Studies on Recent Machine Learning Approaches to Explore Performance, Security Issues and New Dimensions to Deal with the Challenges

Reshamlal Pradhan[1], S R Tandan[2]

[1,2]Dept. of CSE, Dr. C.V.Raman University, Bilaspur (C.G.) India.

ABSTRACT

In the era of Computer technology, Machine learning is centre of attraction for the researchers of data mining. Organizational and individuals information in the Computers and we are going through serious issues of threats and intrusions. Malicious activities are increased in the computer and web usage. With rapid advancement in computer technology and networking services, huge amount of data has been generating, which is difficult to handle by traditional data processing applications. Datasets in the web are composed of structure and unstructured set of data. To deal with the unstructured set of data is a prime area of attention for the researchers. There is need of advancements in the data mining and data processing techniques, to deal with this massive amount of structure and unstructured datasets. Challenges are to improve accuracy and performance of data classification, regression and clustering, to analyze and update data storage, to maintain security and information privacy. Today machine-learning techniques, which are getting key attention, are Feature reduction, Decision tree techniques, Ensemble techniques, Neural Networks, Statistical techniques, Genetic algorithm, Fuzzy logic and big data analytics. In this paper, we are trying to gain attention on some of recent work done in these fields to explore data processing, data analysis and security challenges issues.

Keywords: Data Mining, IDS, Big data, Ensemble techniques.

I INTRODUCTION

Data mining has attracted more attention in recent years as scientific organizations and business organizations are dealing with very huge amount of data. Probably Big Data is the key attention of data mining in current era. Challenges in big data are extracting data, data storage and analysis, searching, querying and updating data, information privacy. Data mining is the process of discovering knowledge and interesting patterns from large amount of data. Today Intrusion detection system (IDS) is a necessary addition to the security infrastructure of most organizations. IDS collect and analyses information from different areas within a computer or network to detect possible security violations defined as attempts to compromise the confidentiality, integrity, availability, or to bypass the security mechanisms of a computer or network.

Domain fields of data mining are business intelligence, scientific discovery, Web search and digital libraries etc. Big data has become crucial for numerous application domains as it deals with the large amount of unstructured data. Rapid growth of cloud services are the reason behind popularity of Big Data.

(a) Process of KDD: Data mining is often referred to as Knowledge discovery of data, which highlights the goal of mining process. To extract knowledge from data following steps are performed in KDD:

Step 1. Data preprocessing
Step 2. Data transformation
Step 3. Data mining
Step 4. Pattern evaluation and presentation.

(b) Types of IDS

Different types of Intrusion detection system are:

(i) NIDS and HIDS: Host based Intrusion detection system (HIDS) monitors only individual workstation or system. HIDS are unaffected by switched network. HIDS can be thought of as an agent which monitors whether anyone or anywhere any unusual or subspecies activity is done. Network based intrusion detection system (NIDS) on the other hand monitors network traffic for particular network segment or device and analysis the network and application protocol activity to identify any sign of suspicious activity [1].

(ii) Misuse and anomaly detection based IDS: Misuse detection or signature based detection technique uses the previous data or pattern for detection; if previous pattern or signature is not available it cannot detect the new attack. Anomaly detection is adaptive in nature. They attempt to identify behaviors that do not conform to normal behavior [1, 2].

(c) Data Mining Techniques: There are varieties of data mining techniques. Using data processing techniques, it perceives and extrapolates knowledge that may scale back the probabilities of fraud detection [5]. These techniques are used for knowledge discovery and pattern recognization in order to detect intrusions and extract information.

(i) Genetic algorithm: Genetic Algorithm is an adaptive search technique initially introduced by Holland [7]. Genetic algorithm operates on a set of individuals called population, where each individual is an encoding of the problem input data and are called chromosomes. The search for best solution is guided by an objective function called fitness function. The selected solution of fitness function replace those of less function ,as they are able to produce new solution that are more fitted in the environment. Fitness function controls the selection of best solution and provides criteria to evaluate the candidate individuals [8].

(ii) **Decision tree:** Decision trees are unit arborous structures that represent decision sets. These choices generate rules that are used to classify data [6]. A decision tree classifies a sample through a sequence of decisions, in which the current decision helps to make the subsequent decision. Such a sequence of decision is represented in a tree structure. The classification of a sample proceeds from the root node to a suitable end life node, where each end life node represents a classification category. The attributes of the samples are assigned to each node, and the value of each branch is corresponding to the attributes [9].Some of decision tree techniques are CHAID, CART, ID3 etc.

(iii) **Artificial Neural Network:** Artificial Neural Network is unit non-liner predictive models that learn through training. Though there are powerful predictive modeling techniques. The auditors simply use them is reviewing records to spot fraud and fraud-like actions, they are higher utilized in things wherever they will be used and reused, like reviewing MasterCard transactions each month to envision for anomalies [6].

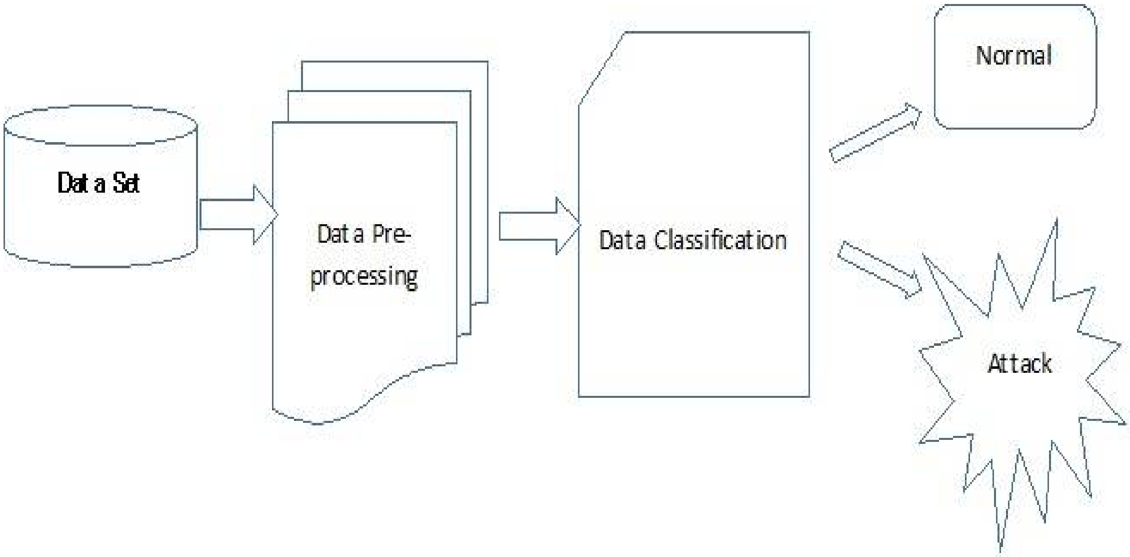

Fig. 1 Classification as IDS technique

(iv) **Statistical Techniques:** Statistical models [22,23] involving a latent structure often sup-port clustering, classification, and other data mining tasks. Because of their ability to deal with minimal information and noisy labels in a systematic fashion, statistical models of this sort have recently gained popularity. Statistical techniques are Bayesian net, support vector machine etc.

(v) **Ensemble Classifier:** The idea behind a Ensemble classifier[24,25] is to combine several machine learning techniques so that the system performance can be significantly improved. Ensemble model is combination of two or more IDS techniques; here combination of techniques is done to detect the attack, increase to avoid the drawbacks of individual models and to achieve high accuracy [26]. Ensemble techniques are Bagging, Boosting and Stacking.

II RELATED WORKS

Data mining is one of the key research area in the field of computational science. There are varieties of data mining techniques used for knowledge discovery. Mining techniques of key attention in current time are ensemble techniques, Genetic algorithm, Fuzzy logic and big data analytics. Data processing of Big data encounters many challenges. To explore data processing and security challenges, novel data processing and analysis techniques, some of recent works in these fields are:

(a) Jemal H. Abawajy, Andrei KelREV, Morshed Chowdhury have worked on "Large iterative Multitier Ensemble Classifiers for security of Big data"[4]. The paper introduces and investigates large iterative multitier ensemble classifiers (LIME) for big data. They are generated automatically as a result of several iterations in applying ensemble meta classifiers. They incorporate diverse ensemble meta classifiers into several tiers simultaneously and integrate them into one iterative system. The four tier LIME classifiers based on Random forest achieved better performance compared with the base classifiers. The four tier LIME classifiers based on Multiboost at the fourth tier, decorate at third tier and bagging at second tier obtained the best outcome with AUC 0.998.

(b) Peiying Tao , Zhe Sun, And Zhixin Sun have worked on "An Improved Intrusion Detection Algorithm Based on GA And SVM"[13]. The Author proposed an alarm intrusion detection algorithm (FWP-SVM-GA) based on the genetic algorithm (GA) and support vector machine (SVM) algorithm for use in human centred smart IDS. First, this paper makes effective use of the GA population search strategy and the capability of information exchange between individuals by optimizing the crossover probability and mutation probability of GA. The convergence of the algorithm is accelerated, and the training speed of the SVM is improved. A new fitness function is proposed that can decrease the SVM error rate and increase the true positive rate. Simulation and experimental results show that the improved intrusion detection technology based on the genetic algorithm (GA) and support vector machine (SVM) proposed in this paper increases the intrusion detection rate, accuracy rate and true positive rate; decreases the false positive rate; and reduces the SVM training time.

(c) Abdel-Rahman Hedar, Mohamed A. Omer, Ahmad F. AI-Sadek And Adel A. Sewisy, proposed "Hybrid Evolutionary Algorithms For Data Classification in Intrusion Detection System"[03]. This research work is based on classification attack for Intrusion detection system. AGAAR is used to reduce features. A Classifier model built by GPLS is used to classify attacks in the NSL-KDD. The classifier trained with the full feature of 20%-NSL-KDD. The classifier was trained with 19.51% of the dataset features. The classifier accuracy improves the result after reducing dimensionality of the dataset. This reduction makes a significant improvement in term of memory and CPU time. This shows that AGAAR can remove the relevant features in intrusion detection. The experiment shows that the GPLS using reduced features is more accurate than that uses all of the features. These classifiers (AGAAR-GPLS) compared with others methods, show better results than many methods. The classification rate increased from 75.98% to 81.44% after reducing the features. In few cases, the results were close for few methods. The reduction of the dataset leads to minimizing the modeling time and computational costs.

(d) Mohammad Saniee Abadeh, Hamid Mohamadi and Jafar Habibi, proposed "Design and analysis of genetic fuzzy systems for intrusion detection in computer networks"[29]. In this paper, the use of different GFS (genetic fuzzy system) approaches is investigated to develop an intrusion detection system capable of detecting intrusive behaviors in a computer network. The characteristic features of the proposed GFSs can be summarized as follows:

(i) As intrusion detection is a high-dimensional classification problem one of the important properties of the proposed GFSs in this paper is that the class labels of all of the rules in the population and in each rule set (in Pittsburgh approach) are the same. This feature allows the algorithm to focus on learning of each class independently. Therefore the genetic fuzzy rule generation algorithm is repeated for each of the classes in the classification problem.

(ii) An initialization procedure is used to generate fuzzy if-then rules directly from the training data set. These rules enable the algorithm to focus on finding fuzzy rules, which are related to a special class. A same procedure is used to reinitialize the population at the end of each generation of the Michigan based GFS.

(iii) The genetic operators (i.e., crossover, and mutation) of the Michigan approach based GFS guaranteed to generate valid individuals. To achieve this, after performing the operator, consequent class of the generated individual is determined. If this class is the same as the parent class then the generated individual is accepted, otherwise the operator is repeated.

(e) Bolón-Canedo, N. Sánchez-Maroño and A. Alonso-Betanzos, proposed "Feature selection and classification in multiple class datasets: An application to KDD Cup 99 dataset"[30]. This paper proposes a method based on the combination of discretization, filtering and classification methods that maintains the performance results of the classifiers but using a reduced set of features. Specifically, it has been applied over the KDD Cup 99 dataset, a benchmark in the intrusion detection field. The proposed method, is based on classifiers like naive Bayes or C4.5, is applicable to large databases, since this machine learning algorithms have the advantages of being faster and more computational efficient than other classifiers used by other authors of the literature, such as SVM's, multilayer perceptions or functional networks. The comparative study denotes that the proposed method achieved a better performance than the other authors' results, specifically in those classes more difficult to detect.

(f) **Mr. Vijay D. Katkar, Mr. Siddhant Vijay Kulkarni, worked on "Experiments on Detection of Denial of Service Attacks using Ensemble of Classifiers"[31].**The classifiers used are Naive Bayesian(NB), Bayesian Network(BN), Sequential Minimal Optimization(SMO), J48(C4.5) and Reduced Error Pruning Tree(REPTree). Result shows that Ensemble of Reduced Error Pruning Tree, Bayesian Network and J48 classifiers with no data pre-processing provided significant accuracy increment with minimal resource requirement. It is also proven that instead of developing a new classifier, one can achieve extremely high accuracy by using Ensemble of these multi-category classifiers.

(g) **Fatma Gumus, C. Okan Sakar, Zeki Erdem and Olcay Kursun, proposed "Online Naive Bayes Classification for Network Intrusion Detection" [32].** The proposed online naive Bayes classification method is first tested on the well-known Fisher-iris dataset which is available at UCI machine learning repository. The dataset has 150 samples, 3 classes, and each instance is represented with 4 features. The dimensionality of the data is reduced to two using the principal component analysis (PCA) to enable visualization on two axes. This method is based on the idea that weighting the most recent samples more allows the classifier to adapt to the recent attacks which may develop over time. The proposed method is time efficient compared to k-NN and more accurate than the linear perceptron.

(h) **Winston et. al. proposed "A Novel Technique to Detect DDoS and Sniffers in Smart Grid"[12].** In this paper, an IDS technique called double layer protection method is used for detection and isolation of sniffers, in first layer detection MD-5 technique is used and in the second layer detection PMD technique is used. DDoS attacks are detected using TTL analysis technique. Tools used for this detection purpose are NS-2 and CISCO. MD-5 safeguards the data integrity by encryption and decryption technique. PMD helps to find the source of the sniffing packets. NS-2 and network analyzer are the tools used for result comparison. Using these techniques and some features it gives the high efficiency.

(i) **Adhikari et. al. proposed Developing a "Hybrid Intrusion Detection System Using Data Mining for Power Systems"[14].** The IDS was trained an evaluated for a three-bus two-line transmission system which implements a two zone distance protection scheme. Twenty five scenarios consisting of stocktickerSLG faults, control actions, and cyber-attacks were implemented on a hardware-in-the-loop test bed. Scenarios were run in a loop 10000 times with randomized system parameters to create a dataset for IDS training and evaluation. The IDS correctly classified 90.4% of tested scenario instances. Evaluation also included a tenfold cross validation to evaluate the detection accuracy of zero-day attack scenarios. The average detection accuracy for zero-day attack scenarios was 73.43%. The common paths mining-based IDS out performs traditional machine learning algorithms and is better suited for the high volume of data present in power system.

(j) **Jaiyen et. al. proposed "Intrusion Detection Model Based on Ensemble Learning for U2R and R2L Attacks"[18].** In this paper, an algorithm is used for increasing the accuracy, and decreases the false alarm rate of U2R and R2L attacks by using Correlation-based feature selection and multiple weak classifiers such as Naïve Bayes, Decision Tree, MLP, k-NN and SVM based on Adaboost algorithm.

(k) **Lei xu, Chunxiao Jiang, Jian wang, Jian yuan and Yong ren have an article on "Information security in Big data: Privacy and Data Mining"[8].** In this paper they have highlights an emerging research topic in data mining known as privacy preserving data mining (PPDM). It focused on how to reduce the privacy risk brought by data mining operations. It reviews the privacy issues related to data mining by using a user role based methodology. Particularly four different types of users are identified which are involved in data mining applications, namely data provider, data collector, data miner, and decision maker. Each user role has its own privacy concerns. Hence the privacy preserving approach adopted by one user role different from those adopted by others.

(l) **Yenduri et. al. proposed "Analyzing Intrusion Detection System: An Ensemble based Stacking Approach"[19].** In this paper, an intrusion detection system uses stacking classifier which has detected different types of intrusions, for achieving good accuracy, precision, recall and ROC values. KDD cup99 dataset and weka data mining tool is used. Accuracy rate is 82.7206%.

(m) **Kulkarni et. al. worked on "Experiments on Detection of Denial of service Attacks using Ensemble of Classifiers"[20].** In this paper J48, Naive Bayesian classifiers and KDD 99 as a dataset are used. Ensemble of Naive Bayesian, J48 and Sequential Minimal Optimization classifiers, when combined with Numeric to Binary Data Pre-processing method provides maximum accuracy of 99.89785%. The same accuracy is also provided by the ensemble of Bayesian network, J48 and Sequential Minimal Optimization.

(n) **Pradhan et. al. worked "Performance Assessment of Robust Ensemble Model for Intrusion Detection using Decision Tree Techniques"[21].** In this paper J48, robust forest as a data mining technique and ensemble classifiers (bagging, boosting, and stacking) are used. Dataset which is used here is NSL-KDD and weka tool. The experiment results shows that Bagging classifiers provides highest accuracy 98.71%, stacking provides accuracy of 98.66% and boosting provides accuracy of 98.60% which is better than the accuracy of individual classifier J48 and Random forest. Not only in accuracy, in precision, have recall and f-measure also had ensemble techniques provided better results than individual classifiers.

(o) **Mr. Vijay D. Katkar, Mr. Siddhant Vijay Kulkarni, worked on "Experiments on Detection of Denial of Service Attacks using Naïve Bayesian Classifier"[27].** This paper evaluates variation in performance of Naive Bayesian classifier for intrusion detection when used in combination with different data pre-processing and feature selection methods. Naive Bayesian classifier performed significantly better when combined with Numeric to Binary data pre-preprocessing. It can be also observed that, instead of going for an improved version of Naive Bayesian classifier or completely different set of multi-classifiers, one can achieve better performance using Naive Bayesian classifier along with Numeric to Binary data pre-processing. Experimental results prove that accuracy of Naive Bayesian classifier is improved and performs better than other classifiers when used in combination with Feature Selection and data pre-processing methods.

III PROBLEM IDENTIFICATION AND FINDINGS

Above literature review explores data processing and security challenges, novel data processing and analysis techniques to deal with security challenges. With the study of these research works the advantage and disadvantage of these literatures are summaries. It is observed that hybrid models, accuracy percentage is high from the base classifier. The author [12] uses MD-5, PMD and TTL techniques for detection of DDoS and Sniffers in smart grid. MD-5 safeguards the data integrity by using cryptography techniques. Detection of sniffers is done using very less bandwidth. Author [4] has uses large iterative multitier ensemble classifier for security of big data. With random forest, adaboost and bagging in Meta classifier in multitier, it results in high accuracy. Author [13] stated with genetic algorithm convergence of the algorithm is accelerated, and the training speed of the SVM is improved. A new fitness function decreases the SVM error rate and increases the true positive rate. The hybrid of cryptographic methods reduces the overhead, EAACK mechanism is used for this. The hybrid of GPLS and AGAAR has the classification rate 81.44%. If both of the models are used as a single classifier the rate will be low. AGAAR is used to reduce the features from the dataset; it removes the relevant features in intrusion detection. The hybrid of hardware and software based model which is used for grid, is mainly apply on high volume of data's. Author [18] uses Adaboost algorithm to create the ensemble of Decision Tree, Naïve Bayes, SVM, and MLP classifiers for detecting U2R and R2L attacks which are difficult to detect. The hybrid of Naive Bayes and MLP produces the highest sensitivity, Decision tree results the least performance. Author [19] uses stacking based ensemble classifier technique which provides the efficient result, accuracy rate is high and dataset which is used here is easily available. The multi-layer hybrid technique uses PCA for feature selection and comparison purpose. Classifiers uses here are fast and highly independent. The hybrid of J48 and Naive Bayesian network with no data pre- processing provides accuracy and less resource requirement. Author [29] investigated the use of different GFS (genetic fuzzy system) approaches to develop an intrusion detection system capable of detecting intrusive behaviors in a computer network.

IV CONCLUSION

In this paper, a literature study of recent work in the field of data mining is presented. In which different hybrid classifiers are analyzed based on their performance and result. It explored data processing and security challenges, novel data processing and analysis techniques to deal with security challenges. We see some of the single classifiers with their drawbacks and strengths. There is a need of hybrid system for better result. It is not necessary that every single model has some drawback but it is observed that hybrid models give better results and performance. Many combinations of machine learning techniques are tested on hybrid models and still there are provisions for different combination of machine learning techniques which can be tested on ensemble (hybrid) models for better result. It is well known that Challenges in big data are extracting data, data storage and analysis, searching, querying and updating data, information privacy. Various dimensions to deal with these issues are ensemble techniques, Genetic algorithm, Fuzzy logic and big data analytics, which are playing key role in the security concern of Big data and data mining.

REFERENCES

[1] Farid Lawan Bello,Kiran Ravulakollu, Amrita (2015) “Analysis and Evaluation of Hybrid Intrusion Detection System models”, in international conference on computers, communication and systems.

[2] James P Anderson “Computer security threat monitoring and surveillance” in technical report, James P Anderson company, fort woshington, Pennsylvania.1980.

[3] Abdel-Rahman Hedar, Mohamed A. Omer and Ahmed F. Al-Sadek, Adel A. Sewisy,(2015) “hybrid evolution algorithm for data classification in IDS”, in IEEE SNPD 2015, June 1-3 2015, Takamatsu, Japan.

[4] Jemal H. Abawajy, Andrei KelREV, Morshed Chowdhury (2014)“ Large iterative Multitier Ensemble Classifiers for security of Big data”, in IEEE Transactions on EMERGING TOPICS IN COMPUTING, Volume 2, No. 3, September 2014.

[5] Lei Li,De-Zhang Yang, Fang-Cheng Shen (2010)“A Novel Rule Based Intrusion detection system Using Data Mining” in the Proc . Of 3rd IEEE International conference on computer science and information technology, pp. 169-172,2010.

[6] S.Revathi, Dr. A .Malathi (2013)“A detailed analysis on NSL-KDD Dataset using various machine learning techniques for intrusion detection”, in International Journal of Engineering &Technology (IJERT),ISSN: 2278-0181,vol. 2 issue 12. December-2013.

[7] DE Goldberg and H Holland, Genetic algorithms and machine learning, Machine learning, Vol. 3(2), 95-99,1988.

[8] Lei xu, Chunxiao Jiang, Jian wang, Jian yuan , Yong ren (2014)“Information security in Big data: Privacy and Data Mining” in IEEE access, The journal for rapid open access publishing, Volume 2, 2014.

[9] Chih-Fong Tsai a, Yu-Feng Hsu b, Chia-Ying Lin c, Wei-Yang Lin d,(2009) “Intrusion detection by machine learning: A review” in Expert Systems with Applications, Elsevier, 36 (2009) 11994–12000, 0957-4174/ 2009.

[10] Amrita anand, Brajesh Patel (2012)“An Overview on Intrusion Detection System and Types of Attacks It Can Detect Considering Different Protocol” in International Journal of Advanced Research in Computer Science and Software Engineering. Volume 2, Issue 8, August 2012.

[11] Sanjiban Sekhar Roy, P Venkata Krishna, Sumanth Yenduri, (2014)“Analyzing Intrusion Detection System: An Ensemble based Stacking Approach”, 978-1-4799-1812-6/14 ©2014 IEEE.

[12] S.Shitharth, Dr.D.Prince Winston,(2016) “A Novel IDS Technique to Detect DDoS and Sniffers in Smart Grid” in 2016 world conference on futuristic trends in research and innovation for social welfare.

[13] Peiying Tao , Zhe Sun, And Zhixin Sun (2018) “An Improved Intrusion Detection Algorithm Based On GA and SVM” in Special section on human-centered smart systems and technologies, IEEE ACCESS, volume 6, 2018.

[14] Shengyi Pan, Thomas Morris, Uttam Adhikari (2015) “Developing a Hybrid Intrusion Detection System Using Data Mining for Power Systems” in IEEE TRANSACTIONS ON SMART GRID, VOL. 6, NO. 6, NOVEMBER 2015.

[15] M. Revathi, T. Ramesh, “Network intrusion detection system using reduced dimensionality” in Indian Journal of Computer Science and Engineering (IJCSE), ISSN: 0976-5166, Vol. 2 No. 1 pp. 61 -67.96.

[16] Mohammed A. Ambusaidi, Xiangjian He, Priyadarsi Nanda, Zhiyuan Tan (2016)“Building an Intrusion Detection System Using a Filter-Based Feature Selection Algorithm” in IEEE TRANSACTIONS ON COMPUTERS, VOL. 65, NO. 10, OCTOBER 2016.

[17] W. Feng, Q. Zhang, G. Hu, J. X. Huang, “Mining network data forIntrusion detection through combining SVMs with Ant colony networks” in Elsevier, Future Generation Computer Systems 37(2014) 127 – 140.

[18] Ployphan Sornsuwit, SaichonJaiyen, (2015)“Intrusion Detection Model Based on Ensemble Learning for U2R and R2L Attacks” in 2015 7th International Conference on Information Technology and Electrical Engineering (ICITEE),chiangmai , Thailand.

[19] Sanjiban Sekhar Roy, P Venkata Krishna, Sumanth Yenduri (2014)"Analyzing Intrusion Detection System: An Ensemble based Stacking Approach" in 978-1-4799-1812-6/14 ©2014 IEEE.

[20] Mr. Vijay D.Katkar, Mr.Siddhant Vijay Kulkarni,(2013) "Experiments on Detection of Denial of Service Attacks using Ensemble Classifier" in 2013 International Conference on Green Computing, Communication and Conservation of Energy (ICGCE), 978-1-4673-6126-2/13/ 2013 IEEE.

[21] Reshamlal Pradhan, Deepak Kumar Xaxa,(2014)" Performance Assessment of Robust Ensemble Model for Intrusion Detection using Decision Tree Techniques", in International Journal of Innovations & Advancement in Computer Science IJIACS ISSN 2347 – 8616 Volume 3, Issue 3 May 2014.

[22] Arun K. Pujari. (2001), Data mining techniques, 4th edition, Universities Press (India) Private Limited.

[23] Jiawei Han, Micheline Kamber, (2006), "Data mining concepts and tech-niques", Second edition, San Francisco, Margan Kaufmann Publishers, USA.

[24] Manish Kumar Nagle, Dr. Setu Kumar Chaturvedi (2013)"Feature Extraction Based Classification Technique for Intrusion Detection System" in International Journal of Engineering Research and Development (August 2013).

[25] Mrutyunjaya Panda, (2011) "A hybrid intelligent approach for network intrusion detection", Proceedia Engineering.

[26] Anup Ashok Patil, ShitalMali (2016)"Hybrid Cryptography Mechanism for SecuringSelf-Organized Wireless Networks" in 2016 3rd International Conference on Advanced Computing and Communication Systems (*ICACCS* -2016), Jan. 22 – 23, 2016, Coimbatore, INDIA.

[27] Mr. Vijay D. Katkar, Mr. Siddhant Vijay Kulkarni, (2013)"Experiments on Detection of Denial of Service Attacks using Naïve Bayesian Classifier" in International Conference on Green Computing, Communication and Conservation of Energy (ICGCE), IEEE(2013).

[28] Reshamlal Pradhan, Deepak Kumar Xaxa,(2014) "Robust Ensemble Model for Intrusion Detection using Data Mining Techniques" in International Journal of Scientific & Engineering Research, Volume 5, Issue 4, April-2014 781 ISSN 2229-5518.

[29] Mohammad Saniee Abadeh, Hamid Mohamadi, Jafar Habibi (2011)"Design and analysis of genetic fuzzy systems for intrusion detection in computer networks" in Expert Systems with Applications 38 (2011) 7067–7075, Elsevier.

[30] V. Bolón-Canedo, N.Sánchez-Maroño, A. Alonso-Betanzos, (2011) "Feature selection and classification in multiple class datasets: An application to KDD Cup 99 dataset" in Expert Systems with Applications 38 (2011) 5947–5957, Elsevier.

[31] Mr. Vijay D. Katkar, Mr. Siddhant Vijay Kulkarni, (2013)" Experiments on Detection of Denial of Service Attacks using Ensemble of Classifiers" in International Conference on Green Computing, Communication and Conservation of Energy (ICGCE), IEEE(2013).

[32] Fatma Gumus, C. Okan Sakar, Zeki Erdem, Olcay Kursun (2014)"Online Naive Bayes Classification for Network Intrusion Detection" [32] in 2014 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2014).

[33] Amira Sayed A. Aziz, Aboul Ella Hassanien, Sanaa El-Ola Hanafy, M.F.Tolba (2013)" Multi-layer hybrid machine learning techniques for anomalies detection and classification approach", 978-1-4799-2439-4/13/ ©2013 IEEE.