International Journal on Recent and Innovation Trends in Computing and Communication ISSN: 2321-8169
Volume: 2 Issue: 10 3253 – 3257

Study of Tree Base Data Mining Algorithms for Network Intrusion Detection

Kailas Shivshankar Elekar
M.E. (IT) 2nd Year, Department of Computer Engineering
Dattakala faculty of Engineering,
Swami Chincholi, Daund, Pune, India
ekailas@gmail.com

Prof. M.M. Waghmare
Department of Computer Engineering
Dattakala faculty of Engineering,
Swami Chincholi, Daund, Pune, India
monica.waghmare@gmail.com

***Abstract*—**Internet growth has increased rapidly due to which number of network attacks have been increased. This emphasis importance of network intrusion detection systems (IDS) for securing the network. It is the process of monitoring and analyzing network traffic for detecting security violations many researcher suggested data mining technique such as classification, clustering ,pattern matching and rule induction for developing an effective intrusion detection system. In order to detect the intrusion, the network traffic can be classified into normal and anomalous. In this paper we have evaluated tree base classification algorithms namely J48, Hoeffding tree, Random Forest, Random Tree, REPTree. The comparison of these tree based classification algorithms is presented in this paper based upon their performance metrics using 10 fold cross validation and KDD- CUP test dataset. This study shows that random forest and J48 are the best suitable tree base algorithms.

Keywords- *Classification; Data Mining; Intrusion Detection; KDD CUP dataset; Network Security;IDS; J48;Hoeffding tree; Random Forest; Random Tree; REPTree; WEKA*

*****

I. INTRODUCTION

Sophisticated hacking attacks are continuously increasing in the internet. In the past hacking personal information, accessing resources is done for just fame, but now a days hacking targets companies, government agencies for various reasons such as profit, protest, challenge or enjoyment. The new kind of attack Advanced Persistent Threat (APT) targets a specific system and analyses vulnerabilities of the system for a long time. APT usually targets organizations or nations for business or political motives. Due to which it is very difficult to detect and prevent APT attacks compared to traditional attacks and could result massive damage.[1][2]

Several techniques, methodologies, policies, and systems have been proposed to alleviate the threat and attacks through the intrusion detection systems (IDS). IDSs can monitor events at the endpoints on the network or host to detect anomaly. There are basically two approaches for IDS , first approach is based on signature matching while the second is to detect the anomaly behavior from the network. Each has its own strength and weakness in order to cope with both known and unknown attacks in an efficient way with high detection precision and speed.

The rest of this paper is organized as follows: section II describes related work. Section III describe data mining, KDD CUP 1999 dataset and WEKA. Section IV shows result of experiments, In Section V results analysis , and we conclude with conclusion and Future work in Section VI.

II. RELATED WORK

The KDD Cup '99 dataset is the most well-known intrusion detection dataset available and researched by many researchers. The network traffic records in the dataset are classified as Normal or one of the four attack types i.e. DOS - denial of service , PROBE - network probe , R2L - remote to local and U2R- user to root attacks. In past various static machine learning algorithms have been evaluated and results are published.

The results of the KDD'99 classifier learning contest, as summarized by Elkan [3], were all variants of the C5 decision tree algorithm (see Quinlan [4]). After the contest a comprehensive set of other algorithms were tested on the KDD Cup 99 data, mostly with comparable results, were presented by Sabhani and Serpen [5], Sung and Mukkamala [6], Chavan, Shah et al. [7] and Peddabachigari, Abramham et al. [8]. The majority of results published are on the KDD Cup '99 `10%' training set only see Sung and Mukkamala [6], Kayacik, Zincir-Heywood et al. [9] and Lee, Shin et al. [10].

Some of the researchers extracted 11,982 records from KDD Cup 10% training dataset and build custom training datasets with 5,092 records and 6,890 test record see Chavan, Shah et al. [7], Chebrolu, Abraham et al. [11] and Chen, Abraham et al. [12].

Chavan, Shah et al. [7] use a decision tree method for ranking of features per class. They reduced number of features from 41 to 13 for 'normal',16 for `probe', 14 for `dos', 15 for `u2r' and 17 for `r2l' for experiment they evaluated it using artificial neural networks and fuzzy inference systems.

Kayacik, Zincir-Heywood et al. [9] investigated the relevance of each feature provided by the KDD Cup '99 intrusion detection dataset in terms of information gain and presented the most relevant feature for each individual attack. Another important result was that 9 features do not make any contribution for intrusion detection.

Tavallaee and Bagheri et al. [15] described the importance of each feature in KDD '99 intrusion detection dataset for detection of DOS, PROBE, U2R L2R and Normal class. They also discuss various problems of KDD Cup '99 datasets and created a revised version of the datasets, called NSL-KDD to address the some of known issues. They modified the class distributions by cleaning the training and testing datasets. This will avoid biasness towards the more frequent records.

Ben Amor et al. [16] performed comparative analysis of decision tree vs naïve bayes and found that decision tree performs slightly better than naïve bayes. They also found that building naïve bayes computational model is faster than of decision tree. Decision trees generally have very high speed of operation and high attack detection accuracy.

Gary Stein et al. [17] suggest that all 41 features are not required for detecting four types of attack i.e. Probe, DOS,U2R and R2L. They performed experiment for each of the above four categories of attack separately using Genetic Algorithm. They found that GA made drastic performance improvements in Probe category attacks. However, performance improvement on R2L and U2R are limited. One of the reason is that proportions of R2L and U2R are very less in the training data compare to testing data.

Nadiammai et al [18] presented a comparative study of attack predication based on accuracy, sensitivity, specificity, time and error using rule based and some function based classifiers.

Hwang et al [19] presented three-tier architecture of intrusion detection system based on traffic filtering such as white list traffic as normal, blacklist traffic as known attacks and rest of the traffic as anomalies using multi-class SVM classifier.

Reddy et al [20] also presented a survey of various data mining techniques for intrusion detection system.

Subramanian et al [21] presented the performance study of decision tree algorithms using NSL-KDD[22] dataset.

Das et al [23] also presented a comparison of the various data mining classification techniques for intrusion detection.

Nagaraju et al [24] evaluated the performance different data mining classification techniques namely CART, Naive Bayesian, and Artificial Neural Network Model classifier using a confusion matrix.

Chakchai So et al [41] performed intrusion detection analysis using both KDD CUP dataset and recent HTTP BOTNET attacks on Decision Tree, Ripper Rule, Neural Networks, Naïve Bayes, k–Nearest–Neighbour, and Support Vector Machine classifiers. They evaluated the performance using standard cross–validation and confusion matrix.

III. DATA MINING, KDD CUP 1999 DATASET AND WEKA DESCRIPTIONS

A. Data mining

Data mining (also known as Knowledge Discovery in Databases - KDD) generally refers to the process of extracting descriptive models from huge amount of data. Frawley[27] defined data mining as "The nontrivial extraction of implicit, previously unknown, and potentially useful information from data". Data mining commonly involves five classes of tasks namely clustering, classification, regression, association rule learning and visualization. The main aim of data mining is forecasting [40]. Data mining-based IDSs require less expert knowledge (only need to label the traffic data to indicate intrusions instead of hand-coding rules) yet provide good performance.

Classification is one of the most commonly used data mining technique. The goal of classification is to build a model/classifier from classified objects in order to classify previously unseen objects as accurately as possible. Depending on the information available on classes and the type of classification, the output of a "classifier" can be presented in different forms, for example in the form of decision trees or rules [37-39]. For analyzing the data and classification of network attacks, five different tree base classification algorithms such as Hoeffding tree, J48, Random Forests, Random Tree, REPTree are studied and evaluated.

Hoeffding tree[28]:A Hoeffding tree (Very Fast decision Tree i.e. VFDT) is based on hoeffding algorithm for building decision tree from real time rapid data feeds. It is one pass algorithm It uses no of concepts and its results are approximate. The main advantage of Hoeffding tree algorithm is it consumes less memory. It uses an incremental approach for new samples. It consumes more memory as tree grows and they waste computational time in checking ties [31].

J48:Java implementation for generating a pruned or unpruned C4.5[29] decision tree[31].

Random Forests[8]: This algorithm was developed by Leo Breiman and Adele Cutler. It is an ensemble learning method for classification and regression. It construct a number of decision trees (CART) at training time and they are not influenced by each other. While predication it sums all predication made by all decision trees. It is best suited for the analysis of complex data structures having large column data with small to moderate data sets [42] .

Random Tree: It construct tree using K randomly chosen attributes at each node. After tree is constructed it does not perform pruning. Also it has s an option to estimate of class probabilities based on a hold-out set [31].

REPTree: This algorithm builds decision or regression tree using information gain or variance. Decision or regression tree is pruned using reduced-error pruning. Sorting only once on numeric attributes values. Splitting the corresponding instances into pieces for Missing values [31].

B. KDD CUP 1999 Dataset

In general, KDD CUP 1999 is based on the intrusion detection simulation of U.S. Air force local area networks via tcpdump [www.tcpdump.org]. The dataset consists of network access behavior including up to 41 attributes as well as heterogeneous access patterns.

In general, KDD CUP consists of four main types attacks categories[32] as given below;

DoS (Denial of Service): This attack can freeze the server operation and activity by acquiring all resources so that the server cannot provide any service, commonly using flooding–based schemes.

PROBE: This attack is used during a preparation stage for other attacks in order to gain valuable information such as enabled ports and services as well as Internet address information.

U2R (User to Root): This attack performs a specific operation in order to penetrate into a system hole/leak such as Buffer Overflow.

R2L(Remote to User): The attack is used to take advantages of related users' safety information or configuration such as SQL Injection.

The details of these attack types in data set are shown in Table I

TABLE I. ATTACK CATEGORY

Type	Attack
DoS	apache2, smurf, neptune, dosnuke, land,pod, back, teardrop, tcpreset, syslogd, crashiis, arppoison,mailbomb, selfping, processtable, udpstorm, warezclient
PROBE	portsweep, ipsweep, queso, satan, msscan, ntinfoscan, lsdomain, illegal-sniffer
U2R	sechole, xterm, eject, ps, nukepw, secret, perl, yaga, fdformat, ffbconfig, casesen, ntfsdos, ppmacro, loadmodule, sqlattack
R2L	dict, netcat, sendmail, imap, ncftp, xlock, xsnoop, sshtrojan, framespoof, ppmacro, guest, netbus, snmpget, ftpwrite, httptunnel, phf, named

TABLE II. NO OF ATTACKS IN DATASET

Type	Training	Test
DoS	395176	229853
PROBE	4107	4166
U2R	58	228
R2L	1125	16189
NORMAL	112332	60593
TOTAL	512798	311029

C. *Waikato Environment for Knowledge Analysis(WEKA)*

It is a collection of machine learning algorithms developed in Java for data mining tasks developed by Machine Learning Group at the University of Waikato New Zealand. The algorithms can either be applied directly to a dataset or called from your own Java code. It contains various tools for data mining activities like data pre-processing, classification, regression, clustering, association rules, and visualization. It is easy to develop new machine learning schemes using this tool. It consists of Explorer, Experimenter, Knowledge flow, Simple Command Line Interface, Java interface [31].

IV. EXPERIMENTS AND RESULTS

In order to evaluate the performance of tree base classification techniques, KDD CUP 1999 10% train dataset [32] available at University of California web site (http://kdd.ics.uci.edu/databases/kddcup99/) was used. All classification models were generated using WEKA 3.7.11 on Intel Dual Core with 4 GB RAM machine with Windows 7 operating system. Their performance is evaluated using 10 fold cross validation and test dataset.

The experiments were performed on full 10% training dataset having 512798 records and test data set having 311029 records.

Performance of all the classifiers is compared based upon accuracy of attack detection and instance predication provided by confusion matrix. The obtained results are shown below

A. *Using Cross Validation*

TABLE III. PERFORMANCE METRICS

Classifiers	Classified Instances	
	Correctly	***Incorrectly***
HoeffdingTree	99.8528	0.1472
J48	99.9684	0.0316
RandomForest	**99.9789**	**0.0211**
RandomTree	99.9602	0.0398
REPTree	99.9524	0.0476

TABLE IV. ATTACK DETECTED

Classifiers	DOS		PROBE	
	Correct	***False + ve***	***Correct***	***False + ve***
HoeffdingTree	395063	113	3944	163
J48	395156	20	***4081***	***26***
RandomForest	**395174**	**2**	4066	41
RandomTree	395151	25	4048	59
REPTree	395141	35	4012	95

Classifiers	U2R		R2L	
	Correct	***False + ve***	***Correct***	***False + ve***
HoeffdingTree	40	18	1038	87
J48	32	26	1078	47
RandomForest	**41**	**17**	**1103**	**22**

Classifiers	U2R		R2L	
	Correct	***False + ve***	***Correct***	***False + ve***
RandomTree	37	21	1100	25
REPTree	28	30	1090	35

Classifiers	NORMAL	
	Correct	***False + ve***
HoeffdingTree	111958	374
J48	112289	43
RandomForest	**112306**	**26**
RandomTree	112258	74
REPTree	112283	49

B. *Using Test Dataset*

TABLE V. PERFORMANCE METRICS

Classifiers	Classified Instances	
	Correctly	***Incorrectly***
HoeffdingTree	90.8358	9.1642
J48	**92.617**	**7.383**
RandomForest	92.4801	7.5199
RandomTree	90.3529	9.6471
REPTree	92.3026	7.6974

TABLE VI. ATTACK DETECTED

Classifiers	DOS		PROBE	
	Correct	***False + ve***	***Correct***	***False + ve***
HoeffdingTree	222200	7653	1345	2821
J48	223688	6165	3142	1024
RandomForest	***223932***	***5921***	***3249***	***917***
RandomTree	216830	13023	2852	1314
REPTree	222904	6949	3060	1106

Classifiers	U2R		R2L	
	Correct	***False + ve***	***Correct***	***False + ve***
HoeffdingTree	***27***	***201***	2	16187
J48	7	221	**946**	**15243**
RandomForest	2	226	890	15299
RandomTree	25	203	1734	14455
REPTree	23	205	1569	14620

Classifiers	NORMAL	
	Correct	***False + ve***
HoeffdingTree	58950	1641
J48	**60281**	**310**
RandomForest	59565	1026
RandomTree	59581	1010
REPTree	59530	1061

V. RESULT ANALYSIS

We use 10-fold cross validation and test data set provided by University of California to test and evaluate the algorithms. In 10-fold cross validation process the data set is divided into 10 subsets. Each time, one of the 10 subsets is used as the test set and the other remaining 9 subsets are used as the training set. Performance statistics are calculated across all 10 trials. Table III summarize the results of all classifiers and Table IV summarize category wise attack detected using 10-fold cross validation. Percentage of attack detection is shown in Table VII and FIG 1.

TABLE VII % OF ATTACK DETECTION USING CROSS VALIDATION

Classifiers	Attack Types				
	DOS	PROBE	U2R	R2L	Normal
HoeffdingTree	99.971	96.031	68.966	92.267	99.667
J48	99.995	**99.367**	55.172	95.822	99.962
RandomForest	**99.999**	99.002	**70.690**	**98.044**	**99.977**
RandomTree	99.994	98.563	63.793	97.778	99.934
REPTree	99.991	97.687	48.276	96.889	99.956

From Above table it is clear that almost all classifier performs well compare to each other and they are slightly better than each other. All classifier achieve more than 90% attack detection ratio in DOS, PROBE, R2L and more than 99% Normal category. Only in U2R category of attack the ratio is less than 75%. This is because of U2R types of attack are very less in training dataset of KDD Cup data set. Compare to other classifier RandomForest performs slightly better than other classifier in DOS, U2R,R2L and Normal Category and in PROBE category it is slightly behind J48 classifier.

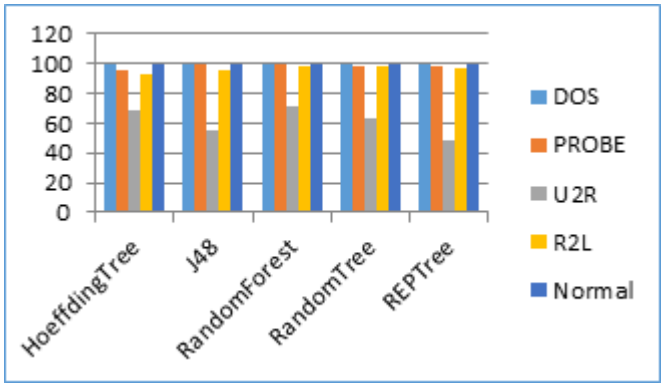

FIG 1 % OF ATTACK DETECTION USING CROSS VALIDATION

Table VIII shows percentage of attack detected by all classifiers using test dataset provided by KDD Cup data set. From Table VIII it is clear that almost all classifier performs well compare to each other in DOS attack category and they achieve more than 95% attack detection ratio except RandomTree. All classifier achieve more than 97% detection ratio in Normal category. Only in U2R and L2R category of attack the ratio is less than 12%. This is because of U2R and L2R types of attack are very less compare to other attack types in training dataset of KDD Cup data set. Compare to other classifier J48 performs slightly better than other classifier in U2R,R2L and Normal Category and RandomForest classifier performs slightly better than other classifier in DOS and PROBE category.

TABLE VIII % OF ATTACK DETECTION USING TEST DATA SET

Classifiers	Attack Types				
	DOS	PROBE	U2R	R2L	Normal
HoeffdingTree	96.670	32.285	11.842	0.012	97.288
J48	97.318	75.420	**3.070**	**5.843**	**99.485**
RandomForest	**97.424**	**77.988**	0.877	5.498	98.303
RandomTree	94.334	68.459	10.965	10.711	98.330
REPTree	96.977	73.452	10.088	9.692	98.246

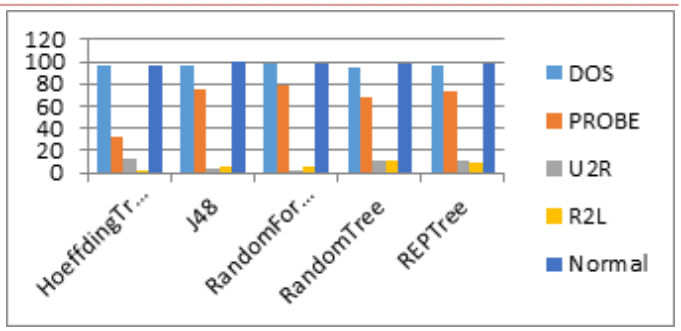

FIG 2 % OF ATTACK DETECTION USING TEST DATA SET

From table VIII it is clear that J48 classifier performs well in U2R, L2R and Normal category and in DOS, PROBE it is slightly behind RandomForest classifier.

VI. CONCLUSIONS AND FUTURE WORK

Tree base data mining classification techniques such as Hoeffding tree, J48, Random Forests, Random Tree, REPTree, were evaluated in this study on network intrusion detection dataset of KDD CUP1999 using Weka 3.7.11 tools. In general, the results show Random Forest using 10 fold cross validation and J48 using test data set classifiers are the best candidates considering their comparative classification accuracy.

The main challenge in intrusion detection is to achieve high detection rate and reduce false alarm rate. Any single classifier alone is not sufficient to achieve high accuracy and low false positive or negative. Therefore more than one classifier can be combined to improve overall performance of attack detection.

ACKNOWLEDGMENT

We wish to express our sincere thanks to our Principal, HOD, Professors and staff members of Computer Engineering Department at Dattakala Faculty of Engineering, Swami Chincholi, Bhigawan. We are very grateful to the authors of various articles on the internet, helping us become aware of the research currently ongoing in this field.

Last but not the least, We would like to thank all our Friends and Family members who have always been there to support and helped us to complete this research work

REFERENCES

[1] J. Feiman, “Hype Cycle for Application Security, 2012”, Gartner Group, July, 2012.

[2] “Advanced Persistent Threat: A Decade in Review”, Command FivePty Ltd, June, 2011..

[3] C. Elkan. "Results of the KDD'99 classiffer learning". SIGKDD Explorations Newsletter, vol. 1, pp. 63-64, 2000.

[4] J. R. Quinlan. "C4.5: programs for machine learning". Morgan Kaufmann Publishers Inc., 1993. ISBN 1-55860-238-0.

[5] M. Sabhnani and G. Serpen. "Application of machine learning algorithms to KDD intrusion detection dataset within misuse detection context". In International Conference on Machine Learning, Models, Technologies and Applications (MLMTA), pp. 209-215. CSREA Press, 2003.

[6] S. Sung, A.H. Mukkamala. "Identifying important features for intrusion detection using support vector machines and neural networks". In Proceedings of the Symposium on Applications and the Internet (SAINT), pp. 209-216. IEEE Computer Society, 2003.

[7] S. Chavan, K. Shah, N. Dave, S. Mukherjee, A. Abraham and S. Sanyal. "Adaptive neuro-fuzzy intrusion detection systems". In Proceedings of the International Conference on Information Technology: Coding and Computing (ITCC), vol. 1, pp.70-74. IEEE Computer Society, 2004.

[8] S. Peddabachigari, A. Abraham, C. Grosan and J. Thomas. "Modeling intrusion detection system using hybrid intelligent systems". Journal of network and computer applications, vol. 30, no. 1, pp.114-132, 2007.

[9] H. Kayacik, A. Zincir-Heywood and M. Heywood. "Selecting features for intrusion detection: A feature relevance analysis on KDD 99 intrusion detection datasets". In Proceedings of the Third Annual Conference on Privacy, Security and Trust (PST), 2005.

[10] C. Lee, S. Shin and J. Chung. "Network intrusion detection through genetic feature selection". In Seventh ACIS International Conference on Software Engineering, Artificial Intelligence, Networking, and Parallel/Distributed Computing (SNPD), pp. 109-114. IEEE Computer Society, 2006

[11] S. Chebrolu, A. Abraham and J. Thomas. "Feature deduction and ensemble design of intrusion detection systems". Computers & Security, vol. 24, no. 4, pp. 295-307, 2005.

[12] Y. Chen, A. Abraham and J. Yang. "Feature selection and intrusion detection using hybrid exible neural tree". In Advances in Neural Networks (ISNN), vol. 3498 of Lecture Notes in Computer Science, pp. 439-444. Springer Berlin / Heidelberg, 2005.

[13] R. Staudemeyer and C. Omlin. "Feature set reduction for automatic network intrusion detection with machine learning algorithms". In Proceedings of the Southern African Telecommunication Networks and Applications Conference (SATNAC).2009.

[14] S. Lakhina, S. Joseph and B. Verma. Feature reduction using principal component analysis for effective anomalybased intrusion detection on NSLKDD". International Journal of Engineering Science and Technology, vol. 2, no. 6, pp. 1790-1799, 2010.

[15] M. Tavallaee, E. Bagheri, W. Lu and A. A. Ghorbani. "A detailed analysis of the KDD CUP 99 data set". In 2009 IEEE Symposium on Computational Intelligence for Security and Defense Applications, Cisda, pp. 1{6. IEEE, Jul. 2009.

[16] Ben Amor, Benferhat, Elouedi, "Naive Bayes vs. Decision Trees in Intrusion Detection Systems, " Proc. of the 2004 ACM symposium on applied computing, 2004, pp. 420–424.

[17] Stein G, Chen B, Wu AS, Hua KA (2005), "Decision tree classifier for network intrusion detection with GA-based feature selection, " In: Proceedings of the 43rd annual southeast regional conference ACM vol 2, pp 136–141.

[18] G. V. Nadiammai and M. Hemalatha, "Perspective analysis of machine learning algorithms for detecting network intrusions, " IEEE Third International Conference on Computing Communication & Networking Technologies (ICCCNT), Coimbatore, India, 2012, pp. 1-7.

[19] T. Hwang, T.Lee, and Y. Lee, "A Three-tier IDS via Data Mining Approach, " 3rd annual ACM workshop on Mining network data, 2007, pp. 1-6.

[20] E. K. Reddy, M. IAENG, V. N. Reddy, and P. G. Rajulu, "A Study of Intrusion Detection in Data Mining, " World Congress on Engineering, vol. III, July 6-8, 2011.

[21] S. Subramanian, V. B. Srinivasan, and C. Ramasa, "Study on Classification Algorithms for Network Intrusion Systems, " pp. 1242-1246, 2012.

[22] NSL-KDD dataset, [Available Online] http://iscx.ca/NSLKDD/

[23] R. Lippmann, J. W. Haines, D. J. Fried, J. Korba, and K. Das, "The 1999 DARPA off-line intrusion detection evaluation, " Computer Networks, vol. 34, no. 4, pp. 579-595, 2000.

[24] P. Srinivasulu, D. Nagaraju, P. R. Kumar, and K. N. Rao, "Classifying the Network Intrusion Attacks using Data Mining Classification Methods and their Performance Comparison, " IJCSNS International Journal of Computer Science and Network Security, vol. 9, no.6, pp. 11-18, 2009.

[25] B. Mukherjee, L.T. Heberlein, and K.N. Levitt, "Network intrusion detection, " IEEE Network, vol.8, no.3, pp. 26–41, 1994.

[26] I. Butun, S.D. Morgera, and R. Sankar, "A Survey of Intrusion Detection Systems in Wireless Sensor Networks, " IEEE Communication Surveys & Tutorials, vol.16, no.1, pp. 266–282, 2014

[27] Frawley W., Piatetsky-Shapiro G., and C. Matheus "Knowledge Discovery in Databases: An Overview, " AI Magazine, 1992, pp. 213-228..

[28] Geoff Hulten, Laurie Spencer, Pedro Domingos "Mining time-changing data streams " ACM SIGKDD Intl. Conf. on Knowledge Discovery and Data Mining, 97-106, 2001..

[29] Ross Quinlan (1993). C4.5: Programs for Machine Learning. Morgan Kaufmann Publishers, San Mateo, CA.

[30] Leo Breiman (2001). Random Forests. Machine Learning. 45(1):5-32.

[31] Weka http://www.cs.waikato.ac.nz/ml/index.html.

[32] KDD CUP 1999 (http://kdd.ics.uci.edu/databases/kddcup99/)

[33] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann and I.H. Witten, "The WEKA Data Mining Software: An Update", ACM SIGKDD Explorations Newsletter, Volume 11 , Issue 1, pp. 10-18, 2009

[34] [34] L. Wenke, S. J. Stolfo, and K. W. Mok, "A data mining framework for building intrusion detection models, " IEEE Symposium on Security and Privacy, 1999, pp. 120-132.

[35] P. Srinivasulu, D. Nagaraju, P. R. Kumar, and K. N. Rao, "Classifying the Network Intrusion Attacks using Data Mining Classification Methods and their Performance Comparison, " IJCSNS International Journal of Computer Science and Network Security, vol. 9, no.6, pp. 11-18, 2009

[36] S. Subramanian, V. B. Srinivasan, and C. Ramasa, "Study on Classification Algorithms for Network Intrusion Systems, " pp. 1242-1246, 2012.

[37] G. Kalyani , A. Jaya Lakshmi, "Performance Assessment of Different Classification Techniques for Intrusion Detection" IOSR Journal of Computer Engineering (IOSRJCE) ISSN: 2278-0661, ISBN: 2278-8727 Volume 7, Issue 5 (Nov-Dec. 2012), PP 25-29

[38] G. Meera Gandhi, "Machine Learning Approach for Attack Prediction and Classification using Supervised Learning Algorithms", International Journal of Computer Science & Communication Vol. 1, No. 2, July-December 2010, pp. 247-250

[39] Adil Baykasoğlua, Lale Özbakirb, "MEPAR-miner: Multi-expression programming for classification rule mining", European Journal of Operational Research, Volume 183, Issue 2, 1 December 2007, Pages 767–784

[40] Jaiwei Han, Micheline Kamber, "Data Mining: Concepts and Techniques (2001)", ISBN 1-55860-489-8 (2001)

[41] Chakchai So, Nutakarn Mongkonchai, Phet Aimtongkham, Kasidit Wijitsopon and Kanokmon Rujirakul " An Evaluation of Data Mining Classification Models for Network Intrusion Detection", DICTAP 2014, Page No: 90 – 94

[42] MichaelWalker http://www.datasciencecentral.com/profiles/blogs/random-forests-algorithm