TOP K RESULT RETRIEVAL IN SEARCHING THE FILE OVER THE ENCRYPTED DATA IN DATA TRANSMISSION SYSTEM

Minimol PS[1] | Karthikeyan[2]

[1](UG Student, Christ the King Engineering College, minimolps95@gmail.com)

[2](Assistant Professor, Christ the King Engineering College, karthikeyann195@gmail.com)

***Abstract*— On focusing the network security and solving the classification problem over encrypted data. A secure k-NN classifier is used over encrypted data in the data transmission. The protocol protects the confidentiality of data, privacy of user's input query, and hides the data access patterns. We achieve the goal by constructing secure verification object for encrypted cloud data. Furthermore, a short signature technique with extremely small storage cost is proposed to guarantee the authenticity of verification object and a verification object request technique is presented to allow the query user to securely obtain the desired verification object. This work is the first to develop a secure k-NN classifier over encrypted data under the semi-honest model.*

***Keywords*— Multi key Words, Private Key, Key generation, Verification**

1. INTRODUCTION

A secure k-NN classifier over semantically secure encrypted data. A novel secure and efficient scheme for k-NN query on encrypted cloud data in which the key of data owner to encrypt and decrypt outsourced data will not be completely disclosed to any query user. A novel privacy-preserving k-NN classification protocol over encrypted data in the cloud. To achieve this, first present a two-party protocol for the exponential mechanism. This protocol can be used as a sub protocol by any other algorithm that requires the exponential mechanism in a distributed setting.

Motivation CLOUD computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction. Driven by the abundant benefits brought by the cloud computing such as cost saving, quick deployment, flexible resource configuration, etc., more and more enterprises and individual users are taking into account migrating their private data and native applications to the cloud server. A matter of public concern is how to guarantee the security of data that is outsourced to a remote cloud server and breaks away from the direct control of data owners. Encryption on private data before outsourcing is an effective measure to protect data confidentiality. However, encrypted data make effective data retrieval a very challenging task. To address the challenge (i.e., search on encrypted data), Song et al. first introduced the concept of searchable encryption and proposed a practical technique.

The BDMRS scheme in the known cipher text model, and the EDMRS scheme in the known background model.1) Design a searchable encryption scheme that supports both the accurate multi-keyword ranked search and flexible dynamic operation on document collection.2) The high efficiency is achieved by "Greedy Depth-first Search" algorithm. The parallel search can be flexibly performed to further reduce the time cost of search process.

2. RELATED WORKS

2.1 Secure Search in Cloud Computing Essentially, the secure search is thus a technique that allows an authorized data user to search over the data owner's encrypted data by submitting encrypted query keywords in a privacy-preserving manner and is an effective extension of traditional searchable encryption to adapt for the cloud computing environment. Motivated by the effective information retrieve on encrypted outsourced cloud data, Wang et al. first proposed a keyword-based secure search scheme and later the secure.

2.2 Verifiable Secure Search in Cloud Computing In practical applications, the cloud server may return erroneous or false search results once he behaves dishonestly for illegal profits or due to possible software/hardware failure of the cloud server. Because of the possible data corruption under a dishonest setting, serval research works have been proposed to allow the data user to enforce query results verification in the secure search fields for cloud computing. Wangetal. applied hash chain technique to implement the completeness verification of query results by embedding the encrypted verification information into their proposed secure searchable index. Sunetal used encrypted index tree structure to implement secure query results verification functionality. In this scheme, when the query ends, the cloud server returns query results along with a minimum encrypted index tree,

then the data user searches this minimum index tree using the same search algorithm as the cloud server did to finish result verification.

Zhengal constructed a verifiable secure query scheme over encrypted cloud data based on attribute-based encryption technique(ABE)in the public-key setting. Sunetal referred to the Merkle hash tree and applied Pairing operations to implement the correctness and completeness verification of query results for keyword search over large dynamic encrypted cloud data. However, these secure verification schemes cannot achieve our proposed fine-grained verification goals. Furthermore, these verification mechanisms are generally tightly coupled to corresponding secure query schemes and have not universality.

3 BACKGROUNDS

To clarify our proposed problems, in this section, we present our system model, threat model, and several preliminaries used to implement our scheme.

3.1 System Model The system model of the secure search over encrypted cloud data usually includes three entities: data owners, data users, and the cloud server, which describes the following scenario: data owners encrypt their private data and upload them to would be maliciously deleted or tampered by the dishonest cloud server. When the query results face the risks that are deleted or tampered, a well-functioning secure query system should provide a mechanism that allows the data user to verify the correctness and completeness of query results. To achieve the results verification goal, we propose to construct secure verification objects for data files that are outsourced to the cloud with encrypted data and secure indexes together.

3.2Threat Model

In this paper, compared with the previous works, an important distinction about the threat model is that the cloud is considered to be an untrusted entity. More specifically, first of all, the cloud server tries to gain some valuable information from encrypted data files, secure indexes, and verification objects (e.g., a misbehaving cloud administrator aims at obtaining these information for possible monetary profits). Then, the cloud server would intentionally return false search results for saving computation resource or communication cost. Further, if the cloud server knows a query results verification mechanism is embedded, he may tamper or forge verification objects to escape responsibilities of misbehavior. Similar to the previous works, both data owners and authorized data users are considered to be trusted in our threat model.

3.3 Preliminaries 3.3.1 Bloom Filter A Bloom Filter is a space- efficient probabilistic data structure which is used to test whether an element is a member of a set. An empty Bloom filter is a bit array of m bits, where all bits are set to 0 initially.

Uploading data

The data owner has a collection of n files to outsource onto the cloud server in encrypted form and expects the cloud server to provide keyword retrieval service to data owner himself or other authorized users. To achieve this, the data owner needs to build a searchable index from a collection of keywords extracted out of files, and then outsources both the encrypted index and encrypted files onto the cloud server responding verification object V Ow. The data user only needs two steps to finish the correctness verification for each data file c in Rw. First, the data user uses the shared key k and the hash functions family H to calculate

.Encryption of data

We propose a new searchable encryption scheme, in which novel technologies in cryptography community are employed, including data encryption and the vector space model. In the proposed scheme, the data owner encrypts the searchable index with data encryption.

When the cloud server receives a query consisting of multi keywords, it computes the scores from the encrypted index stored on cloud and then returns the encrypted scores of files to the data user. Next, the data user decrypts the scores and picks out the top- k highest-scoring files' identifiers to request to the cloud server. The retrieval takes a two-round communication between the cloud server and the data user. "Public-key encryption with fuzzy keyword search: A provably secure scheme under keyword guessing attack,"

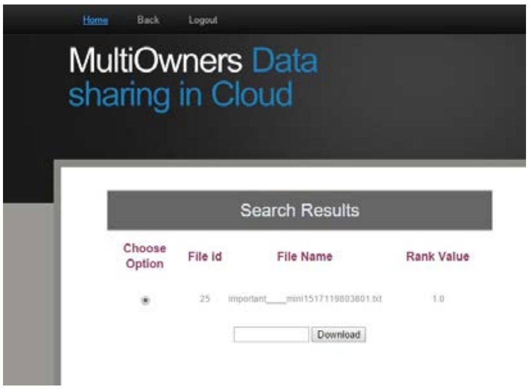
Data Search

The data user is authorized to process multikeyword retrieval over the outsourced data. The computing power on the user side is limited, which means that operations on the user side should be simplified.

The authorized data user at first generates a query. For privacy consideration, which keywords the data user has searched must be concealed. Thus, the data user encrypts the query and sends it to the cloud server that returns the relevant files to the data user. Afterward, the data user can decrypt and make use of the files.

Top k result

Scoring is a natural way to weight the relevance. Based on the relevance score, files can then be ranked in either ascendingly or descendingly. Several models have been proposed to score and rank files in IR community. Among these schemes, we adopt the most widely used one tf-idf weighting, which involves two. The protocol protects the confidentiality of data, privacy of user's input query, and hides the data access patterns. We achieve the goal by constructing secure verification object for encrypted cloud data. Furthermore, a short signature technique with extremely small storage cost is proposed to guarantee the authenticity of verification object and a verification object request technique is presented to allow the query user to securely obtain the desired verification object. This work is the first to develop a secure k-NN classifier over encrypted data under the semi-honest model.



MultiOwners Data sharing in Cloud

Search Results

Choose Option	File Id	File Name	Rank Value
●	25	important____mini1517119803801.txt	1.0

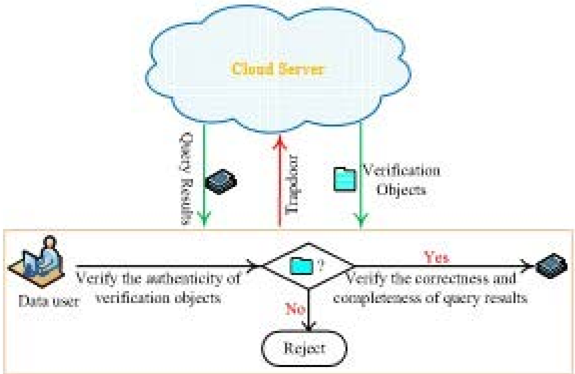
term frequency and inverse document frequency attributes

Both for classification and regression, it can be useful to assign weight to the contributions of the neighbors, so that the nearer neighbors contribute more to the average than the more distant ones. For example, a common weighting scheme consists in giving each neighbor a weight of 1/d, where d is the distance to the neighbor.

The neighbors are taken from a set of objects for which the class (for k-NN classification) or the object property value (for k-NN

Considering the large number of data users and documents in the cloud, it is necessary to allow multi key-word in the search query and return documents in the order of their relevancy with the queried keywords. Scoring is a natural way to weight the relevance. Based on the relevance score, files can then be ranked in either ascendingly or descendingly. Several models have been proposed to score and rank files in IR community. Among these schemes, we adopt the most widely used one tf-idf weighting, which involves two attributes-term frequency and inverse document frequency

4


Cloud Server
Query Results
Trapdoor
Verification Objects
Data user
Verify the authenticity of verification objects
?
Yes
Verify the correctness and completeness of query results
No
Reject

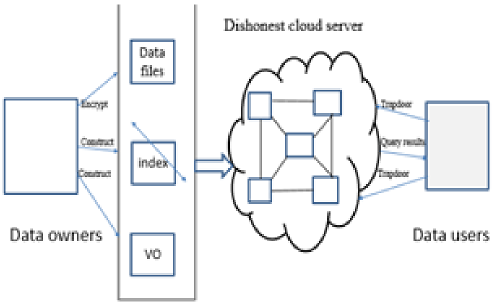
Algorithm Used

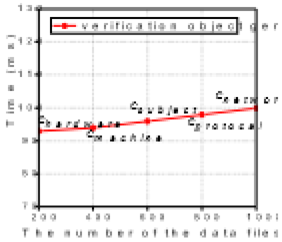
k-nearest neighbors’ algorithm (k-NN)

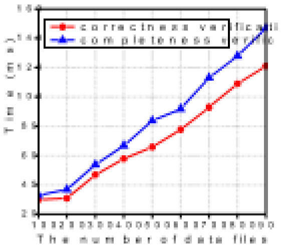
In pattern recognition, the k-nearest neighbors algorithm (k-NN) is a non-parametric method used for classification and regression. In both cases, the input consists of the k closest training examples in the feature space. The output depends on whether k-NN is used for classification or regression: In k-NN classification, the output is a class membership. An object is classified by a majority vote of its neighbors, with the object being assigned to the class most common among its k nearest neighbors (k is a positive integer, typically small). If k = 1, then the object is simply assigned to the class of that single nearest neighbor.
In k-NN regression, the output is the property value for the object. This value is the average of the values of its k nearest neighbors.

regression) is known. This can be thought of as the training set for the algorithm, though no explicit training step is required.
A shortcoming of the k-NN algorithm is that it is sensitive to the local structure of the data.[citation needed] The algorithm is not to be confused with k-means, another popular machine learning technique.

Verifying the Correctness and Completeness of Query Results
When a query ends, the query results set and corresponding verification object are together returned to the query user, who verifies the correctness and completeness of query results based on the verification object. Our proposed query results verification scheme not only allows the query user to easily verify the correctness of each encrypted data file in the query results set, but also enables the data user to efficiently perform completeness verification before decrypting query results. More importantly, for an incomplete query results set, our verification objects can definitely tell the data user that the cloud server has omitted how many qualified data files for a query, which is a significant advantage compared with the previous related works. We give an example to illustrate how to perform query results verifications. Assume that an authorized data user submit an encrypted query keyword w to the cloud server, the cloud server performs query operations and returns back the query results set Rw along with the co.
On focusing the network security and solving the classification problem over encrypted data. A secure k-NN classifier is used over encrypted data in the data transmission. The protocol protects the confidentiality of data, privacy of user’s input query, and hides the data access patterns. We achieve the goal by constructing secure verification object for encrypted cloud data. Furthermore, a short signature technique with extremely small storage cost is proposed to guarantee the authenticity of verification object and a verification object request technique is presented to allow the query user to securely obtain the desired verification object. This work is the first to develop a secure k-NN classifier over encrypted data under the semi-honest model







REFERENCES

[1] C. Gentry, “A fully homomorphic encryption scheme,” Ph.D. dissertation, Stanford University, 2009.

[2] D. Boneh, E. Kushilevitz, R. Ostrovsky, and W. E. Skeith III, “Public key encryption that allows pir queries,” in Advances in Cryptology-CRYPTO 2007. Springer, 2007, pp. 50–67.

[3] D. Boneh, G. Di Crescenzo, R. Ostrovsky, and G. Persiano, “Public key encryption with keyword search,” in Advances in CryptologyEurocrypt 2004. Springer, 2004, pp. 506–522.

[4] D. X. Song, D. Wagner, and A. Perrig, “Practical techniques for searches on encrypted data,” in Security and Privacy, 2000. S&P 2000. Proceedings. 2000 IEEE Symposium on. IEEE, 2000, pp. 44–55.

[5] E.-J. Goh et al., “Secure indexes.” IACR Cryptology eprint Archive,

[6] K. Ren, C. Wang, Q. Wang et al., “Security challenges for the public cloud,” IEEE Internet Computing, vol. 16, no. 1, pp. 69–73, 2012.

[7] O. Goldreich and R. Ostrovsky, “Software protection and simulation on oblivious rams,” Journal of the ACM (JACM), vol. 43, no. 3, pp. 431–473, 1996.

[8] S. Kamara and K. Lauter, “Cryptographic cloud storage,” in Financial Cryptography and Data Security. Springer, 2010, pp. 136–149. vol. 2003, p. 216, 2003.