Review of Soft Computing in Malware Detection

Raman Singh
University Institute of Engineering and Technology,
Panjab University,Chandigarh

Harish Kumar
University Institute of Engineering and Technology,
Panjab University,Chandigarh

R.K. Singla
Department of Computer Science and Application
Panjab University,
Chandigarh

ABSTRACT
Soft computing techniques are widely used in malware detection in these days. These techniques have the ability of learning from the past incidences and can categories normal and abnormal behaviour. In this paper we have reviewed various soft computing techniques. A review of application of these soft-computing techniques in malware detection has also been presented in this paper. Despite so much research, techniques with good accuracy and low false alarm rate are still needs attention.

Keywords
malware, malware detection, soft computing, machine learning, anomaly detection.

1. INTRODUCTION
Malware includes viruses, worms, trojan horses, spy-ware, and adware. A virus is a computer program that attaches itself to a host (e.g., a program file or a hard disk boot record) and spreads when the infected host is moved to a different computer. In August 2010, Postini [1] blocked 188 million viruses. A worm is a computer program that can replicate itself and spread across a network. A trojan horse appears to be a legitimate computer program but has malicious code hiding inside which runs when activated. Spy-ware is malware that collects and sends data copied from the victim's computer, such as financial data, personal data, passwords, etc. Adware, or advertising- supported software, is a computer program that automatically displays ads.

In recent years, network attacks are easy to launch since the tools to execute the attack are freely available on the Internet. Even the novice script users can initiate a sophisticated attack with basic knowledge on network and software technology. To overcome this matter, Intrusion Detection System (IDS) has been used as a vital instrument in defending the network from this malicious activity. With the ability to analyse network traffic and recognize incoming and on-going network attack, majority of network administrator has turn to IDS to help them in detecting anomalies in network traffic [2]. An intrusion detection system dynamically monitors the events taking place in a system, and decides whether these events are symptomatic of an attack or constitute a legitimate use of the system. Figure 1 shows the basic components of IDS. Intrusion model is the component which process information for malware detection. Various soft computing and machine learning approaches are employed in intrusion model for detection of possible attacks. This paper is divided into four different sections. Section-I gives the introduction of the topic. Section-II gives the review of soft computing and machine learning techniques. Various soft-computing techniques adopted for malware detection have been reviewed in section-III. Finally, in section-IV conclusion and future scope in the area of applying soft-computing to detect malware have discussed.

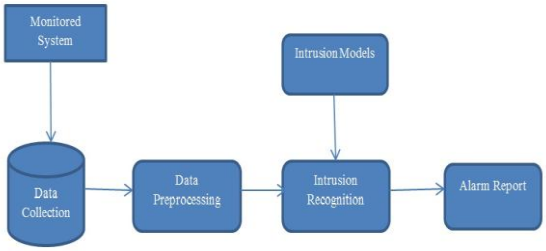

Fig.1. A typical Intrusion detection system

2. SOFT COMPUTING AND MACHINE LEARNING
Soft computing embraces several computational intelligence methodologies, including artificial neural networks, fuzzy logic, evolutionary computation, probabilistic computing, and recently it is extended towards artificial immune systems, belief networks, etc. These members neither are independent of one another nor compete with one another. Rather, they work in a cooperative and complementary way. There are various soft computing and machine learning techniques which are used in malware detection. Some of popular techniques are:

2.1 Artificial Immune System (AIS)
has drawn significant attention as a potential source of inspiration for novel approaches to solve complex computational problems. AIS has feature like highly distributed, adaptive, and self-organizing in nature. Using learning, feature extraction, and pattern recognition, it offers rich metaphors for its artificial counterpart. Unlike other engineered systems, AISs require both immunology and engineering to learn from each other through working in an interdisciplinary manner. A collaborative effort of

several interdisciplinary research scientists has produced a prolific amount of immune inspired algorithms by extracting or gleaning useful mechanisms from the immune system theories, processes and elements [3].

2.2 Fuzzy Set

Fuzzy logic, as a means of modelling the uncertainty of natural language, constructs more abstract and flexible patterns for intrusion detection, and thus greatly increases the robustness and adaptation ability of detection systems. Two research directions are currently active in the fuzzy logic area: (i) algorithms with learning and adaptive capabilities with the purpose of automatically designing fuzzy rules. Popular methods includes association rules, decision trees, evolutionary computation, and artificial neural networks; (ii) to enhance the understand-ability and readability of some machine learning algorithms, such as Support Vector Machine (SVM) or Hidden Markov Model (HMM). The use of fuzzy logic smooths the abrupt separation of normality and abnormality [4].

2.3 Artificial Neural Network (ANN)

learns to predict the behaviour of the various users and daemons in the system. If properly designed and implemented, ANN have the potential to address many of the problems encountered by rule-based approaches. The main advantage of ANN is their tolerance to imprecise data and uncertain information and their ability to infer solutions from data without having prior knowledge of the regularities in the data. This in combination with their ability to generalize from learned data has made them an appropriate approach to Intrusion Detection (ID). In order to apply this approach to ID, it is required to introduce data representing attacks and non-attacks to the ANN to adjust automatically coefficients of this Network during the training phase. ANN can be used in supervised or unsupervised ways [5].

2.4 Decision Tree

are powerful and popular tool for classification and prediction. A decision tree has three main components: nodes, arcs and leaves. Each node is labelled with a feature attribute which is most informative among the attributes not yet considered in the path from the root. Each arc out of a node is labelled with a feature value for the node's feature and each leaf is labelled with a category or class. A decision tree can then be used to classify a data point by starting at the root of the tree and moving through it until a leaf node is reached. The leaf node would then provide the classification of the data point [5].

2.5 Support Vector Machine (SVM)

represent a relatively new supervised learning technique suitable for solving classification problems with high dimensional feature space and small training set size. Although the basic technique was conceived for binary classification, several methods for single and multi-class problems have been developed. Being a supervised method, it relies on two phases: during the training phase, the algorithm ''acquires knowledge" about the classes by examining the training set that describes them. During the evaluation phase, a classification mechanism examines the evaluation set and associates its members to the classes that are available. During the training phase, the target of the algorithm is the estimation of boundaries between the classes described by the samples in the training sets. To describe the method with a very simple example one can think of a two class problem where a single regular surface perfectly divides the features space in two regions, each one fully representative of the corresponding class. Then an exact boundary can be determined and no errors will be reported during the classification phase. This is not always possible and a trade-off between the complexity of the boundary and the error rate must be chosen. The usefulness of SVM has been already demonstrated in several fields: like pattern recognition, where it can provide optimal statistical classification by means of properly chosen decision functions. SVM have recently been applied to identify and counter network DoS attacks showing very high accuracy [6].

2.6 Other Soft Computing Techniques

There are many soft computing based techniques like Genetic algorithms, Evolutionary algorithms, Swarm intelligence based techniques which can also be used for malware detection [7].

3. SOFT COMPUTING IN MALWARE DETECTION

Malwares detection can be broadly classified in to three categories:

- Misuse/signature Detection
- Anomaly Detection
- Hybrid Detection

3.1 Misuse/signature Detection

Signature-based detection schemes recognize intrusions by matching observed data with predefined descriptions of intrusive behavior. Therefore, assign a signature database corresponding to known attacks is specified a priori [8]. Researchers have developed many techniques based on signature. Some are:

3.1.1 Signature Verification with SVM

Network-based length-based signature generator (LESG) is designed for the worms exploiting buffer overflow vulnerabilities. The signatures generated are intrinsic to buffer overflows, and are very difficult for attackers to evade. They further prove the attack resilience bounds even under worst-case attacks with deliberate noise injection [9].

3.1.2 Signature tree generation

Network-based signature generation (NSG) has been proposed as a way to automatically and quickly generate accurate signatures for worms, especially polymorphic worms. An NSG system, PolyTree, defends against polymorphic worms. Signatures from worms and their variants are relevant and a tree structure can properly reflect their familial resemblance. Hence, in contrast to an isolated view of generated signatures in previous approaches, PolyTree organizes signatures extracted from worm samples into a tree structure, called signature tree, based on the formally defined "more specific" relation of simplified regular expression signatures. PolyTree is composed of two components: signature tree generator and signature selector [10].

3.1.3 Logical expression testing criteria

Decision Support has problems in accurately identifying attacks. Therefore, it is important to precisely find out conditions under which IDSs accurately identify attacks or fail to do so. However, no systematic approach has so far been defined and used to study this problem. Recognizing that signatures in essence

provide the specification of an IDS engine, studying the accuracy of an IDS engine becomes a black-box testing problem *[11]*.

3.1.4 F-Sign

F-Sign is designed for automatic extraction of unique signatures from malware files. It is primarily intended for high-speed network traffic filtering devices that are based on deep-packet inspection. Malicious executable is analysed using two approaches: disassembly, utilizing IDA-Pro, and the application of a dedicated state machine in order to obtain the set of functions comprising the executable. The signature extraction process is based on a comparison with a common function repository. By eliminating functions appearing in the common function repository from the signature candidate list, it can minimize the risk of false-positive detection errors. To minimize false-positive rates even further, it proposes intelligent candidate selection using an entropy score to generate signatures. [12]

3.1.5 Bayesian Network

are used for intrusion detection in computer networks. Bayesian system for intrusion detection (Basset) extends functionality of Snort, an open-source network intrusion detection system (NIDS), by incorporating Bayesian networks as additional processing stages. The flexible nature Bayesian network allows it to be used both for misuse-based and anomaly-based detection process. The ultimate goal is to provide better detection capabilities and less chance of false alerts by creating a platform capable of evaluating Snort alerts in a broader context – other alerts and network traffic in general. [13].

3.1.6 Semantic aware signature generation

String extraction and matching techniques have been widely used in generating signatures for worm detection, but how to generate effective worm signatures in an adversarial environment still remains a challenging problem. For example, attackers can freely manipulate byte distributions within the attack payloads and thus inject well-crafted noisy packets to contaminate the suspicious flow pool. Semantics Aware Statistical (SAS) algorithm have been proposed for automatic signature generation. When SAS processes packets in a suspicious flow pool, it uses data flow analysis techniques to remove non-critical bytes. Hidden Markov model (HMM) is further used to the refined data to generate state-transition-graph-based signatures. [14].

3.2 Anomaly Detection

Anomaly Intrusion Detection (ID's) strategy considers abnormal behaviour is rare and tries to model normal rather than anomalous behaviour. These detectors generate an anomaly alarm whenever the deviation between a given observation at an instant and the normal behaviour exceeds a predefined threshold. Another possibility is to model the "abnormal" behaviour of the system and to raise an alarm when the difference between the observed behaviour and the expected one falls below a given limit [8]. Researchers do a lot of work in anomaly detection. In [15], a system is proposed for new extracted features for host-based intrusion detection based on three viewpoints of system activity such as dimension, structure, and contents. Specification based anomaly detection technique allows automatic development of the normal and abnormal behavioural specifications in a form of variable-length patterns classified via anomaly-based approach. Specifically, they use machine-learning algorithm to classify fixed-length patterns generated via sliding window technique to infer the classification of variable length patterns from the aggregation of the machine learning based classification results [16]. Sequential anomaly detection based approach is used on temporal-difference learning [17] A novel approach to network-based anomaly detection based on the analysis of non-stationary properties and "hidden" recurrence patterns occurring in the aggregated IP traffic flows. In the observation of the above transition patterns for detecting anomalous behaviours, Authors in [18] adopted recurrence quantification analysis, a nonlinear technique widely used in many science fields to explore the hidden dynamics and time correlations of statistical time series.

In [19], a scheme is introduced to achieve early attack detection and filtering for the application-layer-based DDoS attack. An extended hidden semi-Markov model is proposed to describe the browsing behaviours of web surfers. In order to reduce the computational amount introduced by the model's large state space, a novel forward algorithm is derived for the online implementation of the model based on the M-algorithm. Entropy of the user's HTTP request sequence fitting to the model is used as a criterion to measure the user's normality. In [20], two different approaches are presented to characterize traffic: (i) a model-free approach based on the method of types and Sanov's theorem, and (ii) a model-based approach modelling traffic using a Markov modulated process. Using these characterizations as reference traffic is continuously monitored and large deviations is used and decision theory results to "compare" the empirical measure of the monitored traffic with the corresponding reference characterization, thus, identifying traffic anomalies in real-time. The model checking is used —a well-established software verification technique—for proactive malware detection. A tool is used that extracts an annotated control flow graph from the binary and automatically verifies it against a formal malware specification [21]. In [22], reputation establishment—one for systems solely consisting of smart insiders and the other for systems in which both smart insiders and naïve attackers are present is proposed .Some other techniques proposed by researchers are Feature-Aided Tracking with Hidden Markov Models [23]. Histogram-Based Traffic Anomaly Detection [24]. Anomaly Detection through a Bayesian Support Vector Machine [25]. Bivariate parametric detection mechanism (bPDM) that uses a sequential probability ratio test, allowing for control over the false positive rate while examining the trade-off between detection time and the strength of an anomaly [26]. Transaction-Pattern-Based Anomaly Detection Algorithm for IP Multimedia Subsystem [27]. BrowserGuard is a Behaviour-Based Solution to Drive-by-Download Attacks [28]. Out of proposed techniques some are statistical based on univariate, multivariate and time series model. Some techniques are knowledge based like Finite state machine (FSM), description language, expert system etc. Some techniques are machine learning based like Bayesian network, markov model, neural network, fuzzy logic genetic algorithm, clustering and outlier detection etc. [29][30]In [31], an approach of integrating Data Mining and Natural Language Processing, namely, the N-Gram_Square root Term Frequency-Inverse Document Frequency (N-Gram_STF-IDF) is proposed in order to detect masquerading.In [32], researcher suggests hierarchical hidden Markov model (HHMM) is used to represent a temporal profile of normal behaviour in a computer system. In [33], HMMPayl, an Intrusion detection System (IDS) is proposed,

where the payload is represented as a sequence of bytes, and the analysis is performed using Hidden Markov Models (HMM). In [34] a system based on the receiver operating characteristic (ROC) is proposed to efficiently adapt ensembles of HMMs (EoHMMs) in response to new data, according to a learn-and-combine approach. When a new block of training data becomes available, a pool of base HMMs is generated from the data using a different number of HMM states and random initializations.Fuzzy measures and fuzzy sets are used in [35] to design simple and robust alert aggregation algorithms for reducing false positives in anomaly detectors through fuzzy alert aggregation. In [36], an approach, called FC-ANN, based on ANN and fuzzy clustering is used to solve the problem and help IDS achieve higher detection rate, less false positive rate and stronger stability. The general procedure of FC-ANN is as follows: firstly fuzzy clustering technique is used to generate different training subsets. Subsequently, based on different training subsets, different ANN models are trained to formulate different base models. Finally, a meta-learner, fuzzy aggregation module, is employed to aggregate these results. In [37], artificial immune system and genetic algorithm based approaches are suggested like vEye, a behavioural foot printing for self-propagating worm detection and profiling, A method for utilizing a genetic algorithm for sparse trees to detect anomalies, using genetic algorithm for the meta-learning step, on labelled vectors of statistical classifiers is proposed in [38]. Each of the statistical classifiers was a 2-bit binary encoding of the abnormality of a particular feature, ranging from normal to dangerous. In [39], researcher use decision tree based light weight intrusion detection using a wrapper approach and then a neuro tree model is employed as the classification engine.In [40], nearest neighbour based approach is used which is a learning model works on the triangle area based nearest neighbours (TANN) in order to detect attacks more effectively. In [41], genetic weighted KNN (K-nearest-neighbour) classifiers is used for anomaly detection. In [42], weighted k-nearest-neighbour classifiers are also used in anomaly detection systems for Denial-of-Service attacks in real time.

3.3 Hybrid Detection

Signature based detection system can only detect attacks which are known to system and signatures are defined. Anomaly based detection has higher False-Positive rate. These techniques can be combined to give better results. This combined technique is known as Hybrid detection scheme [8].In [43], a hybrid model is used not only to correlate alerts as accurately and efficiently as possible but also to be able to boost the model in the course of time. The model consists of two parts: (1) an attack graph-based method to correlate alerts raised for known attacks and hypothesize missed alerts and (2) a similarity-based method to correlate alerts raised for unknown attacks which cannot be correlated using the first part and also to update the attack graph. These two parts cooperate with each other such that if the first part could not correlate a new alert, the second part is applied. A hybrid model based on improved fuzzy and data mining techniques, which can detect both misuse and anomaly attacks is proposed in [44]. A combination of misuse detection system with anomaly detection system (ADS) is proposed in [45]. The hybrid intrusion detection system (HIDS) contains three sub-modules: misused detection module, anomaly detection module and signature generation module. The basis of misused detection module is snort. Anomaly detection module is constructed by using frequent episode rule. And signature generation module is based on a variant of Apriori algorithm. In [46], a hybrid intrusion detection system based on protocol analysis and decision tree algorithms is presented. Performance evaluation of the proposed system is conducted using Generalized Stochastic Petri Nets (GSPN). Simulation results show that this hybrid system can reach a high detection rate. In [47], an approach of combination of DBSCAN with anomaly detection is also proposed.

4. CONCLUSION AND FUTURE SCOPE

The network data is very large, heterogeneous, highly varying and imbalanced. Most of the available machine learning approaches developed for uniformly distributed data and doesn't emphasize on these characteristics of network data that this data is not normally distributed in equal classes. As volume of network traffic data is very huge and has large number of attributes it is practically impossible to run classic machine learning algorithm on whole data. Rather sampling, feature extraction and selection needs to be performed. But these operations may change overall character of data. Some good and accurate approaches for feature selection, extraction and sampling are required. Real time anomaly detection approaches with high accuracy and low false positive rate are required. In future, the work can be carried out in this direction.

5. REFERENCES

[1] Won Kim n, Ok-RanJeong, Chulyun Kim and Jungmin So, "The dark side of the Internet : Attacks, costs and responses", Elsevier's Journal of Information Systems, Volume 36, Issue 3, May 2011, pp 675-705.

[2] Faizal, M.A. Mohd, Z.M. Sahib, S. Robiah, Y. Siti, R.S. and Asrul, H.Y., "Time Based Intrusion Detection on Fast Attack for Network Intrusion Detection System", Second International Conference on Network Applications Protocols and Services (NETAPPS), Kedah, Malaysia, 22-23 Sept. 2010, pp 148 - 152.

[3] Dipankar Dasgupta, Senhua Yu and Fernando Nino, "Recent Advances in Artificial Immune Systems: Models and Applications Review", Journal of Applied Soft Computing, Volume 11, Issue 2, March 2011, pp 1574-1587.

[4] Shelly Xiaonan Wu and Wolfgang Banzhaf," The use of computational intelligence in intrusion detection systems: A review", Elsevier's Journal of Applied Soft Computing, volume 10, issue 1 , January 2010, pp 1-35.

[5] Gulshan Kumar, Krishan Kumar and Monika Sachdeva , "The use of artificial intelligence based techniques for intrusion detection: a review", Journal of Artificial Intelligence Review, 2010, Volume 34, Number 4, pp 369-387.

[6] Alice Este, Francesco Gringoli and Luca Salgarelli, "Support Vector Machines for TCP traffic classification", Journal of Computer Networks, Volume 53, Issue 14, 18 September 2009, pp 2476-2490.

[7] Chet Langin and Shahram Rahimi, "Soft computing in intrusion detection: the state of the art", Journal of Ambient Intelligence and Humanized Computing, 2010, Volume 1, Number 2, pp 133-145.

[8] Huwaida Tagelsir Elshoush and Izzeldin Mohamed Osman, " Alert correlation in collaborative intelligent intrusion detection systems—A survey", Elsevier's Journal of Applied Soft Computing, volume 11, issue 7 , October 2011, pp 4349-4365.

[9] Lanjia Wang, Zhichun Li, Yan Chen, Zhi (Judy) Fu, and Xing Li, " Thwarting Zero-Day Polymorphic Worms With Network-Level Length-Based Signature Generation", IEEE/ACM Transactions on networking, Vol. 18, No. 1, February 2010, pp 53-66.

[10] Yong Tang, Bin Xiao and Xicheng Lu, "Signature Tree Generation for Polymorphic Worms", IEEE Transaction on computers, Vol. 60, No. 4, April 2011, pp 565-579.

[11] Frederic Massicotte and Yvan Labiche, "Specification-Based Testing of Intrusion Detection Engines using Logical Expression Testing Criteria", 10th International Conference on Quality Software, 14-15 July 2010, Zhangjiajie, China , pp 102-111.

[12] Asaf Shabtai, Eitan Menahem, and Yuval Elovici, "F-Sign: Automatic, Function-Based Signature Generation for Malware", IEEE Transaction on system, man and —Part C: Applications and Reviews, Vol. 41, No. 4, July 2011, pp 494-508.

[13] Wojciech Tylman, " Misuse-based intrusion detection using Bayesian networks", Int. J. Critical Computer-Based Systems, Vol. 1, Nos. 1/2/3, 2010, pp 178-190.

[14] Deguang Kong, Yoon-Chan, Jhi Tao Gong, Sencun Zhu, Peng Liu, and Hongsheng Xi, "SAS: semantics aware signature generation for polymorphic worm detection", International Journal of Information Security, Online First, 21 May 2011 and Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, 1, Volume 50, Security and Privacy in Communication Networks, Part 1, pp 1-19.

[15] Inho Kang and Myong K.," JeongA differentiated one-class classification method with applications to intrusion detection", Expert Systems with Applications, In Press, Uncorrected Proof, Available online 6 July 2011, No. of pp 7.

[16] Natalia Stakhanova, Samik Basu and Johnny Wong, "On the symbiosis of specification-based and anomaly-based detection", Elsevier's journal of Computers & Security, Volume 29, Issue 2, March 2010, pp 253-268.

[17] Xin Xu, "Sequential anomaly detection based on temporal-difference learning: Principles, models and case studies", Elsevier's Journal of Applied Soft Computing, Volume 10, Issue 3, June 2010, pp 859-867.

[18] Francesco Palmieri and Ugo Fiore, "Network anomaly detection through nonlinear analysis ", Elsevier's Journal of Computers & Security, Volume 29, Issue 7, October 2010, pp 737-755.

[19] Yi Xie and Shun-Zheng Yu, "A Large-Scale Hidden Semi-Markov Model for Anomaly Detection on User Browsing Behaviors", IEEE/ACM Transaction on networking, Vol. 17, No. 1, February 2009, pp 54-65.

[20] Ioannis Ch. Paschalidis and Georgios Smaragdakis, "Spatio-Temporal Network Anomaly Detection by Assessing Deviations of Empirical Measures", IEEE/ACM Transaction on networking, Vol. 17, No. 3, June 2009, pp 685-697.

[21] Johannes Kinder, Stefan Katzenbeisser, Christian Schallhart, and Helmut Veith, "Proactive Detection of Computer Worms Using Model Checking", IEEE Transaction on dependable and secure computing, Vol. 7, No. 4, October-December 2010, pp 424-438.

[22] Nan Zhang, Wei Yu, Xinwen Fu, and Sajal K. Das, "Maintaining Defender's Reputation in Anomaly Detection Against Insider Attacks", IEEE Transaction on system, man and cybernetics—Part B: Cybernetics, Vol. 40, NO. 3, June 2010, pp 597-611.

[23] Satnam Singh, Haiying Tu, William Donat, Krishna Pattipati and Peter Willett, "Anomaly Detection via Feature-Aided Tracking and Hidden Markov Models", IEEE Transactions on system, man and cybernetics—Part A: System and humans, VOL. 39, NO. 1, January 2009, pp 144-159.

[24] Andreas Kind, Marc Ph. Stoecklin, and Xenofontas Dimitropoulos, "Histogram-Based Traffic Anomaly Detection" IEEE Transaction on network service management, Vol. 6, No. 2, June 2009, pp 110-121.

[25] Vasilis A. Sotiris, Peter W. Tse, and Michael G. Pecht, "Anomaly Detection Through a Bayesian Support Vector Machine", IEEE Transaction on reliability, Vol. 59, No. 2, June 2010, pp 277-286.

[26] Gautam Thatte, Urbashi Mitra, and John Heidemann, "Parametric Methods for Anomaly Detection in Aggregate Traffic" IEEE/ACM Transaction on networking, Vol. 19, No. 2, April 2011 pp 512-525.

[27] Chi-Yuan Chen, Kai-Di Chang, and Han-Chieh Chao, "Transaction-Pattern-Based Anomaly Detection Algorithm for IP Multimedia Subsystem" IEEE Transactions on information forensics and security, Vol. 6, No. 1, Msrch 2011, pp 152-161.

[28] Fu-Hau Hsu, Chang-Kuo Tso, Yi-Chun Yeh, Wei-Jen Wang, and Li-Han Chen, "BrowserGuard: A Behavior-Based Solution to Drive-by-Download Attacks" IEEE Journal on selected areas in communications, Vol. 29, No. 7, August 2011, pp 1461-1468.

[29] P. Garcıa-Teodoro, J. Dıaz-Verdejo, G. Macia-Fernandez, and E. Vazquez, "Anomaly-based network intrusion detection: Techniques, systems and challenges" Elsevier's journal of Computers & Security, Volume 28, Issues 1-2, February-March 2009, pp 18-28.

[30] Prasanta Gogoi, B Borah and D K Bhattacharyya, "Anomaly Detection Analysis of Intrusion Data using Supervised & Unsupervised Approach",Journal of

Convergence Information Technology Volume 5, Number 1, February 2010, pp 95-110.

[31] Dai Geng and Thmohiro Odaka, Jousuke Kuroiwa and Hisakazu Ogura, "An N-Gram and STF-IDF model for masquerade detection in a UNIX environment", Journal in Computer Virology, 2011, Volume 7, Number 2, pp 133-142.

[32] Chunfu Jia and Feng Yang, "An intrusion detection method based on hierarchical hidden Markov models", Wuhan University Journal of Natural Sciences, 2007, Volume 12, Number 1, pp 135-138.

[33] Davide Ariu, Roberto Tronci and Giorgio Giacinto, " HMMPayl: An intrusion detection system based on Hidden Markov Models", Elsevier's Journal of Computers & Security, Volume 30, Issue 4, June 2011, pp 221-241.

[34] Wael Khreich, Eric Granger, Ali Miri, and Robert Sabourin, "Adaptive ROC-based ensembles of HMMs applied to anomaly detection", Elsevier's journal of Pattern Recognition, Volume 45, Issue 1, January 2012, pp 208-230.

[35] Federico Maggi, Matteo Matteucci and Stefano Zanero, " Reducing false positives in anomaly detectors through fuzzy alert aggregation", Elsevier's journal of Information Fusion, Volume 10, Issue 4, October 2009, pp 300-311.

[36] Gang Wang, Jinxing Hao, Jian Ma and Lihua Huang, " A new approach to intrusion detection using Artificial Neural Networks and fuzzy clustering", Elsevier's journal of Expert Systems with Applications, Volume 37, Issue 9, September 2010, pp 6225-6232.

[37] Xuxian Jiang and Xingquan Zhu, " vEye: behavioral footprinting for self-propagating worm detection and profiling", Journal of Knowledge and Information Systems, 2009, Volume 18, Number 2, pp 231-262.

[38] Gulshan Kumar, Krishan Kumar and Monika Sachdeva, " The use of artificial intelligence based techniques for intrusion detection: a review", Journal of Artificial Intelligence Review, 2010, Volume 34, Number 4, pp 369-387.

[39] Siva S. Sivatha Sindhu, S. Geetha and A. Kannan, " Decision tree based light weight intrusion detection using a wrapper approach", Journal of Expert Systems with Applications, In Press, Corrected Proof, Available online 12 July 2011. No. of pp 13.

[40] Chih-Fong Tsai and Chia-Ying Lin, " A triangle area based nearest neighbors approach to intrusion detection", Elsevier's Journal of Pattern Recognition, Volume 43, Issue 1, January 2010, pp 222-229.

[41] Ming-Yang Su, " Using clustering to improve the KNN-based classifiers for online anomaly network traffic identification", Journal of Network and Computer Applications, Volume 34, Issue 2, March 2011, pp 722-730.

[42] Ming-Yang Su, "Real-time anomaly detection systems for Denial-of-Service attacks by weighted k-nearest-neighbor classifiers", Elsevier's Journal of Expert Systems with Applications, Volume 38, Issue 4, April 2011, pp 3492-3498 .

[43] Seyed Hossein Ahmadinejad, Saeed Jalili and Mahdi Abadi, "A hybrid model for correlating alerts of known and unknown attack scenarios and updating attack graphs" Elsevier's journal of Computer Networks, Volume 55, Issue 9, 23 June 2011, pp 2221-2240.

[44] Bharanidharan Shanmugam and Norbik Bashah Idris, "Improved Intrusion Detection System using Fuzzy Logic for Detecting Anamoly and Misuse type of Attacks", International Conference of Soft Computing and Pattern Recognition, Malacca, Malaysia, December 4-7, 2009, pp 212-217.

[45] Yu-Xin, Min Xiao and Ai-Wu Liu, " Research and implementation on snort based gybrid intrusion detection system", Proceedings of the Eighth International Conference on Machine Learning and Cybernetics, Baoding, 12-15 July 2009, pp 1414-1418.

[46] Jie Yang, Xin Chen, Xudong Xiang and Jianxiong Wan, "HIDS-DT: An Effective Hybrid Intrusion Detection System Based on Decision Tree", International Conference on Communications and Mobile Computing, April 12-14, 2010, Shenzhen Guest House, Shenzhen, China, pp 70-75.

S. I. Handra and H. Ciocârlie, "Anomaly Detection in Data Mining. Hybrid Approach between Filtering-and-Refinement and DBSCAN" 6th IEEE International Symposium on Applied Computational Intelligence and Informatics • May 19–21, 2011 • Timioara, Romania, pp 75-83. Timioara, Romania, pp 75-