ISSN: 2454-9940

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

E-Mail :
editor.ijasem@gmail.com
editor@ijasem.org

www.ijasem.org

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

ISSN2454-9940
www.ijsem.org
Vol 12, Issuse.3 Sep 2018

K-NEAREST NEIGHBOR CLASSIFICATIONOVER SEMANTICALLY SECURE ENCRYPTEDRELATIONAL DATA

[#1]**Mr.CHADA SAMPATH REDDY,** *Assistant Professor*

[#2]**Mr.DASARI SHANTHI KUMAR,** *Assistant Professor*

Department of Computer Science and Engineering,

SREE CHAITANYA INSTITUTE OF TECHNOLOGICAL SCIENCES, KARIMNAGAR, TS.

ABSTRACT: Banking, medicine, scientific research, and government organizations are just a few of the many industries that put data mining to use. Data mining applications make heavy use of the classification process. Recent years have seen a proliferation of theoretical and practical answers to the classification challenge in response to rising privacy concerns. Multiple security models have allowed for the development of these fixes. As cloud computing grows in popularity, more and more people are taking advantage of its encryption capabilities and outsourcing data mining jobs to remote servers. Cloud data encryption makes obsolete the present privacy protection classification systems. An answer to the problem of how to label encrypted information is the driving force behind this paper. In particular, we suggest employing a safe k-NN classifier made for encrypted cloud-based data. The proposed protocol's goal is to keep all information, including user searches and access patterns, private. To our knowledge, our study is the first to employ the semi-honest model in order to create a trustworthy k-NN classifier capable of handling encrypted data. We also perform an empirical test of our suggested protocol using a dataset collected in the wild and a wide range of tuning parameters to see how well it performs.

Keyword: - *Security, k-NN classifier, outsourced databases, encryption.*

---

1. INTRODUCTION

Because of cloud computing, organizations no longer employ the same methods they used in the past for storing, accessing, and processing their data. Businesses are increasingly adopting cloud-based processing due to its scalability, cost-effectiveness, and low administration overhead. Many businesses have found that by moving their computers to the cloud, they can better ensure the accuracy of their data. Many companies are reluctant to take use of cloud computing's many benefits due to security and usability issues. For added security, sensitive data should be encrypted before being sent to the cloud. When data is encrypted, however, obtaining it without first breaking the encryption is a major challenge regardless of the security

ISSN2454-9940
www.ijsem.org
Vol 12, Issuse.3 Sep 2018

mechanisms in place. This is due to the fact that encrypted data cannot be deciphered. The following example shows even more reasons why people should worry about their privacy.

Consider the case of an insurance firm that has entrusted data mining and the protection of its client database to a cloud provider. Your company's representative can assess the risk associated with a prospective client by using a categorization framework. The first order of business for the representative is to draft a client information history questionnaire. The client's credit history, age, marital status, and other details will need to be provided on this form. Then, the past information can be stored in the cloud and used to predict the upcoming q class label.

The variable q must be encoded before it is sent to the cloud since it includes sensitive data. Only then can you safeguard your customers' personal information.

The foregoing example shows how important it is to safeguard a user's past while that user is performing data mining on a cloud platform using encrypted data (a process known as DMED). Even when data is encrypted, the cloud can nevertheless learn vital and secret information about certain information products by tracking how users interact with it. Data encryption, user search history protection, and information access pattern concealment are three privacy and security requirements that must be met in order to resolve the DMED in the cloud problem.

The current efforts in privacy-preserving data mining (PPDM) to solve the DMED problem are unsatisfactory, whether they are perturbation-based or focused on safeguarded multi-party computation (SMC). Protecting Personal Data Mining is referred to as PPDM. For extremely sensitive material, information perturbation approaches are insufficient since they do not guarantee semantic protection. The problem is made worse because information extraction is hampered by the faulty data. Security cannot be ensured for all parties involved in a secure multiparty computation since doing so would require giving sensitive information to too many people. Furthermore, a large number of complex calculations are conducted on unencrypted data. In this article, we assume that the encrypted data has been stored in the cloud and provide some unique methods for efficiently addressing the DMED problem. The category problem is crucial since resolving it is one of the most frequent activities in data mining. The k-nearest neighbor classification algorithm is used to decipher encrypted data in a cloud computing environment. This report describes the steps involved and the benefits of various classification schemes.

2. LITERATURE SURVEY

This study presents a new and useful approach to off-site data storage. This technique simplifies data access, works well with patterns, and is trustworthy. A storage client can use this method to protect the confidentiality of their data and access types during the reading, writing, and insertion of new information, even when interacting with a storage service provider who might be nosy or cruel. The service provider can't tell the difference between read and write activities, or link two consecutive accesses. In addition, the

ISSN2454-9940
www.ijsem.org
Vol 12, Issuse.3 Sep 2018

consumer is given strong guarantees that the product will work as intended, and unlawful or unethical business practices are expressly forbidden. We have created a new system that can quickly and accurately handle thousands of queries per second on databases greater than 1 terabyte. It's light years ahead of everything else on the market right now.

With a completely homomorphic security strategy, circuits can be tested with encrypted data without having to know how to decode the data, as discussed in this study. There are three steps to the response, and they must be done in that order exactly. A security strategy that can detect unneeded circuits must be able to verify its own decryption circuit, even if it is more complex than the one being tested. Over the next few weeks, we will look into a particularly self-sufficient way for securing public keys that makes use of perfect lattices, a technique that is often referred to as being "bootstrappable." In this approach, ideal lattices are used. Most of the time, lattice-based cryptosystems rely on low-complexity decryption algorithms. The great majority of these algorithms are built from an NC1 complexity class inner item calculation. Perfect lattices have both preservative and multiplicative homeomorphisms, provided that the public key is perfect in a lattice representing a polynomial band. These homeomorphisms are unique to perfect lattices and are crucial for testing conventional circuits. Using the approach outlined in this article, you can divide the dataset D into n pieces, and then quickly reassemble D from any k of those pieces. Keep in mind that the particulars of the k-1 components tell us nothing about D. This strategy greatly simplifies the process of creating secure and efficient key management methods for use in cryptographic systems. These methods still work even if fifty percent of the items are lost or stolen and all but one has been hacked.

Collecting and keeping track of sensitive personal data via paper documents is a significant security risk. Making a universally accessible computer with a single formula is not only unachievable, but also impossible to achieve. This paper describes a feasible and efficient mathematical strategy that could be used to solve the issue. Sharemind is a virtual machine that allows users to securely share computational resources while maintaining their privacy.

When there are several people sharing a computer, this technique is used to check for security flaws in features. Our answer is one-of-a-kind because of the choices we took when disseminating sensitive information and building the protocol bundle. We've come up with a number of viable options to help students deal with the long chats that often arise during instruction. The three members of the handling group in the SHAREMIND protocol are all people who are thought to be trustworthy and invested in the task at hand. Although it does not welcome malevolent users, the honest-but-curious database's design makes things more pleasant than those of conventional centralized databases.

Data mining methods that don't invade people's privacy are the primary emphasis of this piece. Consider the case when two companies, each with their own proprietary database, are wary about sharing sensitive data and want to employ a data mining method on a merged dataset. Knowing that their personal

ISSN2454-9940
www.ijsem.org
Vol 12, Issuse.3 Sep 2018

information will still be used for research and other purposes motivates people to provide their best effort. The problem you've raised can be fixed by using a well-known and all-encompassing protocol, as it pertains to encrypted multi-party computations. However, data mining algorithms are notoriously difficult to implement because their feedback often consists of huge swaths of highly detailed data. Given the current state of affairs, the standard method has no chance of fixing the problem and must be replaced with creative new approaches. Our major goal is to get as much information on decision trees as we can using the widely used ID3 technique. Our protocol outperforms competing solutions since it calls for a relatively low number of interaction units and a normal amount of bandwidth to send data. Because of this, it is a highly effective choice.

This paper explains how to use past sales of individual products to infer rules about their relationships. Confidentiality of financial dealings is ensured by the rules being developed from randomly generated data. It's a shame that the same regulations that have been made public can be utilized to track down instances of rule breaking. It's bad that the rules can be utilized to identify instances of violation, even though a simple randomization mechanism can be employed to enforce organizational principles and make things more easy. Find out what kinds of privacy violations can happen, and then come up with a randomization technique that is much more effective than regular randomization at preventing such violations. The next step is to get the variance and equation for a neutral support estimator. This allows us to show how to apply these equations to exploration methods and allows us to generate item set facilitations from randomized datasets. Experiments were conducted utilizing the criterion, and the outcomes confirmed its efficacy when used on real-world datasets.

Databases' ability to organize data and facilitate communication between users speeds up the process of addressing convenience concerns. Data warehousing and data mining reduce the likelihood of a comfort crime occurring by centralizing information from several sources. Data mining methods that respect users' confidentiality by revealing only the desired outcome could be the key to fixing this problem. This article describes an approach to working with k-nearest neighbor (k-nn) classes that does not make use of specialized software or hardware. The approach also guarantees that no information about the allocated resources or related data is revealed, only the final category result.

According to the findings of this study, it is crucial to employ data mining methodologies that preserve users' privacy when collecting information from various databases in order to prevent the publication of sensitive details. To explore nearby partitioned databases without compromising users' privacy, we introduce a framework, a full model, and iterative methods based on the k Nearest Neighbor (kNN) classifier. The model, the algorithms, and the iterative algorithms make up the framework.

The debate over whether or not encrypted data should support multidimensional variety queries is explored. The problem stems from people's reliance on apps that enable them upload sensitive information to a remote

server and then use that server's processing capacity to carry out other operations. Freelancers that like to keep their data private utilize these apps. In order to generate a trustworthy identification for the data, the proposed solution incorporates a data-splitting technique termed bucketization. Although the server cannot see the values themselves, it can determine whether or not a given record fulfills the requirement. If the query is evaluated with approximations, the resulting set may contain inaccurate data. The computational cost that our technique imposes is passed on to the end user in the form of the elimination of this data. The goal of this study is to provide a way of bucketization that may be useful in addressing some of the difficulties that arise when dealing with multidimensional data. Analytics on price and disclosure risk for a selected bucketization strategy are made available to us. These figures show the time and money involved in computing for the client and the possibility that sensitive data may be shared with a third party. For marketing purposes, bucketization might be problematic when working with datasets that include several dimensions. The goal is to prevent private data from falling into the wrong hands while keeping the client's computing load (question price) below a threshold set by the user. Owners of information can find the sweet spot between the benefits and costs of sharing data with others thanks to the flexible classification options made available by our service. We also do numerous experiments with both synthetic and actual data to uncover the features of the tradeoffs. Competition in the SaaS sector is heating up, with both Google and Amazon gaining ground. They convert their large data centers into a cloud computing environment and actively seek out companies eager to have their applications hosted on their servers. In order for this service model to function smoothly and securely, it is crucial that all data processed on the system be kept confidential at all times. Unfortunately, the performance of applications like database queries that rely on the protected data is often not improved by traditional security measures despite their best efforts to ensure the data's absolute safety. The issue of doing secure calculations on encrypted databases is discussed in this study. We propose SCONEDB (Secure Computation ON an Encrypted Database), a system that combines speed with security. Our primary focus is on calculating k-nearest neighbors (kNN) in private data sets. In this paper, we develop a novel form of encryption dubbed asymmetric scalar-product-preserving encryption (ASPE). It safeguards a specific kind of scalar objects from damage.

We develop two secure methods for doing kNN calculations on encrypted data using APSE. Both have been proved to successfully halt attacks from individuals with varying degrees of expertise and access to resources, though at various financial and time commitments. Research is conducted extensively to ascertain the effectiveness and efficiency of the methods.

3. CONCLUSIONS

Data mining has many applications, but its primary use is in classification tasks like detecting credit card fraud or counting the number of cancer cells in a patient's body. Several classification algorithms that

ISSN2454-9940
www.ijsem.org
Vol 12, Issuse.3 Sep 2018

safeguard user privacy have been proposed in the past decade. Existing methods are ineffective in a cloud database service when data is encrypted on a remote server. In this research, we show how to apply k-NN classification in a secure manner to encrypted cloud data. Data, user query, and data retrieval methods are all kept confidential by our protocol. We used a variety of parameter configurations to test the efficacy of our technique.

REFERENCES

[1] P. Mell and T. Grance, ―The NIST definition of cloud computing (draft),‖ NIST Special Publication, vol. 800, p. 145, 2011.

[2] S. De Capitani di Vimercati, S. Foresti, and P. Samarati, ―Managing and accessing data in the cloud: Privacyrisks and approaches,‖ in Proc. 7th Int. Conf. Risk Security Internet Syst., 2012, pp. 1–9.

[3] P. Williams, R. Sion, and B. Carbunar, ―Building castles out of mud: Practical access pattern privacy and correctness on untrusted storage,‖ in Proc. 15th ACM Conf. Compute. Common. Security, 2008, pp. 139–148.

[4] P. Williams, R. Sion, and B. Carbunar, ―Building castles out of mud: Practical access pattern privacy and correctness on Untrusted storage,‖ in Proc. 15th ACM Conf. Compute. Common. Security, 2008, pp. 139–148.

[5] C. Gentry, ―Fully homomorphic encryption using ideal lattices,‖ in Proc. 41st Annu. ACM Sympos. TheoryCompute. 2009, pp. 169– 178.

[6] B. K. Samanthula, Y. Elmehdwi, and W. Jiang, ―k-nearest neighbor classification over semantically secureencrypted relational data,‖ e-print arXiv: 1403.5001, 2014.

[7] C. Gentry and S. Halevi, ―Implementing gentry's fully-homomorphic encryption scheme,‖ in Proc. 30th Annu. Int. Conf. Theory Appl. Cryptographic Techno. Adv. Cryptol., 2011, pp. 129–148.

a. Shamir, ―How to share a secret,‖ Common. ACM, vol. 22, pp. 612–613, 1979.

[8] D. Bogdanov, S. Laur, and J. Williamson, ―Sharemind: A framework for fast privacy-preserving computations,‖ in Proc. 13th Eur. Symp. Res. Compute. Security: Compute. Security, 2008, pp. 192–206.

[9] Y. Lindell and B. Pinkas, ―Privacy preserving data mining,‖ in Proc. 20th Annu. Int. Cryptol. Conf. Adv. Cryptol., 2000, pp. 36–54.

[10] Evfimievski, R. Srikant, R. Agrawal, and J. Gehrke, ―Privacy preserving mining of association rules,‖ Inf. Syst., vol. 29, no. 4, pp. 343–364, 2004.