for MRSF as indistinguishability under same-closeness-pattern chosen-plaintext attacks (IND-CLS-CPA), which is a natural relaxation of the standard IND-CPA security definition. In the IND-CPA definition, the attacker will trivially exploit of the leakage of closeness and equality (e.g. access pattern), which is inevitable in MRSF. In this paper, we provide a comprehensive and detailed security analysis to prove that MRSF is IND-CLS-CPA secure. Moreover, the security of MRSF under classic and novel attacks in the known background model.

Specifically, the contributions of this paper can be summarized as follows:

\_ Higher search accuracy. In comparison with the previous ranked multi-keyword search schemes, MRSF achieves higher search accuracy by constructing document indexes with TF-IDF rule. Experiments using real-world dataset demonstrate that MRSF outperforms the compared schemes in search accuracy without introducing excessive computational overheads.

\_ Lightweight Fine-grained access control. Instead of employing the time-consuming encryption and decryption operations in attributed-based schemes, MRSF

provides lightweight access control by extending the polynomial-based access strategy to document indexes and search queries. The access control mechanism

in MRSF is integrated into the improved secure kNN algorithm while it retains the merit of the finegrained access privileges for users.

\_ Higher privacy-preserving level and formal security proof. The optimal security notion of MRSF is defined and theoretically proved as IND-CLS-CPA, i.e., the confidentiality of indexes and search queries are both verified as IND-CLS-CPA secure. Security issues in the Known Background Model are further analyzed to prove that MRSF is resistant to the proposed attacks. In this way, MRSF is eventually proved to be securer than the previous secure kNN based schemes.

**Advantages**

* To avoid privacy leakage from indexes and queries submitted to the cloud, the classic secure kNN technique has been adapted to support MRSF. The changes include the vector extensions, pseudorandom permutation function, and appended random varieties.We designed several extensive experiments to explore the relationship between randomness and search accuracy. Inspired by the system,
* The proposed system gives an **optimal security notion** for MRSF as indistinguishability under same-closeness-pattern chosen-plaintext attacks (IND-CLS-CPA), which is a natural relaxation of the standard IND-CPA security definition. In the IND-CPA definition, the attacker will trivially exploit of the leakage of closeness and equality (e.g. access pattern), which is inevitable in MRSF.

**SYSTEM REQUIREMENTS**

➢ **H/W System Configuration:-**

➢ Processor - Pentium –IV

➢ RAM - 4 GB (min)

➢ Hard Disk - 20 GB

➢ Key Board - Standard Windows Keyboard

➢ Mouse - Two or Three Button Mouse

➢ Monitor - SVGA

**Software Requirements:**

* Operating System - Windows XP
* Coding Language - Java/J2EE(JSP,Servlet)
* Front End - J2EE
* Back End - MySQL