Journal of Theoretical and Applied Information Technology
20th November 2013. Vol. 57 No.2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

DATA MINING IN NETWORK SECURITY - TECHNIQUES & TOOLS: A RESEARCH PERSPECTIVE

[1]D.ASIR ANTONY GNANA SINGH, [2]E.JEBAMALAR LEAVLINE

[1]Bharathidasan Institute of Technology, Anna University, Tiruchirappalli, Department of CSE

[2]Bharathidasan Institute of Technology, Anna University, Tiruchirappalli, Department of ECE

E-mail: [1]asirantony@gmail.com, [2]jebilee@gmail.com

ABSTRACT

This paper presents recent trends and practices in data mining to handle the rising risks and threats in the area of Network security in today's digital age and discusses the various data mining tools for data analysis and prediction, network tools for sniffing and analyzing the networks. This paper proposes a supervised learning based Intrusion Detection System (IDS) to identify the intruders, attackers in a network and covers the most significant advances and emerging research issues in the field of data mining in network security. This will be beneficial to academicians, industrialists and students who incline towards research and development in the area of data mining in network security.

Keywords: *Learning, Intrusion detection, Supervised Learner, Data Mining, Network Security*

1. INTRODUCTION

In this digital age, we can't imagine the world without communication. The human beings need to exchange information for various purposes. Securing the communication is a vast challenge due to the raising threats and attacks against network security.

Securing the network is the major challenge in this information era from the various types of network threats and attacks [1]. The threats are classified based on their behaviour such as *leakage*: unauthorized access of information available in the network [2]. *Tampering:* modifying the information without permission of the author. *Vandalism:* making malfunction over a normal execution of a system. The various types of attacks such as *eavesdropping:* collecting the replica information without obtaining permission to the arbiter. *Masquerading:* making conversation using through others identity without permission of others. *Message tampering:* modifying and altering the information while travel on the communication media. *Man-in-the-middle attack:* is a one type of message interfering in which an attacker interrupt the very first message in an exchange of encrypted keys to establish a secure channel [3][4]. The attacker substitutes compromised keys that enable them to decrypt subsequent messages before reconfiguring them in the correct keys and passing them on. *Replying:* this is one type of attack that stores intercept messages then sends these messages later. This attack may be effective even with authenticated and encrypted messages [3]. *Denial of service:* makes the transmission channels and systems as busy as possible by sending garbage data for denying the service [5].

The knowledge about these attacks is acquired from the huge volume of network data with data mining tools. This knowledge facilitates the security system to identify the attackers or hackers based on their behaviour in a network. The behaviour of the attackers and hackers are studied and identified by two types of learning strategies namely supervised and unsupervised learning.

In data mining based network security approach, the network sniffing or scanning software collects the data about the activities of the attacker. The collected data are learnt by the supervised learning algorithm and the predictive model is built. This model predicts and detects the attackers and hackers.

This paper proposes a supervised learning based intrusion detection system that utilizes the advantages of supervised learning and prediction techniques. Also, a detailed discussion on the step by step procedure of building a predictive model using data mining tools is presented.

Journal of Theoretical and Applied Information Technology
20th November 2013. Vol. 57 No.2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

2. METHODOLOGY

2.1 Collecting Network Historical and Log Data

The network historical and log data are the network activity data. These data are passively monitored, scanned and collected through the various monitoring mechanisms [6][7][8][9][10][11][12] are explored as follows:

Kismet: Kismet is a scanning tool that uses the 802.11 wireless detectors, and permits card based passive monitoring (RF-mon) to sniff any 802.11x standard networks. It displays ARP (Address Resolution Protocol) and DHCP (Dynamic Host Configuration Protocol) traffic, to save files in the file format of Wireshark and TCPDump and display level of activity at some different channels. It decodes and measures the real-time traffic signals. Hackers mostly use the Kismet, since it can be used in any communication network. It helps to detect the intrusions. It runs on Mac and Linux the platforms [13][14].

Snoop: Sun Microsystems developed a common intrusion detection sniffing tool 'Snoop' to function with Solaris platform. It adopts single and multiline format to display the results. It sniffs IPv4 and IPv6 network packets. This tool is similar to TCPDump in displaying and formatting of the files. Snoop is considerably good than TCPdump because of its user friendly interface [15][16].

Wireshark: The Gerald Combs developed first public packet sniffing tool 'Wireshark' earlier known as Ethereal. It is an open source packet sniffer and analyzer and licensed by GNU GPI (General Public License). It works with the FreeBSD, UNIX, Linux, Solaris, OpenBSD, and Windows platforms [17]. It is user friendly to capture, filter and analyze packets. This tool is very flexible since its log files are in different format [14][18][19].

TCPdump: The Lawrence Berkeley National Laboratory developed the TCPdump open source network scanning and repair tools for TCP/IP (Transmission Control Protocol/Internet Protocol) packet networks in 1990. The user intercepts captures and monitors TCP-IP packets during transmission in a network. It works with Unix, Linux, Solaris, BSD (Berkeley Software Distribution), Mac and Windows platforms. It uses the command line to capture and filter log based on certain rules. These log files are not in understandable format [20][21][22].

2.2 Supervised Learning Algorithms to Learn the Historical or Log Data for Building the Predictive Model

The collected historical/log data from the network are learned by the classification algorithms for building the predictive model to identify the hackers and attackers. In this section the most popular classifiers also called as supervised learners namely probabilistic algorithm Nave Bayes(NB), tree based C4.5(J48) and Instance based IB1 (Instance-based) are described.

Naive Bayes(NB): This classification algorithm uses Bayes' theorem and the features of the training dataset as shown in the Table 1 are assumed as independent to the given class labels for building the predictive model [23][24]. This classifier relies on discriminant function as seen in equation (1):

$$f_i(X) = \prod_{j=1}^{N} P(x_j \mid c_i) P(c_i) \quad (1)$$

Let the dataset be D with the Features X = (x_1, x_2, ...x_N) and the Class C with j labels C_j, j = 1, 2, ...N. This algorithm computes the conditional probabilities $P(x_j|c_i)$ and prior probabilities $P(c_i)$ on given training dataset to build the predictive model. $P(c_i)$ are computed by counting data which present in the Class label C_i divides the resultant count based on the number of the training data. The same way is followed to compute the probabilities through observed frequency of feature distribution in x_j within the training dataset which is labelled. The posterior probability is computed on each class to predict the unknown labelled data [25][26][27].

C4.5 (J48): This algorithm uses the decision tree to build the predictive model. The decision tree is constructed in numerous methods. All these methods convert the given dataset in to a tree structure. The nodes of the tree represent the features and the edges represent the association between the features by value of features the lowest level of the node represents the class label [28]. Recursively the value of the features are calculated by the information gain or entropy measure to convert the training datasets in to tree structure. The low entropy and high information gain value of feature is selected repetitive node as head node to split the dataset and convert the dataset into tree structure. The tree structure is used as a rule to

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

predict the unlabeled data in prediction [25][26][27].

IB1: This algorithm uses the nearest neighbour principle to construct the predictive model. In this approach, the distance between the training instance and the given test instance are calculated by the Euclidean distance measure. If more than one instance has the smallest distance to the test instance, the first found instance is used. Nearest neighbor is one of the most significant learning algorithms; it can be adapted to solving wider problems [29]. Let a dataset D has X instances (X_1,X_2,X_3…X_n) and F feature (F1,F2,F3,…Fm) with the class label Cj where j=1,2…K. This algorithm ranks the distance value of the neighbouring instances to predict the unlabeled data X with the Class label. The Euclidean distance measure is used to compute the weight of the neighbours of the instances X. In this way the unlabeled data is predicted by the voting for the weight to calculate the nearest neighbours of the particular class to predict the unknown data. This method is not relay on prior probability as NB algorithms. computation cost is high when the numbers of instance are more since the distance measure is computationally costly. Hence, the feature selection algorithm are used to reduced the dimensionality for the training dataset to effective the computational cost of this algorithm [26][27].

2.3 Data Analysis and Mining Tools

In this section, various open source data mining tools for analyzing and building the predictive model for the unlabeled data classification and prediction are discussed.

WEKA (Waikato Environment for Knowledge Analysis): The university of Waikato New Zealand developed this open source tool in Java technology. This consists of a collection of machine learning algorithms such as Clustering, Feature selection/Attribute subset selection, Classification, Association Rule mining etc. The Weka provides four interfaces namely Explorer, Experimenter, Knowledge Flow, Simple CLI (command-line interface) to work with machine learning algorithm and datasets. The Explorer provides a platform for data exploration. The Experimenter provides a platform to perform Experiments for conducting statistical tests among the learning schemes. The Knowledge Flow provides a Graphical User Interface to implement the functionalities available in explorer. The SimpleCLI provides a simple command-line interface to execute the Weka commands [30].

Orange: The open source Orange tool contains a variety of machine learning and data mining algorithms with routines to data exploration. It works with Python and C++ it works for the functionalities such as decision trees, attribute subset selection, boosting and bagging .It gives a platform for visual programming to use the visual component widgets, this explores the data for analysis. The widgets modularity connects the communication media to exchange the data packets automatically for data analysis. This orange is used for many data analyzing applications [31].

R Tool: Initially, the Ihaka and Gentleman from University of Auckland, New Zealand developed R tool in 1996. R provides a platform to compute the statistics for data analysis. , R works with the Unix, Windows, Mac platform. The formal work flow for a data mining task is carried out through the following steps [32].

- Load a Dataset and select features
- Explore the data in understandable format
- Distribution of Test
- Transform the data to suit the modeling
- Build the Models
- Evaluate the models with dataset
- Review the Log in data mining task.

Keen Tool: The Keen (Knowledge Extraction based on Evolutionary Learning) carry outs many data mining tasks includes regression, supervised, unsupervised and evolutionary learning algorithms. It consist of my library functions for pre and post processing method for data manipulation and soft computing mythology for helps to the scientific research in the area of machine learning. It also supports the fuzzy and genetic algorithm to do research in the area of data mining and various applications related to data analysis [33].

2.4 Network Security Techniques, Mechanisms and Protocols

Numerous security techniques, mechanisms, devices and protocols are available to secure the network from the threats and attacks. The security techniques [34] are Cryptography, Virtual Private Network (VPN) [35], tunnelling [36], Hashing [37], Digital Signature, Bastion Host Configuration Certificate Authority to PKI (Public Key Infrastructure)[38] and so on. The protection mechanisms and devices are Firewalls, Proxy server, Demilitarized Zone (DMZ), Intrusion Detection System [39], Intrusion Prevention System, Network access server: Remote Authentication Dial In User Service

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

(RADIUS)[40], Honey pot, Honey net, Antivirus Software and so on. The protocols are SSL(Secured socket layer) to Secure web, SSH(Secure Shell)[41] to Secure telnet and rlogin or file transfer, S/MIME to (Secure/Multipurpose Internet Mail Extensions) Secure email [42], Secure Information Management to Log Management [43].

2.5 Intrusion Detection System (IDS)

An intrusion occurs when intruder tries to gain entry or disrupt the normal operations of network. Intrusion detection system learns the normal activities of the networks and build the predictive model like human behavioral model [44]. Based on this model it identifies the intruders in a network. Intrusion detection methods classified as Signature-based IDS [45], Statistical anomaly-based IDS, Stateful protocol analysis IDS and Log file monitors.

The Signature-based IDS also called as "knowledge-based IDS" examines network traffic in search of patterns that match Known signatures i.e. preconfigured, predetermined attack patterns. The bottleneck of this approach is that the new type of attacks must be identified and updated in the database and it is a time consuming process.

The statistical anomaly-based IDS is also called as "behaviour-based IDS" [46] . It collects statistical summaries by observing traffic. The normal period of evaluation establishes a performance baseline. The baseline data can include variables such as host memory or CPU (Central Processing Unit) usage, network packet types, and packet quantities. Once the baseline is established, The IDS compares the network activity to this baseline. If it exceeds the baseline then that level is known as "Clipping level", Then IDS system immediately sends an alert to the administrator. The advantage of this type is it can detect new types of attacks, since it looks for abnormal activity of any type and disadvantage is it requires much more overhead and processing capacity than signature based IDS. So, this method is not suitable for heavy packet traffic [47][44].

3. PROPOSED SUPERVISED LEARNING BASED INTRUSION DETECTION SYSTEM

The proposed Supervised learning based intrusion deduction system identifies the attacks and hackers in computer networks. This system collects the historical or log data related network activities (the behaviors of intruders, warms, virus, and attackers) of a particular network by any one of the network sniffing or scanning software. The supervised learner classifiers learns the collected historical and log data then builds a predictive model in order to identify the intruders. The constructed predictive model identifies the attackers and prevents the attackers in the particular network.

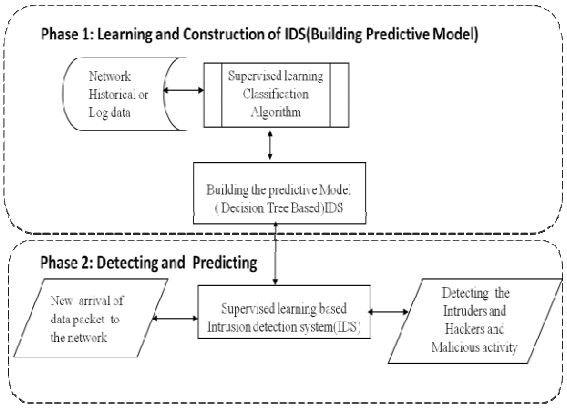

Figure. 1 Architecture of the proposed supervised learning based intrusion detection system

Figure 1 shows the architecture of the proposed supervised learning based intrusion detection system. This system works in two phases. In first phase, the system learns the network historical and log data by the supervised learning classifier NB, IB1, C4.5 and builds the predictive model. In second phase, the predictive model detects and identifies the attackers and hackers in networks.

3. IMPLEMENTATION AND SIMULATION OF THE PROPOSED SUPERVISED LEARNING BASED INTRUSION DEDUCTION SYSTEM

4.1 Experimental Setup & Step-By-Step Procedure

To demonstrate the proposed system, an experimental setup is constructed with the data mining simulation tool Weka and the network historical data is collected as benchmark training dataset from the repository with the format 'arff' (attribute relation file format) format [48] [49]. The supervised learner classifier, NB, IB1 and C4.5 are deployed to learn the network historical dataset to build the predictive model and calculate their accuracy and time taken to build the model.

The step-by-step experimental procedure is described below.

Loading the training dataset: The collected historical or log data from the network are

Journal of Theoretical and Applied Information Technology
20th November 2013. Vol. 57 No.2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

loaded into pre-processing unit of the Weka tool as shown in Figure.2. The range of the dataset in terms of features from 41 to 500 and instances from 213 to 466 as shown in the Table 1 are to train the supervised learner in order to construct the predictive model.

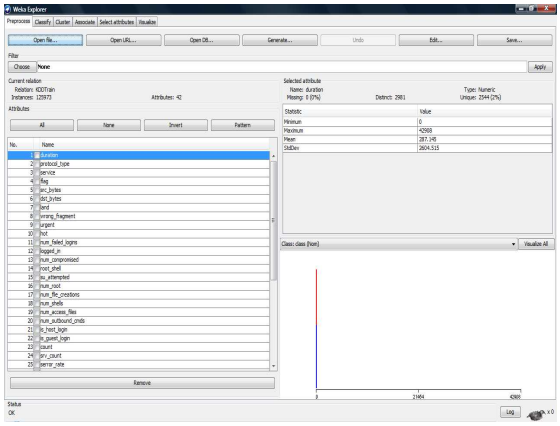

Figure 2 Loading And Pre-Processing The Dataset

Constructing the predictive Model and Testing: The classifier supervised learner C4.5 is chosen to train the dataset as shown in the Figure. 3, and trains the dataset as shown in the Figure.4

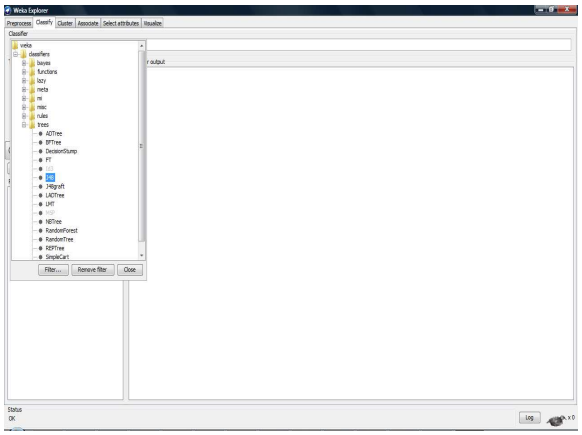

Figure.3 Select Supervised Learning Classifier

The decision tree is constructed to build the predictive model. The predicted model is saved as in Figure.5, and the predictive model is loaded (Figure.6), and the predictive model is deployed a

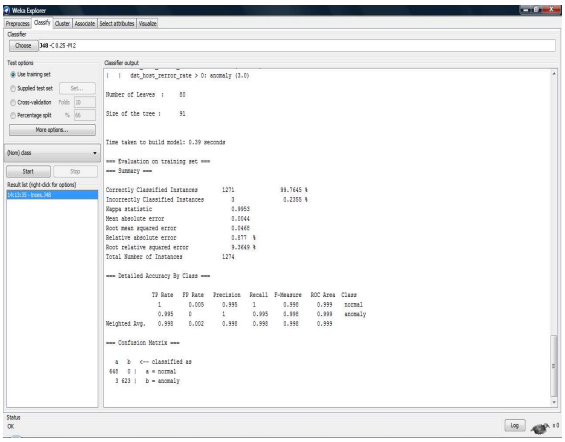

Figure 4 Training And Building The Predictive Model

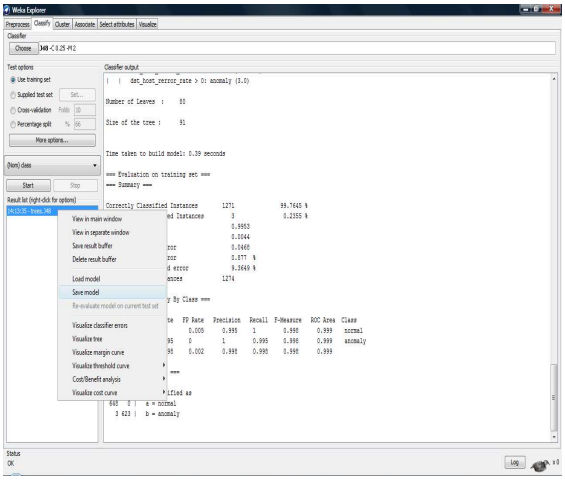

Figure 5 Saving The Predictive Model

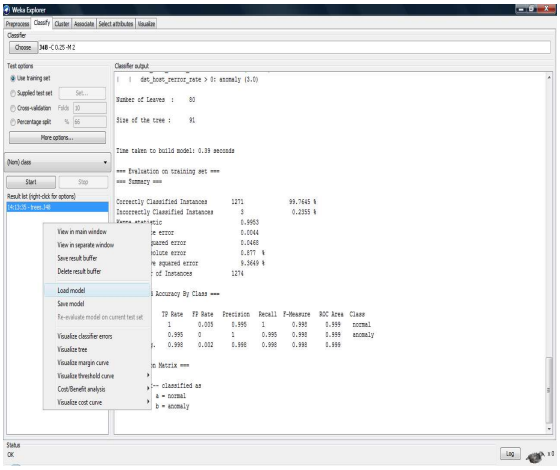

Figure 6 Loading The Predictive Model

Journal of Theoretical and Applied Information Technology
20th November 2013. Vol. 57 No.2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

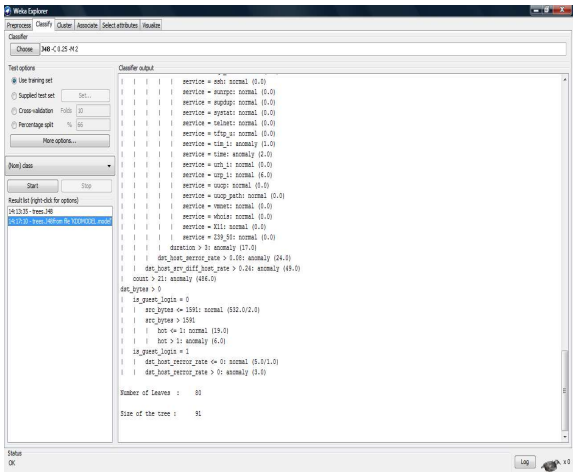

Figure 7 Enabling The Option Output Prediction For Viewing The Prediction

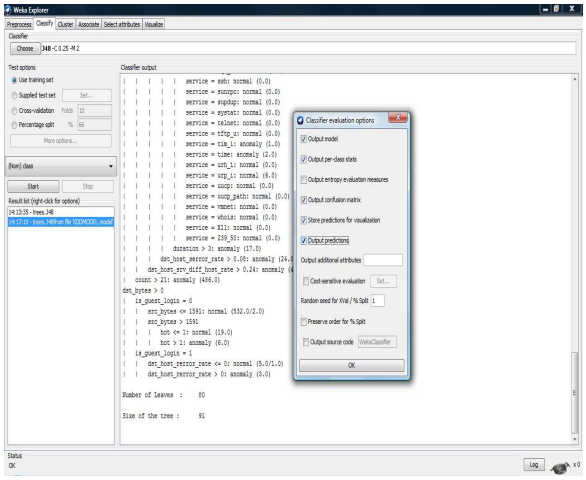

Figure 8 Enabling the output prediction mode

supervised learning based intrusion detection system as shown in Figure7. And enable the output prediction mode to view the predicted unlabeled data as shown in Figure 8. The unlabeled test data is fed in to the build predictive model as shown in Figure 9 and configure or evaluated as shown in Figure 10 the model with the test data to identify the intrudes or attackers. The attackers or hackers are identified as the output of the predictive model as shown in the Figure 11.

5. DISCUSSIONS AND CONCLUSION

This paper explored and analyzed the various challenges of threats and attacks in networks in this recent era, various network sniffing, snooping tools for capturing the network data and log data for analysis and learning, various data mining tools for learning and building the predictive models and various supervised learning classification algorithms for learning the network data and identify the behaviour of the attackers and hacker.

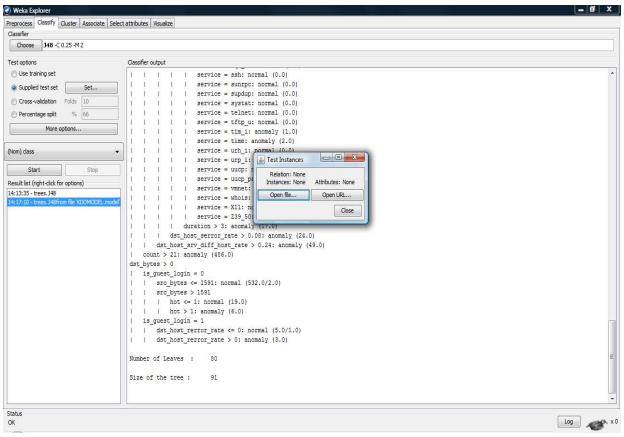

Figure 9 Feeding The Testing Dataset For Prediction

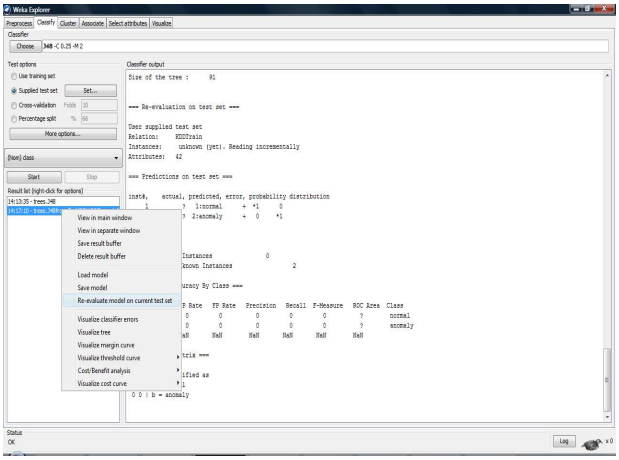

Figure 10 Re-Evaluating The Model On The Current Test Set

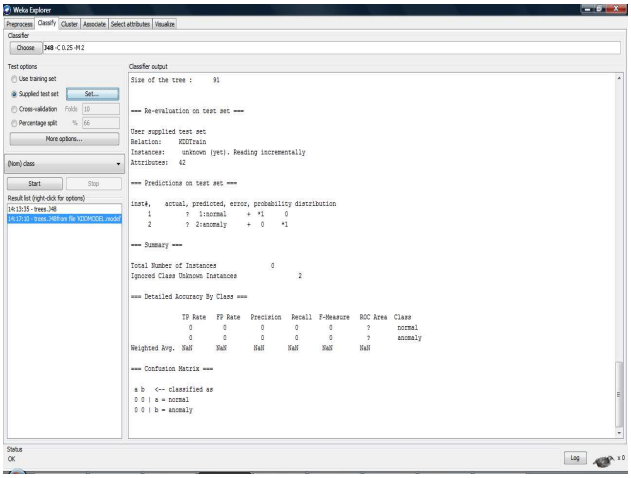

Figure 11 Predicting and identifying the attackers and intruders.

Journal of Theoretical and Applied Information Technology
20th November 2013. Vol. 57 No.2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

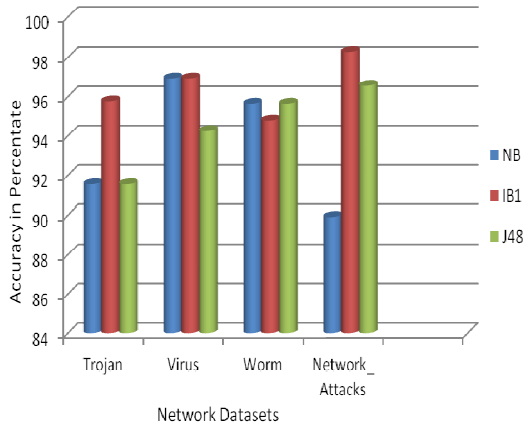

Figure 12 Comparison On Accuracy Of Learners

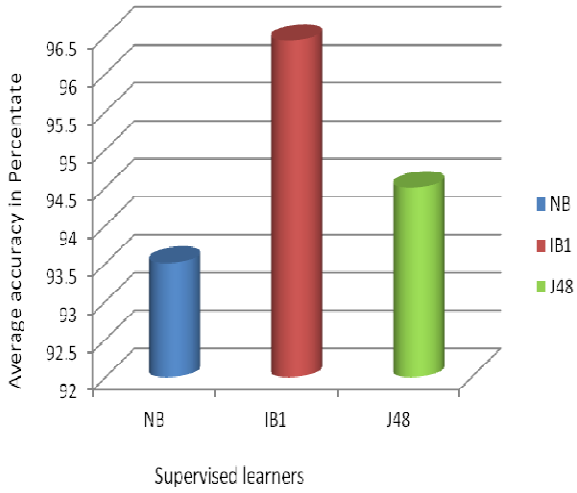

Figure 14 Comparison On Average Accuracy Of Learners

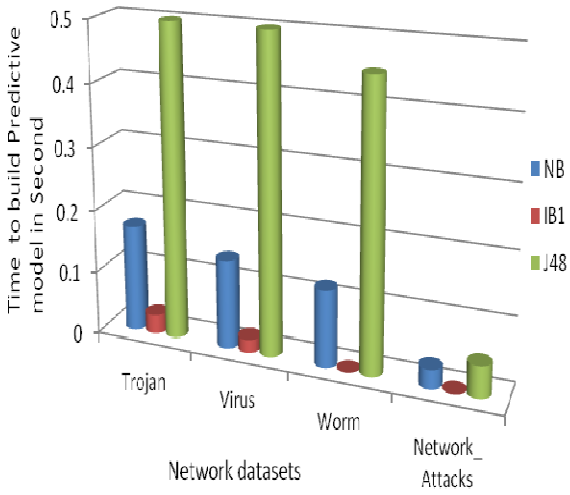

Figure 13 Comparison On Time Taken To Build The Predictive Model By Learners

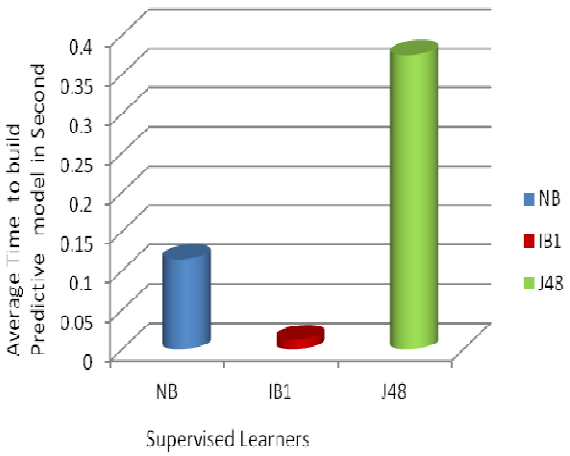

Figure 15 Comparison On Average Time Taken To Build The Predictive Model By Learners

Table 1. Experimental Results On Datasets With Learning Algorithms

Datasets			Accuracy of Supervised learner algorithm in Percentage			Time Taken to build Predictive model in seconds		
Name	Instances	Features	NB	IB1	J48	NB	IB1	J48
Trojan	213	500	91.54	95.77	91.54	0.17	0.03	0.50
Virus	261	500	96.93	96.93	94.25	0.14	0.02	0.50
Worm	230	500	95.65	94.78	95.65	0.12	0.00	0.45
Network_ Attacks	466	41	89.91	98.28	96.56	0.03	0.00	0.05
Average			93.50	96.44	94.50	0.11	0.01	0.37

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

A novel supervised learning based Intrusion detection system has been proposed and implemented with the supervised learning algorithm NB, IB1, J48 (C4.5) and simulated by the data mining tool Weka with standard benchmark training and testing datasets. This simulation results shown in Figure 12 to Figure 15 depict that, in the experiment of accuracy analysis, the Instance based learner IB1 produces high accuracy compared to all other supervised learners. The Naive Bayes NB comparably gives good accuracy than the Tree-based supervised learner J48. The Instance Based supervise learner takes very less time to build the predictive model compared to the other supervised learners and the Naive Bayes NB takes less time to build the predictive model compared to the Tree based supervised learner J48.

We hope, this supervised learning based intrusion detection system provides a suitable way to identify and detect intruders and hackers in real-time networks and this paper is very constructive to the researchers especially working in the area of data mining in network security.

REFRENCES:

[1] M. Schäfer, V. Lenders, and I. Martinovic, "Experimental analysis of attacks on next generation air traffic communication," in *Applied Cryptography and Network Security*, 2013, pp. 253–271.

[2] N. Meng, J. Wang, E. Kodama, and T. Takata, "Reducing data leakage possibility resulted from eavesdropping in wireless sensor network," *International Journal of Space-Based and Situated Computing*, vol. 3, no. 1, pp. 55–65, 2013.

[3] T. Denning, T. Kohno, and H. M. Levy, "Computer security and the modern home," *Communications of the ACM*, vol. 56, no. 1, pp. 94–103, 2013.

[4] T.-H. Lin, C.-Y. Lin, and T. Hwang, "Man-in-the-Middle Attack on 'Quantum Dialogue with Authentication Based on Bell States'," *International Journal of Theoretical Physics*, pp. 1–5, 2013.

[5] Z. Tan, P. Nanda, R. P. Liu, A. Jamdagni, and X. He, "A System for Denial-of-Service Attack Detection Based on Multivariate Correlation Analysis," *IEEE Transactions on Parallel and Distributed Systems*, vol. 99, no. 1, p. 1, 2013.

[6] A. M. Rajeswari, G. V. Aishwarya, V. A. Nachammai, and C. Deisy, "Temporal outlier detection on quantitative data using unexpectedness measure," in *Intelligent Systems Design and Applications (ISDA), 2012 12th International Conference on*, 2012, pp. 420–424.

[7] G. Han, J. Jiang, W. Shen, L. Shu, and J. Rodrigues, "IDSEP: a novel intrusion detection scheme based on energy prediction in cluster-based wireless sensor networks," *IET Information Security*, vol. 7, no. 2, pp. 97–105, 2013.

[8] H. Zhao and Y. Shi, "Detecting Covert Channels in Computer Networks Based on Chaos Theory," 2013.

[9] G.-H. Tu, C. Peng, H. Wang, C.-Y. Li, and S. Lu, "How Voice Calls Affect Data in Operational LTE Networks," 2013.

[10] B. G. Gohil, R. K. Pathak, and A. A. Patel, "Federated Network Security Administration Framework," 2013.

[11] C. Thomas and N. Balakrishnan, "Issues and Challenges in Intrusion Detection with Skewed Network Traffic," 2013.

[12] G. Ruiz Utgés, "Vulnerability assessment of distributed systems," B.S. thesis, 2013.

[13] E. J. Morgan, M. G. Shean, F. Alizadeh-shabdiz, and R. K. Jones, *Continuous Data Optimization of Moved Access Points in Positioning Systems*. 2013.

[14] F. Li, M. Li, R. Lu, H. Wu, M. Claypool, and R. Kinicki, "Tools and techniques for measurement of ieee 802.11 wireless networks," in *Modeling and Optimization in Mobile, Ad Hoc and Wireless Networks, 2006 4th International Symposium on*, 2006, pp. 1–8.

[15] D. Dasgupta and H. Brian, "Mobile security agents for network traffic analysis," in *DARPA Information Survivability Conference & Exposition II, 2001. DISCEX'01. Proceedings*, 2001, vol. 2, pp. 332–340.

[16] P. Li, C. Li, and T. Mohammed, "Building a repository of network traffic captures for information assurance education," *Journal of Computing Sciences in Colleges*, vol. 24, no. 3, pp. 99–105, 2009.

[17] U. Banerjee, A. Vashishtha, and M. Saxena, "Evaluation of the Capabilities of WireShark as a tool for Intrusion Detection," *International Journal of Computer Applications*, vol. 6, no. 7, pp. 1–5, Sep. 2010.

[18] S. Zeadally, E. Yaprak, Y. Li, and X. Che, "A Survey of Network Performance Tools for Computer Networking Classes,"

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

presented at the Computers and Advanced Technology in Education.

[19] R. Shimonski, *The Wireshark Field Guide: Analyzing and Troubleshooting Network Traffic*. Newnes, 2013.

[20] J. Therdphapiyanak and K. Piromsopa, "An analysis of suitable parameters for efficiently applying K-means clustering to large TCPdump data set using Hadoop framework," in *Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON), 2013 10th International Conference on*, 2013, pp. 1–6.

[21] N. T. Anh and R. Shorey, "Network sniffing tools for WLANs: merits and limitations," in *Personal Wireless Communications, 2005. ICPWC 2005. 2005 IEEE International Conference on*, 2005, pp. 389–393.

[22] F. Fuentes and D. C. Kar, "Ethereal vs. Tcpdump: a comparative study on packet sniffing tools for educational purpose," *Journal of Computing Sciences in Colleges*, vol. 20, no. 4, pp. 169–176, 2005.

[23] L. Jiang, Z. Cai, H. Zhang, and D. Wang, "Naive Bayes text classifiers: a locally weighted learning approach," *Journal of Experimental & Theoretical Artificial Intelligence*, vol. 25, no. 2, pp. 273–286, 2013.

[24] J. Weston, F. Pérez-Cruz, O. Bousquet, O. Chapelle, A. Elisseeff, and B. Schölkopf, "Feature selection and transduction for prediction of molecular bioactivity for drug design," *Bioinformatics*, vol. 19, no. 6, pp. 764–771, 2003.

[25] J. Novaković, P. \vSTRBAC, and D. Bulatović, "Toward optimal feature selection using ranking methods and classification algorithms," *The Yugoslav Journal of Operations Research ISSN: 0354-0243 EISSN: 2334-6043*, vol. 21, no. 1, 2011.

[26] D. Asir Antony Gnana Singh, S. Appavu Alias Balamurugan, and E. Jebamalar Leavline, "Towards higher accuracy in supervised learning and dimensionality reduction by attribute subset selection-A pragmatic analysis," in *Advanced Communication Control and Computing Technologies (ICACCCT), 2012 IEEE International Conference on*, 2012, pp. 125–130.

[27] A. G. Singh, D. Asir, A. Balamurugan, S. Appavu, and E. J. Leavline, "An empirical study on dimensionality reduction and improvement of classification accuracy using feature subset selection and ranking," in *Emerging Trends in Science, Engineering and Technology (INCOSET), 2012 International Conference on*, 2012, pp. 102–108.

[28] E. P. Xing, M. I. Jordan, and R. M. Karp, "Feature selection for high-dimensional genomic microarray data," in *ICML*, 2001, vol. 1, pp. 601–608.

[29] M. Kuramochi and G. Karypis, "Gene classification using expression profiles: a feasibility study," *International Journal on Artificial Intelligence Tools*, vol. 14, no. 04, pp. 641–660, 2005.

[30] 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*, vol. 11, no. 1, pp. 10–18, 2009.

[31] J. Dem\vsar, B. Zupan, G. Leban, and T. Curk, *Orange: From experimental machine learning to interactive data mining*. Springer, 2004.

[32] G. J. Williams, "Rattle: A data mining GUI for R," *The R Journal*, vol. 1, no. 2, pp. 45–55, 2009.

[33] J. Alcalá-Fdez, L. Sánchez, S. García, M. J. del Jesús, S. Ventura, J. M. Garrell, J. Otero, C. Romero, J. Bacardit, and V. M. Rivas, "KEEL: A software tool to assess evolutionary algorithms for data mining problems," *Soft Computing*, vol. 13, no. 3, pp. 307–318, 2009.

[34] D. A. A. G. Singh and E. J. Leavline, "IATARPA: Implementation of anonymity threat avoidance routing protocol architecture for MANET," in *Advanced Computing (ICoAC), 2011 Third International Conference on*, 2011, pp. 321–326.

[35] L.-H. Gong, Y. Liu, and N.-R. Zhou, "Novel Quantum Virtual Private Network Scheme for PON via Quantum Secure Direct Communication," *International Journal of Theoretical Physics*, pp. 1–9, 2013.

[36] James Hoagland, "The Teredo Protocol Tunneling Past Network Security and Other Security Implications - Google Search," Symantec Advanced Threat Research, United States, 2006.

[37] M. O. Pervaiz, M. Cardei, and J. Wu, "Routing security in ad hoc wireless networks," in *Network Security*, Springer, 2010, pp. 117–142.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

[38] T. H.-J. Kim, L.-S. Huang, A. Perring, C. Jackson, and V. Gligor, "Accountable key infrastructure (AKI): A proposal for a public-key validation infrastructure," in *Proceedings of the 22nd international conference on World Wide Web*, 2013, pp. 679–690.

[39] C. Chung, P. Khatkar, T. Xing, J. Lee, and D. Huang, "NICE: Network Intrusion Detection and Countermeasure Selection in Virtual Network Systems," 2013.

[40] W. Liu, J. Zhong, and J. Lin, "Secure Access Scheme Research and Design Based on the Internet of Things," *Applied Mechanics and Materials*, vol. 278, pp. 1818–1821, 2013.

[41] H. Xia and J. C. Brustoloni, "Hardening Web browsers against man-in-the-middle and eavesdropping attacks," in *Proceedings of the 14th international conference on World Wide Web*, 2005, pp. 489–498.

[42] S. Roosa and S. Schultze, "Trust Darknet: Control and Compromise in the Internet&# x2019; s Certificate Authority Model," 2013.

[43] I. Ray, K. Belyaev, M. Strizhov, D. Mulamba, and M. Rajaram, "Secure Logging As a Service—Delegating Log Management to the Cloud," 2013.

[44] D. A. A. G. Singh and E. J. Leavline, "Competency-Based Calisthenics of Learning Outcomes for Engineering Education." International Journal of Education and Learning Vol. 2, No. 1, March, 2013

[45] M. Ghorbanian, B. Shanmugam, G. Narayansamy, and N. B. Idris, "Signature-based hybrid Intrusion detection system (HIDS) for android devices," in *Business Engineering and Industrial Applications Colloquium (BEIAC), 2013 IEEE*, 2013, pp. 827–831.

[46] A. Benham, H. Read, and I. Sutherland, "Network Attack Analysis and the Behaviour Engine," *Int. J. Com. Net. Tech*, vol. 1, no. 2, pp. 103–117, 2013.

[47] J. A. Santos, M. G. de Mendon\cca, and C. V. Silva, "An exploratory study to investigate the impact of conceptualization in god class detection," in *Proceedings of the 17th International Conference on Evaluation and Assessment in Software Engineering*, 2013, pp. 48–59.

[48] M. Tavallaee, E. Bagheri, W. Lu, and A.-A. Ghorbani, "A detailed analysis of the KDD CUP 99 data set," in *Proceedings of the Second IEEE Symposium on Computational Intelligence for Security and Defence Applications 2009*, 2009.

[49] "Datasets." [Online]. Available: http://nexginrc.org/Datasets/Default.aspx. [Accessed: 18-Sep-2013].