Performance Analysis of Intrusion Detection Systems Implemented using Hybrid Machine Learning Techniques

Purushottam R. Patil
Ph.D Research Scholar
CSE, Faculty of Engg. & Tech.
Jodhpur National University
RJ, INDIA

Yogesh Sharma, PhD
Dean & Professor (Maths.)
Faculty of Computer Appl.
Jodhpur National University
RJ, INDIA

Manali Kshirasagar, PhD
Vice President (Academic)
ADCC Infocad Ltd.
IT Park, Nagpur
MS, INDIA

ABSTRACT
Intrusion Detection System (IDS) are said to be more effective when it has both high intrusion detection (true positive) rate and low false alarm (false positive). But current IDS when implemented using data mining approach like clustering, classification alone are unable to give 100 % detection rate hence lack effectiveness. In order to overcome these difficulties of the existing systems, many researchers implemented intrusion detection systems by integrating clustering and classification approach like k-means and Fuzzy logic, K-means and genetic algorithm, some of the researcher also tried use of Decision tree and Neural Network to detect unknown attacks. In this paper analysis of such Hybrid systems which are implemented by using the benchmark dataset compiled for the 1999 KDD intrusion detection contest, by MIT Lincoln Labs.

Keywords
Intrusion detection system (IDS), Detection rate in IDS, False alarm Rate, Classification, Prediction, MIT KDD'99 dataset.

1. INTRODUCTION
1.1 Intrusion Detection System (IDS)
It identifies known and unknown attacks on a communication network and takes necessary actions for the systems network connections. They are the set of approaches that are used to detect suspicious activity at network and host level. An attacker places networks or hosts in jeopardy, without intruding into the hosts [1]. The attacks on a famous website, such as Yahoo, E-bay, and E*TRADE, are good examples [2].

This type of Denial of Service (DoS) and Distributed Denial of Service (DDoS) attacks [3, 4, 5] will cause more damage for the following reasons. There are so many DoS/DDoS tools that even unskilled users can use easily.

The accurate and rapid detection of network traffic anomaly is crucial to enhance the effective operation of a network. It is often difficult to detect the time when the faults occur in a network. A successful DoS/DDoS attack shows its impact quickly and makes it difficult to trace back to the intruder. Moreover, the bandwidth consumption by the attacks influences network performance. Even on highly over-provisioned links, malicious traffic causes an increase in the average DNS latency by 230% and an increase in the average web latency by 30% [6]. The monitoring result of NG-MON [7], there is more serious latency deficiency (up to 500%) in enterprise networks that contains a target or bypassing machine of the attacks. These menaces need to make provisions against DoS/DDoS attacks.

IDS Design approaches includes a) Misuse based or Signature based and b) Anomaly based. In a misuse based IDS, intrusions are detected by looking for activities that correspond to know signatures of intrusions or vulnerabilities occurs whenever [3]. While an anomaly based IDS detect intrusions by searching for anomalous network traffic. The anomalous traffic pattern can be defined either as the violation of accepted verge for frequency of events in a connection or as a user's violation of the consistent profile developed for normal behavior.

An anomaly detection approach generally consists of two steps: the first is called training phase wherein a normal traffic profile is generated; the second is called anomaly detection, here the studied profile is applied to the current traffic to look for any deviations. In Literature many anomaly detection mechanisms has been proposed to detect such deviations, which can be classified into data-mining methods, statistical methods and machine learning based methods.

Therefore, the goal of this paper is to analyse 13 related systems published by examining what methodology have been used for software implementation and what should be considered for future work.

This paper is organized as follows. Section 2 provides an overview of machine learning approach and briefly compiles a number of related approaches for intrusion detection. Section 3 Analysis of related work based on the types of classifier design, the chosen baselines etc. Conclusion and discussion for future research are given in Section 4.

2. MACHINE LEARNING TECH.
2.1 Classification and Prediction
As per data mining literature Classification is two-step process Learning and Classification

a) Learning: Training data are analyzed by a classification algorithm and the learned model or classifier is described in the form of classification rules.

b) Classification: Test data are used to measure the accuracy of the classification rules. If the accuracy is considered acceptable, the rules can be applied to the classification of new data tuples because the class label of each training tuple is provided, this step is also known as supervised learning. It distinction with unsupervised learning (clustering), in which the class label of each training tuple is not known, and the number or set of classes to be learned may not be known in prior. The certainty of

a classifier on a given test set is the percentage of test set tuples that are correctly classified by the classifier.

2.1.1 Decision trees

A decision tree takes a specimen through a sequence of decisions, in which the current decision helps to make the future decision. Such a sequence of decisions is represented in form of trees. The classification of a specimen proceeds from the root node to a suitable end leaf node, where each end leaf node represents a classification group. The attributes of the specimens are assigned to each node, and the value of each branch is corresponding to the. Examples are CART (Classification and Regressing Tree), C4.5, ID3 [20].

2.1.2 K-Nearest Neighbor ((k-NN)

It is simplest and conventional nonparametric approach to classify specimens. It computes the approximate distances between different points on the input vectors, and then nominates the unlabeled point to the class of its K-nearest neighbors. In the process of create k-NN classifier, k is an important parameter and different k values will cause different conducts. If k is considerably huge, the neighbors which used for prediction will make large classification time and influence the accuracy of prediction. It is called precedent based learning, and it is different from the preparatory learning approach. Thus, it does not contain the model training stage, but only searches the examples of input vectors and classifies precedents. Therefore, k-NN ‘‘on-line” trains the examples and finds out k-nearest neighbor of the precedent [16].

2.1.3 Artificial Neural Networks (ANN)

The neural network is information processing units which to mimic the neurons of human brain. Multilayer feed forward is the widely used architecture in many pattern recognition problems. A MLP network consists of an input layer including a set of sensual nodes as input nodes, one or more hidden layers of computation nodes, and an output layer of computation nodes. Each interconnection has associated with it a scalar weight which is adjusted during the training phase.

2.1.4 Support Vector Machines (SVM)

SVM first maps the input vector into a higher dimensional feature space and then obtain the optimal separating hyper-plane in the multidimensional feature space. Moreover, a decision boundary, i.e. the separating hyper-plane, is determined by support vectors rather than the whole training specimens and thus is extremely robust to outliers. It mainly designed for binary classification. The SVM also provides a user specified criterion called penalty factor. It allows users to make a concession between the number of misclassified specimens and the width of a decision boundary.

2.1.5 Genetic Algorithms (GA)

An Evolutionary technique uses the computer to implement the natural selection and evolution. This concept comes from the ‘‘adaptive survival in natural organisms”. The algorithm starts by randomly generating a large population of candidate programs. Some type of fitness measure to evaluate the conduct of each individual in a population is used. A large number of iterations are performed to select fittest chromosomes. Crossover and Mutation operation makes recombination for new population [20].

2.1.6 Rough Set Approach

It can be used for classification to explore structural links within estimated or noisy data. Therefore continuous-valued attributes must be normalizing before its use. Rough set theory is based on the establishment of equivalence classes within the given training data. All of the data tuples forming an equivalence class are obscure, that is, the specimens are identical with respect to the attributes representing the data.

2.1.7 Fuzzy Logic (FL)

It is based on the concept of the fuzzy aspect to occur frequently in reality. FL considers the set membership values for inference and the values range between 0 and 1. That is the degree of truth of a statement can range between 0 and 1 and it is not constrained to the two truth values (i.e. true, false).

2.2 Cluster Analysis

It groups objects either physical or abstract into classes of similar objects. A cluster is a set of data objects that are similar to each other within the same cluster and are dissimilar to the objects in other clusters. A cluster can be treated generally as one group and so may be considered as a form of data compression. Clustering can be classifies into partitioning method, hierarchical methods, density-based methods, grid-based methods, model-based methods.

2.2.1 A Partitioning Method

It starts with creation of initial set of k partitions, where r k is the number of partitions to. It then uses an iterative remotion approach that tries to improve the partitioning by moving objects from one group to another. Examples include k-means, k-Medoids, CLARANS, and their improvements [20].

2.2.2 A Hierarchical Method

This method first creates a hierarchical decomposition of the given set of data objects. It can be either bottom-up or top-down, based on how the hierarchical decomposition is done. To compensate for the rigor of merge or split, the quality of hierarchical cluster can be improved by analyzing object linkages at each hierarchical partitioning or by first performing micro clustering and then operating on the micro clusters with other clustering approaches, such as iterative relocation [20].

2.2.3 A Density-Based Method

It clusters objects based on the notion of density. It either grows clusters according to the density of neighborhood objects (e.g. DBSCAN) or according to some density function (e.g. DENCLUE). OPTICS is a density based method that generates an enhance ordering of the clustering structure of the data [20].

2.2.4 A Grid-Based Method

This method first quantizes the object space into a finite number of cells that form a grid structure, and then performs clustering on the grid structure (e.g. STING is a typical example on statistical information stored in grid cells. Wave Cluster and CLIQUE are two clustering algorithms that are both grid based and density-based [20].

2.2.5 A Model-Based Method

This method cerebrate a model for each of the clusters and finds the best fit of the data to that model (e.g. EM algorithm), conceptual clustering (e.g. COBWEB), and neural network approaches (e.g. SOM). In this method MLE (maximum likelihood estimation) is used to find the parameter inside the probability model. Since the probability function is a mixture

summation of a couple of probability function, it makes the conventional method infeasible to find the maximum value [25].

3. REVIEW OF RELATED WORK

In this section, authors have compiled 13 research articles. The software implementation platform or programming languages may be JAVA, MATLAB etc. but they all used benchmark dataset compiled for the 1999 KDD intrusion detection contest, by MIT Lincoln Labs. Some researchers tested systems on varying number of attributes; Numbers of attributes selected also affect the system performance. Average Detection Rate calculates Detection rate of Normal, Probe, DoS, U2R, R2L. Table 1 Is compilation of research articles in ascending order of year from 2004 to 2015

Table 1: Compilation of Hybrid Approaches for Anomaly Detection

Sr. No.	Ref. No.	Researchers Name , Year	Design Approach	Detection Rate (%)	False Alarm Rate (%)
1	9	Sampada Chavan, Khusbu Shah, Neha Dave, Ajith Abraham, Sugata sanyal **(2004)**	Evolving Fuzzy Neural Network	Avg. : 92	NA
2	10	Srilatha Chebrolu, Ajith Abraham, Johnson P. Thomas **(2005)**	Ensemble And Base Classifier	Normal, Probe, DoS : 100 U2R: 84, R2L: 99.47	NA
3	12	Dong Song, Malcolm A, NurZincir-Heywood **(2005)**	Genetic Programming	NA	NA
4	13	Adel Nadjaran Toosi, M. Kahani **(2007)**	Adaptive Neuro Fuzzy Inference System	NA	NA
5	15	K. M. Faraoun and A. Boukelif **(2007)**	MLFF-NN & K-Means	Avg. : 92	6.21
6	16	Ajith Abraham, Ravi Jain, Johnson Thomas, Sang Yong Hana **(2007)**	Fuzzy Rule-Based Classifiers	Avg. : 100	NA
7	17	Alireza Osareh, Bita Shadgar **(2008)**	Neural Network & SVM	NN(Avg.) : 66.3 SVM (Avg.) : 73.3	NN(Avg.): 1.62 SVM(Avg.): 0.92
8	18	Marjan B, E. Salahi, M. Khaleghi **(2009)**	Decision Tree & Som	Avg. : 97.13	NA
9	19	Vivek Patole, V. Pachghare, Dr. Parag Kulkarni **(2009)**	SOM	NA	NA
10	21	Hai Nguyen, Katrin Franke and Slobodan Petrovic **(2010)**	C4.5 & BayesNet	C4.5 (Avg): 99.41, BayesNet (Avg.) : 98.91	NA
11	22	P Jongsuebsuk, Wattanapongsakorn N , C. Charnsripinyo **(2013)**	Fuzzy Genetic Algorithm	Avg. : 97	NA
12	23	Sharmila Kishor Wagh, Dr.Satish R. Kolhe **(2014)**	Self Learning Semi supervised	DoS: 99.25,U2R:70, R2L: 66.66, Probe: 96.88	Avg.: 0.102
13	24	Samaneh Rastegari, Chiou-Peng Lam, Philip Hingston **(2015)**	ESR-NID GA-based Learning	Avg. : 98.4	NA

4. CONCLUSION

In this paper 13 research articles on Intrusion detection Systems implemented using clustering and classification hybrid techniques are analyzed. The Dataset used by all researchers is DARPA KDD'99. The detection rate of attack classes like Normal, Probe, DOS, U2R, R2L ranges from 66 % to 100 % and False Alarm rate is minimized up to 0.102. Number of attributes selection for processing also effect conduct of the system, lower the numbers of attributes better the conduct. In future hybrid of such machine learning approaches will achieve 100 % detection rate, False alarm rate to 0% hence more adaptive and efficient systems can be designed.

5. REFERENCES

[1] Myung-Sup Kim, Hun-Jeong Kang, Seong-Cheol Hong, Seung-Hwa Chung, and James W. Hong 2003 "A Flow-based Method for Abnormal Network Traffic Detection",

[2] CNN, Cyber-attacks batter web heavyweights, February 2000,http://www.cnn.com/2000/TECH/computing/02/09 cyber.attacks.01Drew Dean, Matt Franklin, and Adam Stubblefield

[3] 2001. "An algebraic approach to IP trace back," Proc. of Network and Distributed System Security Symposium, NDSS '01, San Diego, California.

[4] L. John Ioannidis and Steven M. Bellovin February 2002 ,"Implementing pushback: Router-based defense against DDoS attacks," Proc. of Network and Distributed System Security Symposium, NDSS '02, San Diego, California.

[5] Stefan Savage, David Wetherall, Anna Karlin, and Tom Anderson 2000 "Practical network support for IP traceback," Proc. of the 2000 ACM SIGCOMM, Stockholm, Sweden.

[6] Kun-chan Lan, Alefiya Hussain, and Debojyoti Dutta April 2003, "Effect of Malicious Traffic on the Network," Proc. of PAM 2003, San Diego, California,.

[7] Se-Hee Han, Myung-Sup Kim, Hong-Taek Ju, and James W. Hong 2002 "The Architecture of NG-MON: A Passive Network Monitoring System," Lecture Notes in Computer Science 2506, 13th IFIP/IEEE International Workshop on Distributed Systems: Operations and Management (DSOM 2002), Montreal, Canada.

[8] Dewan Md. Farid and Mohammad Zahidur Rahman 2010, "Anomaly Network Intrusion Detection Based on Improved Self Adaptive Bayesian Algorithm", Journal Of Computers, Vol. 5, No. 1.

[9] Sampada Chavan, Khusbu Shah, Neha Dave, Sanghan Mitra Mukherjee, Ajith Abraham and Sugata Sanyal 2004. "Adaptive Neuro-Fuzzy Intrusion detection Systems",ITCC'04, IEEE.

[10] Srilatha Chebrolu, Ajith Abraham and Johnson P. Thomas 2005," Feature deduction and ensemble design of intrusion detection systems", Computers & Security.

[11] Muna M. Taher Jawhar and Monica Mehrotra January-June 2010,"Anomaly Intrusion Detection System using Hamming Network Approach", International Journal of Computer Science & Communication, Vol. 1, No. 1.

[12] Dong Song, Malcolm I. Heywood, Nur Zincir-Heywood 2005,"Training Genetic Programming on Half a Million Patterns: An Example From Anomaly Detection, IEEE transactions on evolutionary computation,Vol. 9, No. 3.

[13] Adel Nadjaran Toosi and Mohsen Kahani, 2007 "A New approach to intrusion detection based on an evolutionary soft computing model using neuro-fuzzy classifiers", Elsevier B.V.

[14] Preeti Aggarwal,Sudhir Kumar Sharma, 2015 "Analysis of KDD Dataset Attributes-class wise For Intrusion Detection",Procedia Computer Science, Elsevier

[15] [K. M. Faraoun and A. Boukelif, 2005 "Neural Networks Learning Improvement using the K-Means Clustering Algorithm to Detect Network Intrusions" International Journal of Computational Intelligence" Vol. 3 .

[16] Ajith Abrahama, Ravi Jain, Johnson Thomas, Sang Yong Hana 2007 " D-SCIDS: Distributed soft computing intrusion detection system" Journal of Network and Computer Applications, pp.81–98.

[17] Alireza Osareh, Bita Shadgar 2008," Intrusion Detection in Computer Networks based on Machine Learning Algorithms", IJCSNS International Journal of Computer Science and Network Security, Vol.8, No.11,

[18] Marjan Bahrololum, Elham Salahi and Mahmoud Khaleghi, December 2009 "An Improved Intrusion Detection Approach based on two Strategies Using Decision Tree and Neural Network, "Journal of Convergence Information Technology Vol. 4, No. 4.

[19] Mr. Vivek A. Patole, Mr. V. K. Pachghare and Dr. Parag Kulkarni 2010 ,"Self Organizing Maps to Build Intrusion Detection System", International Journal of Computer Applications.

[20] Jiawai Han and Mitcheline Kamber 2006,"Data Mining Concepts and approachs",2e,Elsevier.

[21] .Hai Nguyen, Katrin Franke and Slobodan Petrovic 2010, " Improving Effectiveness of Intrusion Detection by Correlation Feature Selection",IEEE

[22] Jongsuebsuk P, Wattanapongsakorn N Charnsripinyo C. 2013," Real-time intrusion detection with fuzzy genetic algorithm", 10th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON), IEEE.

[23] Sharmila Kishor Wagh, Dr.Satish R. Kolhe 2014, "Effective Intrusion Detection System Using SemiSupervised Learning", IEEE.

[24] Samaneh Rastegari, Chiou-Peng Lam, Philip Hingston 2015 ,"A Statistical Rule Learning Approach to Network Intrusion Detection", IEEE.

[25] HaiJiang Steven Shi 2005," Model-based Clustering" University of Waterloo, Canada,.