Scientific Journal of Impact Factor (SJIF): 4.72

e-ISSN (O): 2348-4470
p-ISSN (P): 2348-6406

International Journal of Advance Engineering and Research Development

Volume 4, Issue 7, July -2017

C5.0 Classification over Encrypted Rational Data

Ms. Supriya Borse[1], Drs. Mrs. Neeta Deshpande[2]

M. E. II (Comp. Dept.)[1], Head of Comp. Dept. DYPCOE[2]

Abstract— *Cloud itself is secure but when the data is outsource, it needs more security. Encryption is one of the technique to secure cloud data. Before outsourcing the data to cloud, we have encrypted it using Paillier Cryptosystem at the server side and then classification techniques are applied. Data Mining is used in many applications such as banking, medicine, scientific research and among government agencies. Classification is one of the commonly used tasks in data mining applications. Data stored on the cloud in encrypted form is classified to decrease the memory requirement, improve the efficiency of the system. There are many classification algorithms such as K-NN, C5.0, SVM, Naive Bayes, Random Forest. we have proposed C5.0 classifier using privacy preserving protocol(PPC5.0), which improves the performance of the system.*

Keywords: *Data mining, Cloud Computing, C5.0 Classifier, Outsourced databases, Security, Encryption.*

1. INTRODUCTION

Recently 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) can be rapidly provisioned and released with negligible management effort or service provider interaction[1]. Cloud model is composed of five essential characteristics, three service models, and four deployment models. Today's digital infrastructure supports innovative ways of storing processing, and disseminating data [2].

In fact, we can store our data in remote access servers, access reliable and coherent services provided by third parties, and use computing power available at multiple locations across the networks. To protect data stored in such an untrusted server model, security systems should offer user's assurances of data access privacy and confidentiality. As a first line of defense, to ensure confidentiality, all data and associated data about data can be encrypted at the client side using non-malleable encryption, before being stored on the server. The data remains encrypted throughout its lifetime on the server and is decrypted by the client upon retrieval [3].

Example: Suppose an insurance company outsourced its encrypted customers database and relevant data mining tasks to a cloud. When an agent from the company wants to identify the risk level of a prospective new customer, the agent can use a classification method to determine the risk level of the customer. First, the agent needs to generate a data record 'q' for the customer containing certain personal information of the customer, e.g., credit score, age, marital status, etc. Then this record can be sent to the remote server i.e. cloud, and the cloud will compute the class label for q. Nevertheless, since q contains sensitive information, to protect the customer's privacy, q should be encrypted before sending it to the cloud [3].

The above example shows that data mining over encrypted data (DMED) on a cloud also needs to protect a user's record when the record is a part of a data mining process. Classification is one important task in many applications of data mining such as business and health-care. Recently, performing data mining in the cloud attracted significant attention. In cloud computing, data owner outsources his/her data to the cloud. However, from user's perspective, privacy becomes an important issue when sensitive data needs to be outsourced to the cloud. The direct way to guard the outsourced data is to apply encryption on the data before outsourcing. Privacy Preserving Data Mining (PPDM) is defined as the process of extracting/deriving the knowledge about data without compromising the privacy of data [3][19]. Using encryption as a way to achieve the data confidentiality may cause another issue at the cloud during the query evaluation [20].

In existing work, author has proposed a novel PPkNN protocol, a secure k-NN classifier over semantically secure encrypted data. To overcome the drawback of existing system we have implemented the PPC5.0 protocol, which also increases the performance of the system.

2. RELATED WORKS

Sara Foresti et al. [2] Characterize different aspects of the privacy problem in emerging scenarios. Illustrate risks, solutions, and open problems related to ensuring privacy of users accessing services or resources in the cloud, sensitive information stored at external parties, and accesses to such an information. Ensuring proper privacy and protection of the information stored, communicated, processed, and disseminated in the cloud as well as of the users accessing such information is one of the grand challenges of our modern society [2].

Peter Williams et al. [3] Introduce a new practical mechanism for remote data storage with efficient access pattern privacy and correctness. A storage client can deploy this mechanism to issue encrypted reads, writes, and inserts to a potentially curious and malicious storage service provider, without revealing information or access patterns. The provider is unable to establish any correlation between successive accesses, or even to distinguish between a read and a write. Moreover, the client is provided with strong correctness assurances for its operations.

Gemplus et al. [4] This paper investigates a novel computational problem, namely the Composite Residuosity Class Problem, and its applications to public-key cryptography. Authour propose a new trapdoor mechanism and derive from this technique three encryption schemes: a trapdoor permutation and two homomorphic probabilistic encryption schemes computationally comparable to RSA. Cryptosystems, based on usual modular arithmetic, are provably secure under appropriate assumptions in the standard model.

Bharath K. Samanthula et al. [5] Focus on solving the classification problem over encrypted data. In particular, author propose a secure k-NN classifier over encrypted data in the cloud. The proposed k-NN protocol protects the confidentiality of the data, users input query, and data access patterns. To the best of our knowledge, work is the first to develop a secure k-NN classifier over encrypted data under the standard semi-honest model. Also, empirically analyze the efficiency of solution through various experiments [5]. Classification is one of the commonly used tasks in data mining applications.

Craig Gentry et al. [6] Describe a working implementation of a variant of Gentry's fully homomorphic encryption scheme. Smart and Vercauteren implemented the underlying somewhat homomorphic scheme. Show a number of optimizations that allow us to implement all aspects of the scheme, including the bootstrapping functionality.

Dan Bogdanov et al. [7] Gathering and processing sensitive data is a difficult task. In fact, there is no common recipe for building the necessary information systems. This paper, present a provably secure and efficient general-purpose computation system to address given problem. The protocols of SHAREMIND are information-theoretically secure in the honest-but-curious model with three computing participants. Although the honest-but-curious model does not tolerate malicious participants, it still provides significantly increased privacy preservation when compared to standard centralized databases.

Rakesh Agrawal et al. [8] Author propose a-novel reconstruction procedure to accurately estimate the distribution of original data values. By using these reconstructed distributions, its able to build classifiers whose accuracy is comparable to the accuracy of classifiers built with the original data. Since the primary task in data mining is the development of models about aggregated data, author has develop accurate models without access to precise information in individual data records, which consider the concrete case of building a decision-tree classifier from trending data in which the values of individual records have been perturbed.

pinkas et al. [9] Addresses the issues of privacy preserving data mining, consider a scenario in which two parties owing confidential database wish to run a data mining algorithm on the union of their database, without reducing any unnecessary information.

Peng Zhang et al. [10] this paper, combine the two strategies of data transform and data hiding to propose a new randomization method, Randomized Response with Partial Hiding (RRPH), for distorting the original data. Then, an effective naive Bayes classifier is presented to predict the class labels for unknown samples according to the distorted data by RRPH. Shown in the analytical and experimental results, our method can obtain significant improvements in terms of privacy, accuracy, and applicability.

Ramakrishnan Srikant et al. [11]Discuss a framework for mining association rules from transactions consisting of categorical items where the data has been randomized to preserve privacy of individual transactions. While it is feasible to recover association rules and preserve privacy using a straightforward "uniform" randomization, the discovered rules can unfortunately be exploited to find privacy breaches. We analyze the nature of privacy breaches "and propose a class of randomization operators that are much more effective than uniform randomization in limiting the breaches.

Roberto J. Bayardo et al. [12] This paper proposes and evaluates an optimization algorithm for the powerful de-identification procedure known as anonymization. A anonymized dataset has the property that each record is indistinguishable from at least others. Even simple restrictions of optimized anonymity are NP-hard, leading to significant computational challenges. Author present a new approach to exploring the space of possible anonymizations that tames the combinatorics of the problem, and develop data-management strategies to reduce reliance on expensive operations such as sorting.

Haibo Hu et al. [13] Paper discuss a holistic and efficient solution that comprises a secure traversal framework and an encryption scheme based on privacy homomorphism. The framework is scalable to large datasets by leveraging an index based approach. Based on this framework, we devise secure protocols for processing typical queries such as k-nearest neighbor queries (kNN) on R-tree index. Moreover, several optimization techniques are presented to improve the efficiency of the query processing protocols. Solution is verified by both theoretical analysis and performance study.

Murat Kantarcoglu et al. [14] This paper presents a method for privately computing k-nn classification from distributed sources without revealing any information about the sources or their data, other than that revealed by the final classification result.

3. PROBLEM DEFINITION

Existing work on privacy-preserving data mining (PPDM) (either perturbation or secure multi-party computation (SMC) based approach) cannot solve the DMED problem. Perturbed data do not possess semantic security, so data perturbation techniques cannot be used to encrypt highly sensitive data. Also the perturbed data do not produce very accurate data mining results. Secure multi-party computation based approach assumes data are distributed and not encrypted at each participating party. In addition, many intermediate computations are performed based on non-encrypted data. As a result, in this paper, we proposed novel methods to effectively solve the DMED problem assuming that the encrypted data are outsourced to a cloud.

4. OUR CONTRIBUTION

In this paper, we propose a novel PPC5.0 protocol, a secure C5.0 classifier over semantically secure encrypted data. Specifically, we focus on the classification problem to improve the accuracy of the system. Each classification technique has their own advantage, to be concrete, paper concentrates on executing the C5.0 classification method over encrypted data in the cloud computing environment.

5. SYSTEM FLOW CHART

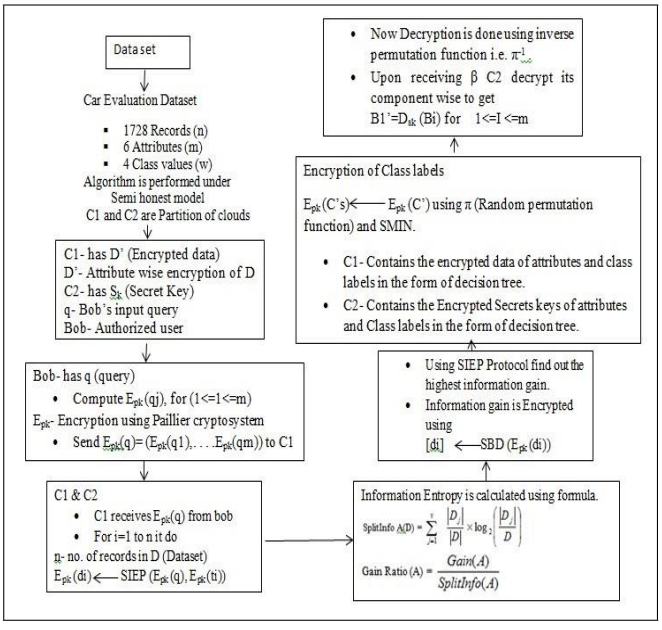

Figure.1. Flow Chart for Proposed System

6. SYSTEM OVERVIEW

Owner Dataset: Owner encrypts the data and that data will store on cloud for maintaining the privacy they want to provide login credential and then they will enter into system. At the time of encryption owner submits the encrypted secret key to cloud and encryption done using public key. Cloud: Cloud is divided into two parts, i.e. C1 and C2. C1 stores the encrypted data and C2 stores the Secret keys generated by encryption algorithm.

Authorized User: User sends the query which is encrypted form that will improve the privacy of system. Clouds collect that query decrypt the query by using secret key and generated result will be send as encrypted that will secure the system. Output of the system does not recognize by end user. Classification: C5.0 classifier is used to classify the encrypted data in cloud environment for classification, need to provide the training to classifier after based on the system will predict class label. Data is stored in the form of decision tree.

Data mining is a knowledge discovery process that analyzes data and produce useful pattern from it. Classification is the technique that uses pre-classified examples to classify the required results. Decision tree is used to model classification process [6]. Existing work on privacy-preserving data mining (PPDM) (either perturbation or secure multi-party computation (SMC) based approach) cannot solve the DMED problem [21]. We are proposing C5.0 algorithm for accurate and efficient mining result in cloud storage along with privacy preserving protocols. C5.0 is the decision tree classifier, which will classify the result set with high accuracy and low memory usage [12].

7. PRIVACY PRESERVING PROTOCOLS

We present a set of generic sub-protocols that will be used in constructing our proposed C5.0 protocol. All of the below protocols are considered under two-party semi-honest setting. In particular, we assume the exist of two semi-honest parties P1 and P2 such that the Paillier secret key s_k is known only to P2 whereas p_k is treated as public [4] [16].

1) **Secure Information Entropy (SIE) protocol:** In this protocol it picks ith attribute in dataset like$d_1,d_2,...d_i$. Find out the highest Information Gain using formula[18]

$$SplitInfoA(D) = \sum_{j=1}^{v} \frac{|D_j|}{|D|} * log_2 \frac{|D_j|}{D}$$

$$GainRatio(A) = \frac{Gain(A)}{SplitInfo(A)}$$

Using the Split information and Gain Ratio we have proposed a Secure Information Entropy (SIE) Protocol. The split information value represents the potential information generated by splitting the training data set D into v partition, corresponding to v outcomes an attribute A. Creates the decision node split based on highest information gain[18][19].

2) **Secure Bit-Decomposition (SBD) Protocol:** Here P_1 with input $E_{pk}(z)$and P_2 securely compute the encryptions of the individual bits of z, where $0 \leq z < 2^l$. The output $[z] = h E_{pk}(z_1,...E_{pk}(z_l))$is known only to P1. Here z_1 and z_l are the most and least significant bits of integer z, respectively [21].

3) **Secure Minimum (SMIN) Protocol:** In this protocol: In this protocol P1 holds private input (u ,v) and p2 holds s_k where
u' = ([u],$E_{pk}(s_u)$) and v' = ([v],$E_{pk}(s_v)$).

4) **Secure Bit-OR (SBOR) protocol:** P1 with input
$E_{pk}(o_1)$,$E_{pk}(o_2)$) and $p2$ securely compute $E_{pk}(o_1 \vee o_2)$ where o_1 and o_2 are 2 bits. The output $E_{pk}(o_1 \vee o_2)$ is known only to P1[21][22].

5) **Secure Multiplication (SM) Protocol:** This protocol considers P1 with input ($E_{pk}(a)$,$E_{pk}(b)$) and outputs $E_{pk}(a * b)$ to P1, where a and b are not known to P1 and P2. During this process, no information regarding a and b is revealed to P1 and P2[21].

6) **Secure minimum out of n numbers:** In this protocol, consider P1 with n encrypted vectors ([d_1], …,[d_n]) along with their respective encrypted secrets and P2 with s_k. Here [d_i] = h$E_{pk}(d_i,1),...,_E pk(d_i,l)$i where ($d_i$,1) and ($d_i$,$l$) are the most and least significant bits of integer d_i respectively, for $1 \leq i \leq n$ [20].

8. PAILLIER CRYPTOSYSTEM

The Paillier cryptosystems an additive homomorphic and probabilistic asymmetric encryption scheme whose security is based on the Decisional Composite Residuosity Assumption[22]. Let E_{pk} be the encryption function with public key p_k given by (N, g) and D_{sk} be the decryption function with secret key s_k given by a trapdoor function λ(that is, the knowledge of the factors of N). Here, N is the RSA modulus of bit length K.
Paillier encryption scheme exhibits the following properties:

Homomorphic Addition:

$D_{sk}(E_{pk}(a + b)) = D_{sk}(E_{pk}(a) * E_{pk}(b) mod N^2)$

Homomorphic Multiplication:

$D_{sk}(E_{pk}(a * b)) = D_{sk}(E_{pk}(a)^b mod N^2)$

Semantic Security: The encryption scheme is semantically secure [23][24]. Briefly, given a set of cipher texts, an adversary cannot deduce any additional information regarding the corresponding plaintexts [25][26].

9. SYSTEM ANALYSIS

we have proposed the PPC5.0 protocol, a secure C5.0 classifier[21] over semantically secure encrypted data. In this protocol, once the encrypted data are outsourced to the cloud, Alice does not participate in any computations. Therefore, no information is revealed to Alice. In particular, the protocol meets the following privacy requirements:

- Contents of *D* or any intermediate results should not be revealed to the cloud.
- Authorized user query q should not be revealed to the cloud.
- c_q should be revealed only to authorized user. In addition, no information other than c_q should be revealed to Bob.
- Data access patterns, such as the records corresponding to the k-nearest neighbors of *q*, should not be revealed to Bob and the cloud (to prevent any inference attacks). Since to overcome the drawback of existing system i.e
- Encryption is done by using only partial homomorphic scheme
- If the data size is large then processing speed will become slow
- Time consuming, as time required to search nearest neighbor is more.

we have implemented the C5.0 classifier over encrypted rational data along with privacy preserving protocol i.e (PPC5.0) to improve the performance of the system.

10. ADVANTAGES OF THE SYSTEM

- The main advantage is interpretability. Decision trees are “white boxes” in the sense that the acquired knowledge can be expressed in a readable form, while KNN (K-Nearest Neighbor), SVM (Support Vector Machine) are generally black boxes, i.e. you cannot read the acquired knowledge in a comprehensible way.
- k-NN and SVM are used for continuous value inputs, unlike Decision Trees that is applicable for continuous and categorical inputs. If you deal with a problem where inputs are categorical values (discrete values) even in part then you have to apply the decision tree.

11. DATASET AND EXPERIMENTAL SETUP

For our experiments, we used the Car Evaluation dataset from the UCI KDD archive [21]. It consists of 1,728 records (i.e., n=1728) and six attributes (i.e., m=6). Also, there is a separate class attribute and the dataset is categorized into four different classes (i.e., w=4). We encrypted this dataset attribute-wise, using the Paillier Cryptosystem whose key size is varied in our experiments, and the encrypted data were stored on our machine.

12. RESULT ANALYSIS

We perform Classification on encrypted data using privacy preserving protocols so, classification is done within the encrypted data and secret keys. The query given by the authorized user is also encrypted using Paillier cryptosystem and stored on the cloud using Secure Minimum protocol.

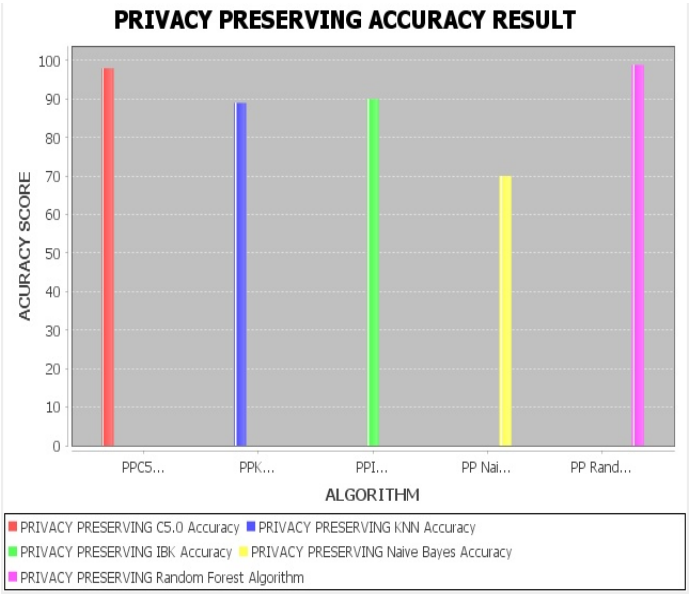

Figure. 2. Accuracy Comparison graph for different classification algorithm .

13. CONCLUSION

Data mining is the process of extracting information from large database or source information from database. Many of the earlier data mining methods extract specific information in exact manner, but in order to preserve user data it not work well. To protect user privacy, various privacy preserving classification techniques have been proposed over the past decade. The existing technique are not applicable to outsourced database environment, where the data reside in encrypted form on a third-party server. The novel privacy preserving C5.0 classifier improves the efficiency of the system. Encryption technique used is RSA with the secure minimum protocol which, improves the performance of the system.

Since in this paper, we are using well known C5.0 classifier and developed privacy-preserving protocol for it over encrypted data. As a future work, we will investigate and extend our research to other classification algorithm.

14. REFERENCES

[1] Samanthula, Bharath K., Yousef Elmehdwi, and Wei Jiang. "K-nearest neighbor classification over semantically secure encrypted relational data."IEEE transactions on Knowledge and data engineering 27.5 (2015): 1261-1273.

[2] di Vimercati, Sabrina De Capitani, Sara Foresti, and Pierangela Sama- rati. "Managing and accessing data in the cloud: Privacy risks and approaches."2012 7th International Conference on Risks and Security of Internet and Systems (CRiSIS). IEEE, 2012.

[3] Williams, Peter, Radu Sion, and Bogdan Carbunar. "Building castles out of mud: practical access pattern privacy and correctness on untrusted storage." Proceedings of the 15th ACM conference on Computer and communications security. ACM, 2008.

[4] Paillier, Pascal. "Public-key cryptosystems based on composite degree residuosity classes." International Conference on the Theory and Appli- cations of Cryptographic Techniques. Springer Berlin Heidelberg, 1999.

[5] Samanthula, Bharath K., Yousef Elmehdwi, and Wei Jiang. "K-nearest neighbor classification over semantically secure encrypted relational data." IEEE transactions on Knowledge and data engineering 27.5 (2015): 1261-1273.

[6] Gentry, Craig, and Shai Halevi. "Implementing Gentrys fully- homomorphic encryption scheme." Annual International Conference on the Theory and Applications of Cryptographic Techniques. Springer Berlin Heidelberg, 2011.

[7] Agrawal, Rakesh, and Ramakrishnan Srikant. "Privacy-preserving data mining." ACM Sigmod Record. Vol. 29. No. 2. ACM, 2000.

[8] Lindell, Yehuda, and Benny Pinkas. "Privacy preserving data mining." Annual International Cryptology Conference. Springer Berlin Heidelberg,2000.

[9] Zhang, Peng, et al. "Privacy preserving naive bayes classification." International Conference on Advanced Data Mining and Applications. Springer Berlin Heidelberg, 2005.

[10] Zhang, Peng, et al. "Privacy preserving naive bayes classification." International Conference on Advanced Data Mining and Applications. Springer Berlin Heidelberg, 2005.

[11] Evfimievski, Alexandre, et al. "Privacy preserving mining of association rules." Information Systems 29.4 (2004): 343-364.

[12] Bayardo, Roberto J., and Rakesh Agrawal. "Data privacy through optimal kanonymization." 21st International Conference on Data Engineering (ICDE'05). IEEE, 2005.

[13] Hu, Haibo, et al. "Processing private queries over untrusted data cloud through privacy homomorphism." 2011 IEEE 27th International Conference on Data Engineering. IEEE, 2011.

[14] Kantarcolu, Murat, and Chris Clifton. "Privately computing a distributed k-nn classifier." European conference on principles of data mining and knowledge discovery. Springer Berlin Heidelberg, 2004.

[15] Xiong, Li, Subramanyam Chitti, and Ling Liu. "K-nearest neighbor classification across multiple private databases." Proceedings of the 15th ACM international conference on Information and knowledge management. ACM, 2006.

[16] Qi, Yinian, and Mikhail J. Atallah. "Efficient privacy preserving k-nearest neighbor search." Distributed Computing Systems, 2008. ICDCS'08. The 28th International Conference on. IEEE, 2008.

[17] Agrawal, Rakesh, et al. "Order preserving encryption for numeric data." Proceedings of the 2004 ACM SIGMOD international conference on Management of data. ACM, 2004.

[18] Hacigm, Hakan, et al. "Executing SQL over encrypted data in the database-service-provider model." Proceedings of the 2002 ACM SIG- MOD international conference on Management of data. ACM, 2002.

[19] Hore, Bijit, et al. "Secure multidimensional range queries over out- sourced data." The VLDB JournalThe International Journal on Very Large Data Bases 21.3 (2012): 333-358.

[20] Wong, Wai Kit, et al. "Secure kNN computation on encrypted databases." Proceedings of the 2009 ACM SIGMOD International Con- ference on Management of data. ACM, 2009.

[21] AL-Nabi, DL Abd, and Shereen Shukri Ahmed. "Survey on classification algorithms for data mining: comparison and evaluation." International Journal of Computer Engineering and Intelligent Systems 4.8 (2013): 18-27.

[22] Paillier, Pascal. "Public-key cryptosystems based on composite degree residuosity classes." International Conference on the Theory and Appli- cations of Cryptographic Techniques. Springer Berlin Heidelberg, 1999.

[23] Goldreich, Oded. "Foundations of cryptography": volume 2, basic applications. Cambridge university press, 2009.

[24] S. Goldwasser, S.Micali, and C. Rackoff. "The knowledge complexity of interactive proof systems. SIAM Journal of Computing". 18:186208, February 1989.

[25] Doshi, Nishant, and Devesh C. Jinwala. "A novel approach for privacy homomorphism using attribute based encryption." Security and Communication Networks 9.17 (2016): 4451-4467.

[26] Damgrd, Ivan, and Mads Jurik. "A length-flexible threshold cryptosystem with applications." Australasian Conference on Information Security and Privacy. Springer Berlin Heidelberg, 2003.

[27] S. De Capitani di Vimercati, S. Foresti, and P. Samarati. "Managing and accessing data in the cloud: Privacy risks and approaches". In 7th International Conference on Risk and Security of Internet and Systems (CRiSIS), pages 1 9, 2012.

[28] Bujlow, Tomasz, Tahir Riaz, and Jens Myrup Pedersen. "A method for classification of network traffic based on C5. 0 Machine Learning Algorithm." Computing, Networking and Communications (ICNC), 2012 International Conference on. IEEE, 2012.

[29] PANG, Su-lin, and Jizhang GONG. "C5. 0 classification algorithm and application on individual credit evaluation of banks." Systems Engineering Theory & Practice 29.12 (2009): 94-104.

[30] Sugumaran, V., V. Muralidharan, and K. I. Ramachandran. "Feature selection using decision tree and classification through proximal support vector machine for fault diagnostics of roller bearing." Mechanical systems and signal processing 21.2 (2007): 930-942.

[31] Bujlow, Tomasz, Tahir Riaz, and Jens Myrup Pedersen. "A method for classification of network traffic based on C5. 0 Machine Learning Algorithm." Computing, Networking and Communications (ICNC), 2012 International Conference on. IEEE, 2012.

[32] PANG, Su-lin, and Ji-zhang GONG. "C5. 0 classification algorithm and application on individual credit evaluation of banks." Systems Engineering Theory & Practice 29.12 (2009): 94-104.

[33] Madzarov, Gjorgji, Dejan Gjorgjevikj, and Ivan Chorbev. "A multi-class SVM classifier utilizing binary decision tree." Informatica 33.2 (2009). [34] Sugumaran, V.

[34] Muralidharan, and K. I. Ramachandran. "Feature selection using decision tree and classification through proximal support vector machine for fault diagnostics of roller bearing." Mechanical systems and signal processing 21.2 (2007): 930-942.

[35] Al-Harbi, S., et al. "Automatic Arabic text classification." (2008).

[36] Lee, Kathy, et al. "Twitter trending topic classification." 2011 IEEE 11th International Conference on Data Mining Workshops. IEEE, 2011.

[37] Javidi, Mohammad Masoud, and Ebrahim Fazlizadeh Roshan. "Speech emotion recognition by using combinations of C5. 0, neural network (NN), and support vector machines (SVM) classification methods." J. Math. Comput. Sci 6 (2013): 191-200.

[38] Agrawal, Rakesh, and Ramakrishnan Srikant. "Privacy-preserving data mining." ACM Sigmod Record. Vol. 29. No. 2. ACM, 2000.

[39] Kantarcolu, Murat, and Chris Clifton. "Privately computing a distributed k-nn classifier." European conference on principles of data mining and knowledge discovery. Springer Berlin Heidelberg, 2004.

[40] Xiong, Li, Subramanyam Chitti, and Ling Liu. "K-nearest neighbor classification across multiple private databases." Proceedings of the 15th ACM international conference on Information and knowledge manage- ment. ACM, 2006.

[41] Quinlan, J. Ross. "Improved use of continuous attributes in C4. 5." Journal of artificial intelligence research 4 (1996): 77-90.

[42] Samanthula, Bharath K., Yousef Elmehdwi, and Wei Jiang. "K-nearest neighbor classification over semantically secure encrypted relational data" IEEE transactions on Knowledge and data engineering 27.5 (2015): 1273.

[43] Lindell, Yehuda, and Benny Pinkas. "Privacy preserving data mining." Annual International Cryptology Conference. Springer Berlin Hei- delberg, 2000.

