Secure K-NN Query and multiple key for secure medical data retrieval

Anusha C[1]and J. Bhuvana[2]

[1]Research Scholar, School of Computer Science and Information Technology, JAIN (Deemed to be University),
Bangalore, India

[2]Associate Professor, School of Computer Science and Information Technology, JAIN(Deemed to be University),Bangalore, India

ABSTRACT

The nearest neighbors query is a fundamental database query with a wide range of applications such as classification and clustering. Massive medical data is increasingly being outsourced to the cloud in encrypted form to take advantage of the benefits of cloud computing, with the promise of anonymity and privacy. Prior research has assumed that query users (QUs) are completely trustworthy and have access to the data owner's (DO) key, which is used to encrypt and decrypt outsourced data. In many cases, the assumptions are impractical because many users are either untrustworthy or unaware of the key. With multiple keys, a unique. Using his own key, the DO encrypts and decrypts outsourced data. QUs retrieve data in part with one key and in full with another private key. This ensures that data sharing is secure. The goal of our research is to develop a new strategy for doing secure k-NN queries on encrypted cloud data with multiple keys. Using his own key, the DO encrypts and decrypts medical data. A distributed two trapdoors public-key cryptosystem (DT-PKC) and a set of secure two-party computing protocols are used to build the proposed scheme, which not only protects data confidentiality and query privacy but also enables the offline data owner.

1. INTRODUCTION

The k-nearest neighbors (k-NN) query is a basic query operation in most real world datasets, which aims to compute k nearest neighbors or points in the given dataset and a given query point. Researchers have researched many applications on cloud security and privacy issues and also adding k-NN queries on outsourced encrypted cloud data. The general approach is for the data owner (DO) to encrypt data before outsourcing; during query execution, authorized query users (QUs) perform a complex series of encryption and decryption operations.

- ❖ To store the Encrypted data on cloud.
- ❖ Retrieve the Encrypted Cloud Data Using query

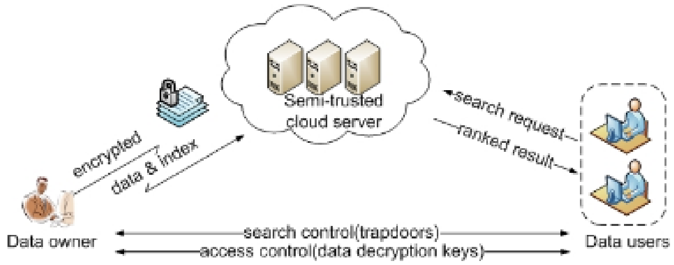
Semi-trusted cloud server
encrypted data & index
search request
ranked result
Data owner
Data users
search control(trapdoors)
access control(data decryption keys)

Fig1: System Architecture of Search on Encrypted dataset in cloud storage

2. LITERATURE SURVEY

Service providers like Google and Amazon are moving into the SaaS (Software as a Service) business. They turn their huge infrastructure into a cloud-computing environment and aggressively recruit businesses to run applications on their platforms. To enforce security and privacy on such a service model, we need to protect the data running on the platform. Unfortunately, traditional encryption methods that aim at providing "unbreakable" protection are often not adequate because they do not support the execution of applications such as database queries on the encrypted data. In this paper we discuss the general problem of secure computation on an encrypted database and propose a SCONEDB (Secure Computation ON an Encrypted Data Base) model, which captures the execution and security requirements. [3].

Data are stored to a third party in cloud environments and query processing is also done by the third party to reduce the expense to maintain the system. Although there are lots of advantages in using independent third parties in query processing, security problems become more crucial since we cannot completely trust the third parties which can be easily corrupted or malfunctioning. The security problems with un trusted third parties are multifaceted in several areas such as privacy, authentication, and recovery. For privacy, the third party should not be able to know what the user's query is since the query itself describes the user's interest. For authentication, the user should be able to verify that the information from the third party is not tampered since the correctness of the query results depends upon the correctness of the information from the third party. For recovery, when the result is found to be forged by an adversary, we should be able to find the adversary and get a correct result by removing the adversary. To address these challenges, we propose several schemes. First, with respect to secure kNN query processing and secure proximity detection, we give novel schemes based on Mutable Order Preserving Encryption (MOPE) and Secure Point Evaluation Method (SPEM). Second, for authenticated top-k aggregation, we suggest novel schemes using Three Phase Uniform Threshold Algorithm, Merkle Hash Tree, and Condensed-RSA. Third, for detecting malicious nodes, we propose novel algorithms based on Additively Homomorphism Encryption and Transmission [4].

The advantages of cloud service while preserving security and privacy, huge data are increasingly outsourced to cloud in encrypted form. Unfortunately, most conventional encryption schemes cannot smoothly support encrypted data analysis and processing. As a significant topic, several schemes have been recently proposed to securely compute k-nearest neighbors (k-NN) on encrypted data being outsourced to cloud server (CS). However, most existing k-NN search methods assume query users (QUs) are fully-trusted and know the key of data owner (DO) to encrypt/decrypt outsourced database. It is not realistic in many situations. In this paper, we propose a new secure k-NN query scheme on encrypted cloud data. Our approach simultaneously achieves: (1) data privacy against CS: the encrypted database can resist potential attacks of CS, (2) key confidentiality against QUs: to avoid the problems caused by key-sharing, QUs cannot learn DO's key, (3) query privacy against CS and DO: the privacy of query points is preserved as well, (4) query controllability: QUs cannot launch a feasible k-NN query for any new point without approval of DO. [1].

There have led to an increase in the capability to store and record personal data (microdata) in the cloud. In most cases, data providers have no/little control that has led to concern that the personal data may be beached. Microaggregation techniques seek to protect microdata in such a way that data can be published and mined without providing any private information that can be linked to specific individuals. An optimal microaggregation method must minimize the information loss resulting from this replacement process. The challenge is how to minimize the information loss during the microaggregation process. [2].

User privacy has been a major concern against the widespread adoption of the cloud technology. A full-fledged cloud data service should effectively support data utilization tasks, especially flexible data search functionalities, while simultaneously achieve user privacy assurance and meet practical system-level performance requirements. In this position paper, we identify the importance and challenges of designing privacy-assured, \textit{flexible} and \textit{practically efficient} search mechanisms for outsourced cloud data services. In particular, we focus on two representative types of flexible search functionalities: ranked keyword search, and search over structured data. Although these functionalities are already prevalent in information retrieval in the plaintext domain, realizing them in the encrypted domain requires non-trivial effort and is relatively new. In light of this, we first describe several existing technical approaches proposed by us and other researchers, and identify their advantages and limitations [9].

3. EXISTING SYSTEM

An asymmetric scalar-product-preserving encryption (ASPE) to preserve scalar product between the query vector and any vector for distance comparison, which is sufficient to find k-NN. Instead of finding exact nearest neighbor, Improved ASPE scheme was proposed to solve the problem without sharing key with query users. Instead, query users interact with the data owner to derive a query encryption. That is, these schemes require data owner to be constantly online. Further improved ASPE scheme proposed by Zhu et al. ,which can support the offline data owner.

3.1 Disadvantages of the existing system

All the above schemes have assumed that the query users are fully-trusted and have the access to the key for encrypting and decrypting outsourced data. It will bring about several problems in the real world.

- Cloud platform can totally break the outsourced database once the key is obtained from any compromised query user. It is obvious that each query user could be one of the lucrative targets for attackers.
- Data owner may have no enough trust on each query user in many applications which will limit the scope of these schemes.
- Once query users receive the key, their query processing will not be controlled by data owner any more, and it is difficult to revoke the access even they are deemed to be untrustworthy. In general, these schemes with key-sharing are still far from being practical in most instances.

APSE these schemes disclose more or less information about data owner’s key. ASPE method cannot be proved to resist the chosen-plaintext attack (CPA).

4. METHODOLOGY

Cloud server module

A cloud server has plenty of storage space to store and manage data from all valid query users. A cloud server also stores all intermediate and final results in encrypted form during the protocol implementation process. Furthermore, a cloud server can perform certain computations on encrypted data.

In the system, a cloud service provider provides online computation services. As a result, the cloud service provider can offload the calculation task to the cloud platform and collaborate with it to find the k-NN for the query user while maintaining privacy.

Data owner module

The DO generates data, encrypts it with his public key, and then sends it to a cloud platform for storage. Let D denote the original database of the data owner. We assume that DO's database has n records, denoted by D = p1, p2,... pn, and that each point is an m-dimensional vector, i.e. pi = (p_{i1}; p_{i2},... p_{im}), for all I = 1, 2,... n. It is assumed that data owner delegated D′ and the future query processing service to the CP. Each Query User has its own set of m-dimensional query points. QU would like to retrieve the top k records in D that are closest to the query point q = (q1, q2,... qm) based on the Euclidean distance.

Secure distance computing

Each query user has some private query points. For the query point q = (q1, q2, … qm), QU would like to retrieve the top k records that are closest to the query point in D according to the Euclidean distance, $\|pi, q\| = \surd\Sigma^{m}_{j=1} (pi,j - qj)^{2}$. QU initially sends his query q (in encrypted form) to cloud provider. After this, CP and CSP involve in a set of sub-protocols to securely compute the Euclidean distance then retrieve the k-NN in D′ and return encrypted result to the query user. At the end of our scheme, only the corresponding query user can decrypt the result points.

Secure K-NN retrieval

Two secure protocols Secure Minimum and Secure Minimum Index of n numbers, based here to build the secure k-NN query retrieval scheme. The aim of this protocol is for CP and CSP jointly compute the encryption and the encrypted index of the minimum number which will be known only to CP. Any cipher text can be decrypted using decryption algorithm with strong private key SK.

5. CONCLUSION

To conclude, The nearest neighbors query is a fundamental database query with a wide range of applications such as classification and clustering. Massive medical data is increasingly being outsourced to the cloud in encrypted form to take advantage of the benefits of cloud computing, with the promise of anonymity and privacy. Prior research has assumed that query users (QUs) are completely trustworthy and have access to the data owner's (DO) key, which is used to encrypt and decrypt outsourced data. In many cases, the assumptions are impractical because many users are either untrustworthy or unaware of the key. The problem of supporting k-NN query over encrypted cloud data is handled, while the data owner cannot share his key with query users. For this a new solution is proposed with multiple keys to solve the key sharing problems thoroughly. At the core of the scheme, a series of novel secure protocols is proposed based on Twin-Cloud structure and A distributed two trapdoors public-key cryptosystem DT-PKC cryptosystem. Proposed scheme can protect the data confidentiality and query privacy.

6. ACKNOWLEDGEMENT

I should convey my real tendency and obligation to Dr M N Nachappa and Dr. J. Bhuvana undertaking facilitators for their effective steerage and consistent inspirations all through my assessment work. Their ideal bearing, absolute co-action and second discernment have made my work gainful.

7. REFERENCES

[1]. Zhu, Y., Huang, Z., & Takagi, T. (2016). Secure and controllable k-NN query over encrypted cloud data with key confidentiality. Journal of Parallel and Distributed Computing, 89, 1-12. https://doi.org/10.1016/j.jpdc.2015.11.004.

[2]. M. E. Kabir, A. N. Mahmood, H. Wang and A. K. Mustafa, "Microaggregation Sorting Framework for K-Anonymity Statistical Disclosure Control in Cloud Computing," in IEEE Transactions on Cloud Computing, vol. 8, no. 2, pp. 408-417, 1 April-June 2020, doi: 10.1109/TCC.2015.2469649..

[3] W. K. Wong, D. W.-l. Cheung, B. Kao, and N. Mamoulis Secure kNN computation on encrypted databases IEEE Transactions on Parallel and Distributed Systems, vol. 25, no. 9, pp. 2245–2254, Sept 2014.

[4]. S. Choi, G. Ghinita, H. -S. Lim and E. Bertino, "Secure kNN Query Processing in Untrusted Cloud Environments," in IEEE Transactions on Knowledge and Data Engineering, vol. 26, no. 11, pp. 2818-2831, Nov. 2014, doi: 10.1109/TKDE.2014.2302434.

[5]. Z. Yan, W. Ding, X. Yu, H. Zhu and R. H. Deng, "Deduplication on Encrypted Big Data in Cloud," in IEEE Transactions on Big Data, vol. 2, no. 2, pp. 138-150, 1 June 2016, doi: 10.1109/TBDATA.2016.2587659.

[6]. S. Bugiel, S. Nurnberger, A. Sadeghi, and T. Schneider Twin clouds: An architecture for secure cloud computing in 2015 IEEE Conference on Computer Communications (INFOCOM). IEEE, 2015, pp. 2659–2667

[7]. N. Cao, Z. Yang, C. Wang, K. Ren and W. Lou, "Privacy-Preserving Query over Encrypted Graph-Structured Data in Cloud Computing," 2011 31st International Conference on Distributed Computing Systems, 2011, pp. 393-402, doi: 10.1109/ICDCS.2011.84.

[8]. B. Yao, F. Li, and X. Xiao, “Secure nearest neighbor revisited,” in Data Engineering (ICDE), 2013 IEEE 29th International Conference on. IEEE, 2013, pp. 733–744.

[9]. M. Li, S. Yu, W. Lou and Y. T. Hou, "Toward Privacy-Assured Cloud Data Services with Flexible Search Functionalities," 2012 32nd International Conference on Distributed Computing Systems Workshops, 2012, pp. 466-470, doi: 10.1109/ICDCSW.2012.41