International Journal of Innovative Research in Computer Science & Technology (IJIRCST)
ISSN: 2347-5552, Volume-5, Issue-3, May 2017
DOI: 10.21276/ijircst.2017.5.3.7

Data Mining Over Encrypted Data of Database Client Engine Using Hybrid Classification Approach

Bhagyashree Ambulkar, Prof. Gunjan Agre

Abstract— **Data mining has been used in various areas, for example crime agencies, retail industries, financial data analysis, telecommunication industry, biological, among government agencies, etc. Several application handle very delicate data. So these data remains secure and private. In data mining, Classification could be the one of the major task. Going back two full decades various privacy issues are occurs so that many conceptual and feasible solutions to the classification problem have been developed. Similarly daily cloud user is increment tremendously and they have a big possibility to process the offload the information an encrypted form. The information in the cloud has been in encrypted form, recent privacy preserving classification systems are not feasible. In this paper, our proposed hybrid method provides privacy -preserving classifier for encrypted data of relational database and also achieves the marginally better performance for extracting information using k-NN algorithm from encrypted data of relational databases. This paper describes AES encryption technique which is highly secure and efficient.**

***Keywords*— AES Encryption Algorithm, kNN classification, Privacy Preserving Classification, RTree.**

I. INTRODUCTION

Now a days, cloud computing has become the essential feasibility for data owners to outsource their data. Subsequently accessing delicate information from cloud becomes major privacy issue. When information is very delicate, the information should be encrypted before outsourcing to the cloud. So, cloud computing world is improving their working methods related to information especially in the way they store, access and process information. When the information is in encrypted form, any data mining task becomes very difficult without ever unscrambling the information.

Also it is essential to protect the client's accessing patterns while retrieving information from encrypted data of database on cloud as a part of data mining process. Encryption techniques are applied for providing security and privacy for data. When user extracts useful information form encoded data of database, the searching performance is degraded. In this paper we recommend the AES algorithm for providing security and privacy to data and propose the hybrid classification technique to improve searching performance.

Manuscript received May 21, 2017

Bhagyashree Ambulkar, M. Tech. Scholar**,** Department of Computer Science and Engineering, Nagpur Institute of Technology, Nagpur, Nagpur, Maharashtra, India, 9763316989.

Prof. Gunjan Agre, Department of Computer Science and Engineering, Nagpur Institute of Technology, Nagpur, Nagpur, Maharashtra, India, 9503887753.

II. RELATED WORK

Bharath K. Samanthula et. al. [2][14] focus on solving the classification problem over encrypted data. Author proposes a model for a secure k-NN classifier over encrypted data in the cloud. The proposed protocol provides the security for data, confidentiality to user's input query, and data access patterns keep secret. In this proposed method, first developed a secure k-NN classifier over encrypted data under the semi-honest model and then analyze the efficiency of proposed protocol using a real-world datafset under different parameter settings.

E.Vani et. al. [3][14] proposes a novel secure k-nearest neighbour query convention over encrypted data. The many privacy preserving techniques cannot applicable to outsource data of database when resides on cloud in encrypted form. The proposed method protects the customer's record. PPkNN has a more complex issue and it can't be understood without altering the secure k-nearest neighbour procedures over encrypted data. The author extends the previous work to provide a new solution to the PPkNN classifier problem over encrypted data.

Zhihua Xia et. al. [4][14] present a secure multi-keyword ranked search scheme over encrypted cloud data which simultaneously supports dynamic update operations like deletion and insertion of documents. Author constructs a special tree-based index structure and propose a "Greedy Depth-first Search" algorithm to provide efficient multi-keyword ranked search. The secure kNN algorithm is used to encode the index and query vectors, and meanwhile ensure accurate relevance score calculation between encrypted index and query vectors. To fight with statistical attacks, phantom terms are added to the index vector for blinding search results. The proposed scheme can achieve sub-linear search time and deal with the deletion and insertion of documents flexibly due to the use of special tree-based index structure.

Yousef Elmehdwi et. al. [5][14] focus on solving the k-nearest neighbor (kNN) query problem over encrypted database outsourced to a cloud: a user issues an encrypted query record to the cloud, and the cloud returns the k closest

records to the user. Author first presents a basic scheme and proves that such a naive solution is not secure. To provide better security, author proposes a secure kNN protocol that protects the confidentiality of the data, user's input query, and data access patterns. Also, to analyze the efficiency of protocols author go through various experiments. These results indicate that proposed secure protocol is very efficient on the user end, and this lightweight scheme allows a user to use any mobile device to perform the kNN query.

Z. Wang et. al. [6] presents fast query over encrypted character data in database. The operation of encryption and decryption significantly degrades query performance. To solve such a problem, author proposes an approach that can implement SQL query on the encrypted character data. The proposed approach not only stores the encrypted character data, but also converts the character data into the characteristic values via a characteristic function, and stores them in an additional field. When user query the encrypted character data, the principle of two-phase query is applied. The proposed approach firstly implements a coarse query over the encrypted data in order to filter the records not related to the querying conditions. Secondly, decrypt the rest records and implement a refined query over them again.

Zheng Wang et. al. [7] proposes a novel approach that can quickly execute SQL query on the encrypted data. The proposed method not only encrypts character data, but also turns the it into characteristic values via a characteristic function and stores them as additional fields. Also it encrypts the numerical data and creates its B+ tree index before the encryption in order to keep the ordering of each record in the index.

III. PROPOSED SYSTEM

Existing methods on Privacy preserving data mining cannot resolve the Data Mining on Encrypted Data problem. The traditional search process on encrypted data is performed by decrypting the whole data and then retrieves the data. This whole task takes so much time and also reduces the performance of searching [8]. The proposed hybrid classification method provides good solution for privacy preserving and marginally better performance of data retrieval.

The proposed method consists of two phases. In first phase, classification is performed over encrypted data to create the class labels and second phase executes searching task over classified data using RTree algorithms.

A. AES Encryption Algorithm

AES is based on a design principle known as a substitution-permutation network, a combination of both substitution and permutation, and is fast in both software and hardware [18]. The silent feature of AES algorithm is the different key lengths That key length is decided only by knowing how much period of time the security to be required and about the cost. AES has a fixed block size of 128 bits, and a key size of 128, 192, or 256 bits. The number of rounds in AES is variable and depends on the length of the key. AES uses 10, 12, 14 rounds for 128-bit, 192-bit, 256-bit keys respectively. Each of these rounds uses a different round key, which is calculated from the original key used for AES encryption [17]. Encryption consists of following operations [19]:

Step1: Initial Round
- Expand the 16-byte key to get the actual key block to be used
- Do one time initialization of the 16-byte Plain Text Block (called as State)
- XOR the state with the Key block

Step2: Rounds
- Substitute each of the plain text Bytes: This is the first transformation as we are substituting with the byte. We illustrate it with hexadecimal digits.
- Shift Row k of plain text by k bytes: This is the transforming step in which the first row is kept unchanged. Last three rows is shifted cyclically in a convinced number of steps i.e. row 1 is shifted by 1 byte, row 2 is shifted by 2 bytes and row 3 by 3 bytes. .
- Perform Mix Columns Operation: In this step mixing the columns is performed by merging four bytes of every column.
- Add Round Key: In this step subkey is combined with state. For each round, a subkey is derived from the main key. The size of each subkey is the same size as the state. This step XORs the subkey with the each byte of the state.

Step3: Final Round
- Sub Bytes.
- Shift Rows.
- Add Round Key.

The decryption process of encrypted data is carried out in reverse order [19].

B. Classification Algorithm

Classification is the process of finding a set of models or functions that describes and distinguishes data classes or concepts for the purpose of predicting the class of objects whose class labels are unknown. [20]. Classification is works in two steps. The first step is learning step or training phase where a classification model is constructed and second step consist of classification where model is used to predict class labels for given data [20]. The model is determined by analysing a set of training data where class labels are known. Consider each training instance has n attributes: F1, F2..... Fn-1 with one addition class attribute C which defines class of training instance or sample associated with other attributes. Now consider new instance where class attribute is unknown. The model determines the class label for this new instance based on its other attributes.

k-nearest neighbor (KNN) classification does not build classifier model in advance like other classification techniques. The simple idea is that the most similar tuples most belongs to the same class. Based on some pre-defined distance, the k nearest training samples of the sample to be classified and allocate the verity of classes of those k

samples to the new sample data [1]. For k, the value is preselected. A common similarity function is based on the Euclidian distance between two data tuples [21]. For two tuples, P = (p1, p2… pn-1) and R = (r1, r2… rn-1), the Euclidian similarity function is

$$d_2(P, R) = \sqrt{\sum_{i=1}^{n-1} (p_i - r_i)^2}$$

IV. HYBRID KNN CLASSIFICATION ALGORITHM

Step1: Upload dataset D.
Step2: Apply the AES encryption technique on dataset D using 128 bit size key ki , De = encrypt (D, ki).
Step3: Store encoded data D_e on cloud1 and secure key ki on cloud2.
Step4: Apply k-NN classification on encrypted data which creates class labels.
Step5: Read the test data D_t.
Step6: Encode the test data D_t and send to cloud1.
Step7: k-NN searches the similar pattern for k training tuples that are closed to unknown tuples or test data . Euclidian distance metric is used to find closeness between tuples, $d(D_e, D_t)$
Step8: Retrieve the k nearest tuples D_k using RTree.
Step9: Retrieve the corresponding secure key from cloud 2.
Step10: Decrypt D_k.
Step11: Display result.

V. RESULT

The previous research focus on solving the classification problem over encrypted data using k-NN classifier. The existing protocol provides the security for data, confidentiality of user's input query, and the data access patterns keep secret from intruder. Our proposed hybrid k-NN classification method solves the classification problem over encrypted data as well as provides the better performance to retrieve the data from database. In proposed method, AES encryption algorithm is used to provide the privacy and security to outsource data and RTree is used to fast retrieval of data needed by the user. The Fig. 1 indicates the time required for classification. The Fig. 2 shows the execution time required for user's query.

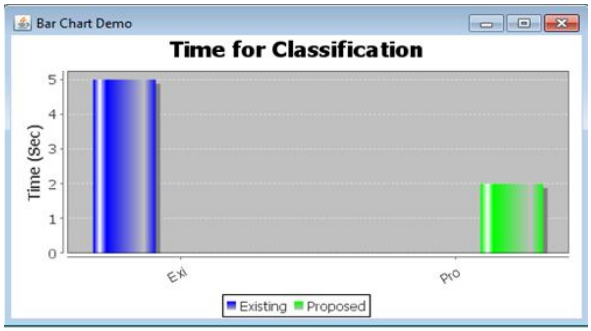

Fig 1: Time required for user's query

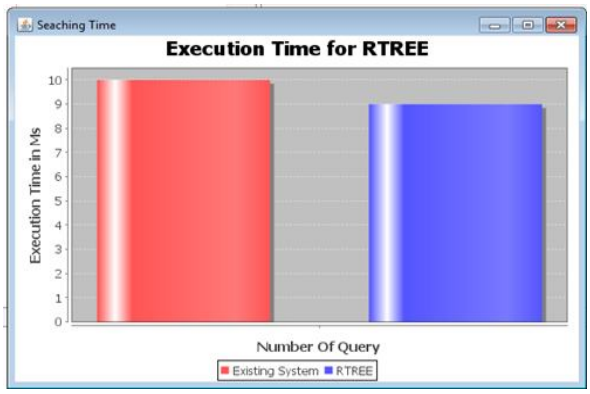

Fig 2: Execution time required for user's query

VI. CONCLUSION

Various types of privacy preserving classification techniques have been introduced from last decades. These methods are not applicable to outsourced databases. The proposed method facilitates efficient computation of nearest neighbour and improves the searching performance over encrypted data of database using RTree. Our proposed system provides the confidentially to data, user's input query, hides the access patterns and improve search performance.

REFERENCES

[1] Bhagyashree Ambulkar, Prof. Gunjan Agre, "Fast Search Processing Over Encrypted Relational Data Using K-Nearest Neighbour Algorithm", International Journal on Recent and Innovation Trends in Computing and Communication Volume: 5 Issue: 4 , ISSN: 2321-8169 , 398 – 401 , May 2017

[2] Bharath K. Samanthula,Yousef Elmehdwi,Wei Jiang,"k-Nearest Neighbor Classification over Semantically Secure Encrypted Relational Data", IEEE Transactions On Knowledge And Data Engineering, Vol. 27, No. 5, May 2015.

[3] E.Vani, S.Veena, D.John Aravindar, "Query Processing Using Privacy Preserving k-NN Classification Over Encrypted Data", International Conference On Information Communication And Embedded System (ICICES), 978-1-5090-2552-7, 2016.

[4] Zhihua Xia, Xinhui Wang, Xingming Sun, Qian Wang,"A Secure and Dynamic Multi-Keyword Ranked Search Scheme over Encrypted Cloud Data",IEEE Transactions On Parallel And Distributed Systems, Vol. 27, No. 2, February 2016.

[5] Yousef Elmehdwi, Bharath K. Samanthula, Wei Jiang, "Secure k-Nearest Neighbor Query over Encrypted Data in Outsourced", ICDE IEEE Conference,978-1-4799-2555-1/14/$31.00 © 2014.

[6] Z. Wang , J. Dai, W. Wang and B. L. Shi," Fast Query over Encrypted Character Data in Database" Communications in Information and Systems, pp. 289-300.

[7] Z. Wang, W. Wang and B. Shi , " Storage and Query over Encrypted Character and Numarical Data in Database", Proceedings of the 2005 The Fifth International Conference on Computer and Information Technology, pp. 77-81,2005

[8] Manish Sharma , Atul Chaudhary, Santosh Kumar, " Query Processing Performance and Searching over Encrypted Data by using an Efficient Algorithm",

International Journal of Computer Applications (0975 – 8887) Volume 62– No.10, January 2013

[9] Kavyashree J, Deepika N, "Secured Access to Cloud Data through Encryption and Top-K Retrieval Using Multiple Keywords", International Journal of Science and Research (IJSR) ISSN (Online): 2319-706

[10] Maleq Khan, Qin Ding and William Perrizo, "K-Nearest Neighbor Classification on Spatial Data Streams Using P-Trees".

[11] http://www.bowdoin.edu/~ltoma/teaching/cs340/spring08/Papers/Rtree-chap1.pdf

[12] Jinka Sravana , Suba. S, "Applying R Trees In Non Spatial Multidimensional Databases", International Journal Of Technology Enhancements And Emerging Engineering Research, Vol 2, Issue 7 28 Issn 2347-4289

[13] Bhagyashree Ambulkar, Prof. Gunjan Agre, "Fast Search Processing Over Encrypted Relational Data Using K-Nearest Neighbor Algorithm",International Journal on Recent and Innovation Trends in Computing and Communication (IJRITCC), May 2017 Volume 5 Issue 5.

[14] J. Han and M. Kamber, "Data Mining Concepts and Techniques", Elevier, 2011.

[15] Bruce Schneier; John Kelsey; Doug Whiting; David Wagner; Chris Hall; Niels Ferguson; Tadayoshi Kohno; et al. (May 2000). "The Twofish Team's Final Comments on AES Selection" (PDF). Nahan Rahman M.K., "Inviolable Data mining in Cloud using AES and Paillier Cryptosystem", International Journal of Advanced Research in Computer and Communication Engineering (IJARCCE) International Conference on Recent Trends in Computing and Communication (ICRTCC 2015) Cochin College of Engineering & Technology Vol. 4, Special Issue 1, June 2015, ISSN (Online) 2278-1021, ISSN (Print) 2319-5940

[16] http://www.tutorialspoint.com/cryptography/advanced_encryption_standard.htm

[17] Bruce Schneier; John Kelsey; Doug Whiting; David Wagner; Chris Hall; Niels Ferguson; Tadayoshi Kohno; et al. "The Twofish Team's Final Comments on AES Selection" , May 2000

[18] Atul Kahate, "Cryptography and Network Security", TMH Publication

[19] Jiawei Han, Micheline Kamber, Jian Pei, "Data Mining : Concepts and Techniques", Morgen Kaufmann" 2001

[20] T. Cover and P. Hart, "Nearest Neighbor pattrern classification", IEEE Trans. Information Theory, 1996

[21] Hong Rong, Huimei Wang, Jian Liu, and Ming Xian, "Privacy-Preserving k-Nearest Neighbor Computation in Multiple Cloud Environments",IEEE, 2169-3536 (c) 2016.

Bhagyashree Ambulkar is M. Tech Scholar in Department of Computer Science and Engineering, Nagpur Institute of Technology, Nagpur, Maharashtra. She received Master of Computer Application (MCA) degree in 2005 from SGBU, Amravati, MS, India. Her research interests are Data Mining, Web Mining, etc.

Gunjan Agre is Assistant Professor in Department of Computer Science and Engineering, Nagpur Institute of Technology, Nagpur, Maharashtra. She received Master in Technology (M.Tech.) in Computer Science and Engineering in 2015 form R.T.M.N.U. University, Nagpur, MS, India. Her research interests are Data mining, Web Crawling and Computer Network.

