International Journal of Advances in Engineering and Management (IJAEM)
Volume 3, Issue 5 May 2021, pp: 875-879 www.ijaem.net ISSN: 2395-5252

Privacy Protection of Decentralized Data Using Skyline and Black Chain Privacy

Prof.R.Raja,[1] M.Tech., Arockia Barvin.K,[2] Gokul.T,[3] Kathiresan.S[4]
Dhanalakshmi Srinivasan Engineering College (Autonomous), Perambalur.

Submitted: 05-05-2021 Revised: 17-05-2021 Accepted: 20-05-2021

ABSTRACT: Outsourcing data and computation to cloud server provides a cost-effective way to support large scale data storage and query processing. However, due to security and privacy concerns, sensitive data (e.g., medical records) need to be protected from the cloud server and other unauthorized users. One approach is to outsource encrypted data to the cloud server and have the cloud server perform query processing on the encrypted data only. It remains a challenging task to support various queries over encrypted data in a secure and efficient way such that the cloud server does not gain any knowledge about the data, query, and query result. In this paper, we study the problem of secure skyline queries over encrypted data. The skyline query is particularly important for multi-criteria decision making but also presents significant challenges due to its complex computations. We propose a fully secure skyline query protocol on data encrypted using semantically-secure encryption. As a key subroutine, we present a new secure dominance protocol, which can be also used as a building block for other queries. Furthermore, we demonstrate two optimizations, data partitioning and lazy merging, to further reduce the computation load.

I. INTRODUCTION

As an emerging computing paradigm, cloud computing attracts increasing attention from both research and industry communities. Outsourcing data and computation to cloud server provides a cost effective way to support large scale data storage and query processing. However, due to security and privacy concerns, sensitive data need to be protected from the cloud server as well as other unauthorized users. A common approach to protect the confidentiality of outsourced data is to encrypt the data. To protect the confidentiality of the query from cloud server, authorized clients also send encrypted queries to the cloud server. It illustrates our problem scenario of secure query processing over encrypted data in the cloud. The data owner outsources encrypted data to the cloud server. The cloud server processes encrypted queries from the client on the encrypted data and returns the query result to the client. During the query processing, the cloud server should not gain any knowledge about the data, data patterns, query, and query result. Fully homomorphic encryption schemes ensure strong security while enabling arbitrary computations on the encrypted data. However, the computation cost is prohibitive in practice. Trusted hardware such as Intel's Software Guard Extensions (SGX) brings a promising alternative, but still has limitations in its security guarantees. Many techniques have been proposed to support specific queries or computations on encrypted data with varying degrees of security guarantee and efficiency (e.g., by weaker encryptions). Focusing on similarity search, secure k-nearest neighbor (kNN) queries, which return k most similar (closest) records given a query record, have been extensively studied.

II. LITERATURE SURVEY

2.1 Anonymous Usage of Location-Based Services Through Spatial and Temporal Cloaking, Authors: B. Rogers, S. Chhabra, Y. Solihin, and M. Prvulovic

Advances in sensing and tracking technology enable location-based applications but they also create significant privacy risks. Anonymity can provide a high degree of privacy, save service users from dealing with service providers' privacy policies, and reduce the service providers' requirements for safeguarding private information. This paper presents a middleware architecture and algorithms that can be used by a centralized location broker service. The adaptive algorithms adjust the resolution of location information along spatial or temporal dimensions to meet specified anonymity constraints based on the entities who may be using location services within a given area. Using a model based on automotive traffic counts and cartographic material, we estimate the realistically expected spatial resolution for different anonymity constraints. The median resolution generated by our algorithms is 125 meters. Thus, anonymous location-based requests for urban areas would have the same accuracy currently needed for E-911 services; this would provide sufficient resolution for wayfinding, automated bus routing services and similar location-dependent services.

Algorithm

- Location-based services(LBS)
- Global Positioning System (GPS)

Disadvantages

- An adaptive quad tree-based algorithm that decreases the spatial resolution of location information to meet a specified anonymity constraint.

Advantages

- It removes the need to evaluate potentially complex service provider privacy policies

2.2 Dummy Based Privacy Preservation in Continuous Querying Road Network Services, Authors: Fincy Francis1, Aparna M.S, Anitta Vincent

Dummies are useful for improving success rate because they can be always available, however, using dummies have two main challenges. The first is how to generate a dummy that is indistinguishable from a real user especially on road networks which have varied movement trends. Secondly, dummies can be used to launch attacks on the location based server by malicious clients which affects the business of the service providers. In this paper, we propose a novel client orientated privacy preserving scheme for continuously querying road network service that is also capable of protecting location based servers from attacks. We employed an offline trajectory clustering algorithm that clustered users' trajectory and used the derived parameters to generate a cost effective reusable realistic dummies on road network. To overcome malicious clients using dummies to launch attacks on location based servers, we developed a privacy preserving verification protocol capable of checking the activities of all clients in privacy preserving manner to curb such attacks. We tested the efficiency of our algorithm with some defined evaluation metrics, and it provided an effective privacy protection, satisfied clients at all times within an excellent processing time at a reasonable dummy processing cost when continuously querying road network services.

Algorithm

- Privacy Preservation Algorithm;
- Clustering Algorithm.

Disadvantages

- How to generate a dummy that is indistinguishable from a real user especially on road networks which have varied movement trends

Advantages

- Increase robustness for on-line signature verification .

III. SYSTEM ANALYSIS

3.1 Existing System

A secure and efficient query framework called Q-Shield to enable scalable multi-user utilization of outsourced data. It adopts Intel SGX to establish hardware-assisted enclaves in the un-trusted cloud so as to protect the confidentiality and integrity of sensitive data run inside. Besides, Q-Shield incorporates a generic SQL-style query model such that it is capable of handling majority of common data query tasks. Furthermore, since cloud applications scenarios often use different data models, Q-Shield exploits a widely-adopted and flexible document-oriented data model to enable compatibility and support high generality.In order to tackle the multi-user query control issue, a lightweight secret sharing scheme is exposed in existing procedures. The core idea is to let the data owner assign an attested enclave a key share and each authorized data user unique key. The encryption key can be reconstructed and used for decryption is delivered to the enclave that holds per query. In existing the enclave for a small time-interval cannot greatly reduce the possibility of being exposed through side-channel attacks; the non-trivial key leakage problem can then be solved.

Disadvantages

- The consumers involved in the cloud services has no knowledge about their data whether the data has been confidential.
- There is a chance for data to be accessed by anonymous users around the distributed servers.
- By the cloud owner side it reduces the guarantee of data that has been in the server which results in reduced data privacy.
- The latency of the entire system has been quiet high due to low truthfulness or confidentiality of the data that has been stored

3.2 Proposed System

In the proposed system along with the encrypted skyline queries we implement illusion data, intruder breach and the encrypted database. The user add or upload his data to the distributed database along with the skyline queries and the additional features added in the process are the data are added in distributed server in an illusion format, so that we can prevent the data from the breach of intruders and along with that the database has been encrypted so that there will be quiet better security to the user data. Along with that intruder alert message will be given to the user from the data owner if any abnormal activities occur in the distributed server.

Advantages

- Enormously increases the authentication procedures of entire data that has been stored in the distributed servers.
- User has a great influence on his data or the information that has been outsourced to the distributed or te cloud server.
- Along with the data stored in the owner side the users who stores their data has an enormous features to provide perfect authentication to their information that has been decentralized.
- Reduces the intruder breach up to the core which enhances the security features.

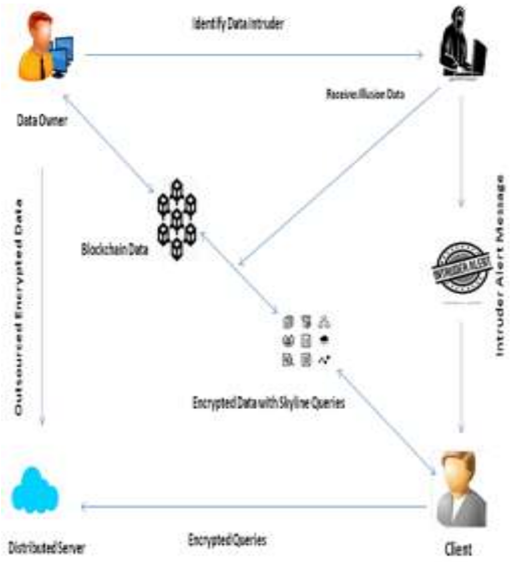
Software Description

PHP – Overview

PHP is a recursive acronym for "PHP: Hypertext Preprocessor". PHP is a server side scripting language that is embedded in HTML. It is used to manage dynamic content, databases, session tracking, even build entire e-commerce sites. The PHP Hypertext Preprocessor (PHP) is a programming language that allows web developers to create dynamic content that interacts with databases. PHP is basically used for developing web based software applications. This tutorial helps you to build your base with PHP.

IV. SYSTEM DESCRIPTION

Architecture Diagram



Module Description

Skyline

A **skyline query** returns the objects that cannot be dominated by any other objects. In the case of a dataset consisting of multidimensional objects, an object dominates another object if it is as good in all dimensions, and better in at least one dimension. Given a dominance relationship in a dataset, a skyline query returns the objects that cannot be dominated by any other objects. In the case of a dataset consisting of multidimensional objects, an object dominates another object if it is as good in all dimensions, and better in at least one dimension. The definition of skyline queries in multidimensional datasets is identical with the known maximum vector problem. In these early works, skyline computation was an algorithmic problem in nature, and all data were assumed to reside in memory. However, nowadays we face big datasets which are stored in secondary memory. Having the data on disk(s), the proposed algorithms for skyline query processing are separated in two categories: index-based algorithms and non-index-based algorithms.

Secure Query Processing On Encrypted Data

Fully homomorphic encryption schemes enable arbitrary computations on encrypted data.Even though it is shown that we build encryption schemes, to provide better security the data has to be encrypted. Many techniques are proposed to support computations on encrypted data with security guarantee and efficiency. We are aware of intruder in any formal work on secure skyline queries over encrypted data with semantic security. An important research has been made to answer the problem that users may be interested for skyline queries in subspaces of the data. In a framework is proposed which uses skyline groups and decisive subspaces, to compute the skyline in any required subspace. Upon this framework an efficient algorithm is proposed, named SKYEY, which applies a top-down approach to recursively compute the skyline in subspaces. Pre-sorting strategies and multidimensional roll-up and drill-down analysis reduce the set of objects to be searched. A similar approach, the SKYCUBE, is proposed in, which is the union of the skylines of all possible non-empty subsets of a given set of dimensions. Several computation sharing strategies are used, based on effectively identifying the computation dependencies among multiple related skyline queries. Bottom-Up and Top-Down algorithms are proposed to compute the SKYCUBE efficiently

Secure Multi-party Computation (SMC)

Secure multi-party computation (also known as secure computation, multi-party computation (MPC), or privacy-preserving computation) is a subfield of cryptography with the goal of creating methods for parties to jointly compute a function over their inputs while keeping those inputs private. Unlike traditional cryptographic tasks, where cryptography assures security and integrity of communication or storage and the adversary is outside the system of participants (an eavesdropper on the sender and receiver), the cryptography in this model protects participants' privacy from each other.

The foundation for secure multi-party computation started in the late 1970s with the work on mental poker, cryptographic work that simulates game playing/computational tasks over distances without requiring a trusted third party. Note that traditionally, cryptography was about concealing content, while this new type of computation and protocol is about concealing partial information about data while computing with the data from many sources, and correctly producing outputs.

Secure Multi-Party Computation (SMC) is an important subset of cryptography. It has the potential to enable real data privacy. SMC seeks to find ways for parties to jointly compute a function using their inputs, while keeping these inputs private.

SMC was proposed for multi-party and refers to problem of users with low security for their data.So we implement blockchain technology to ensure protection to the user data along with the skyline queries. The user provides authentication for their data and every data has been interconnected with their own hash functions and asymmetric keys.The goal is to obtain an encrypted result of a function on the input without disclosing the original input to the administrator.

Stegnography Data

Steganography is the art of hiding data in a seemingly innocuous cover medium. For example – any sensitive data can be hidden inside a digital image. Steganography provides better security than cryptography because cryptography hides the contents of the message but not the existence of the message. Steganography is the art of hiding information and an effort to conceal the existence of the embedded information. It serves as a better way of securing message than cryptography which only conceals the content of the message not the existence of the message. Original message is being hidden within a carrier such that the changes so occurred in the carrier are not observable. In this paper we will discuss how digital images can be used as a carrier to hide messages. This paper also analyses the performance of some of the steganography tools. Steganography is a useful tool that allows covert transmission of information over an over the communications channel. Combining secret image with the carrier image gives the hidden image. The hidden image is difficult to detect without retrieval. This paper will take an in-depth look at this technology by introducing the reader to various concepts of Steganography, a brief history of Steganography and a look at some of the Steganography technique. The data are stored with different digit orderings, blended in a single hybrid data.Apart from proper authentication scheme the intruder receives only the illusioned structure of the data that has been stored in the distributed server.

Blockchain Privacy

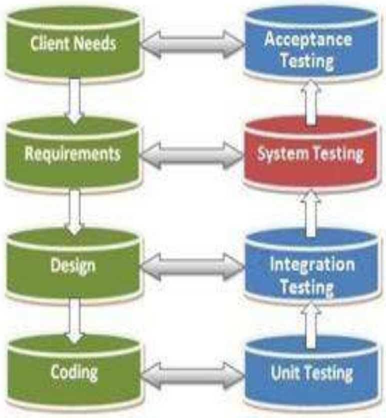
A **blockchain**, originally **block chain**, is a growing list of records,called blocks, that are linked using cryptography. Each block contains a cryptographic hash of the previous block, a timestamp, and transaction data (generally represented asa Merkle tree). By design, a blockchain is resistant to modification of the data. It is "an open, distributed ledger that can record transactions between two parties efficiently and in a verifiable andpermanent way". For use as a distributed ledger, a blockchain is typically managed by a peer-to-peer network collectively adhering to a protocol for inter-node communication and validating new blocks. Once recorded, the data in any given block cannot be altered retroactively without alteration of all subsequent blocks, which requires consensus of the network majority. Although blockchain records are not unalterable, blockchains may be considered secure by design and exemplify a distributed computing system with high Byzantine faulttolerance. Decentralized consensus has therefore been claimed with a blockchain.

V. TESTING

Test Case

File level deduplication will save a relatively large memory space. In general, file level deduplication view multiple copies of same file. It stores first file and then it links other references to the first file. Only one copy will be stored. In testing, even though file names are same, the system can able to detect deduplication. If we upload the same file by using different names, it will view only the content and not names. Thus redundant data is avoided.

In registration phase, the user may not registered before and type their information. So if the user is new user, the alert message will display that the user is not registered before.



There are two types of software testing.
1. Black box testing
2. White box testing

VI. CONCLUSION

In this paper, we proposed a fully secure skyline protocol on encrypted data using two non-colluding cloud servers under the semi-honest model. It ensures semantic security in that the cloud servers knows nothing about the data including indirect data patterns, query, as well as the query result. In addition, the client and data owner do not need to participate in the computation.We also presented a secure dominance protocol which can be used by skyline queries as well as other queries. Furthermore, we demonstrated two optimizations, data partitioning and lazy merging, to further reduce the computation load. Finally, we presented our implementation of the protocol and demonstrated the feasibility and efficiency of the solution. Along with this we introduce more new techniques like intruder breach, illusion data occurrences and the encrypted data as well as the encrypted data and database. So that the data that has been saved in the server are with quiet better privacy and security.

VII. FUTURE ENHANCEMENT

As for future work, we plan to optimize the communication time complexity to further improve the performance of the protocol. Additional features like encryption and decryption using cryptography as well as images and videos can also be implemented in case of illusion data. Stegnography techniques can also be implemented for better future enhancement.

REFERENCES

[1]. K. Fawaz and K. G. Shin, “Location privacy protection for smartphone users,” in Proceedings of the ACM SIGSAC Conference on Computer and Communications Security, 2014, pp. 239–250.

[2]. K. Fawaz, H. Feng, and K. G. Shin, “Anatomization and protection of mobile apps’ location privacy threats,” in fUSENIXg Security Symposium, 2015, pp. 753–768.

[3]. B. Niu, Q. Li, X. Zhu, G. Cao, and H. Li, “Achieving k-anonymity in privacy-aware location-based services,” in IEEE INFOCOM, 2014, pp. 754–762.

[4]. P.-R. Lei, W.-C. Peng, I.-J. Su, C.-P. Chang et al., “Dummy-based schemes for protecting movement trajectories,” Journal of Information Science and Engineering, vol. 28, no. 2, pp. 335–350, 2012.

[5]. T. Hara, A. Suzuki, M. Iwata, Y. Arase, and X. Xie, “Dummy-based user location anonymization under real-world constraints,” IEEE Access, vol. 4, pp. 673–687, 2016.