Indian Journal of Science and Technology, *Vol 7(S6), 11–14, October 2014*

ISSN (Print) : 0974-6846
ISSN (Online) : 0974-5645

Heuristic Approach of Supervised Learning for Intrusion Detection

H. J Shanthi[1]* and E. A. Mary Anita[2]

[1]AMET University, Chennai, India; shanthi_harold@yahoo.co.in
[2]S.A.Engineering College, Chennai, India

Abstract

The objective of the study is to design a heuristic supervised learning algorithm, that locates Intrusion and improves performance of network. In this paper we have proposed the algorithm using existing supervised learning approach and evaluated IDS for MANETs. The trained dataset of KDDCUP is loaded and known four attacks are taken for evaluation. We assume that discussed four attacks dominate the network traffic. The algorithm is iterative in nature to produce optimum results. The performance of proposed supervised algorithm is evaluated under different network traffic and mobility patterns for Dos, PRB, R2L and U2R attacks. The results indicate high accuracy for almost all the four attacks. The proposed algorithms, results show high accuracy on discussed four attacks of KDD 99. It also produces low high positive rate.

Keywords: Intrusion Detection, Supervised Learning, MANET, Heuristic IDS Algorithm

1. Introduction

In Mobile Adhoc Network (MANET)[1] various Intrusion Detection System (IDS) strategies are available for the enhancement of network performance. The IDS system can be classified into two types: anomaly detection and misuse detection[2].

The anomaly detection in IDS focuses on detecting unusual pattern of activities in network traffic. The pattern recognition is based on behavior of users, network traffic, and resource with the normal patterns. The misuse detection is generally based on misused systems, signatures from known system policy. The pattern recognition with machine learning is used to detect the intrusions.

The learning can be unsupervised and supervised. In unsupervised learning, patterns are learnt based on statistical collection of inputs and its reflection in network. They are learnt from unlabeled examples. In supervised machine learning, labeled data is needed for training. The supervised learning is learnt from correct and labeled example for new task. In actual networks the availability of labeled data is costly, enormous amount of network, host data and expert in labeling is required.

Zhang and Lee[3] proposed the first (high-level) IDS approach specific for ad hoc networks. They proposed a distributed and cooperative anomaly-based ID, which provides an efficient guide for the design of IDS in wireless ad hoc networks.

Huang and Lee[4] extended their previous work by proposing cluster-based IDS, in order to combat the resource constraints that MANETs face. They use a set of statistical features that can be derived from routing tables and they apply the classification decision tree induction algorithm. The proposed system is able to identify the source of the attack, if the identified attack occurs within one-hop.

Ye and Chen[5] also proposed an anomaly detector based on the chi-square test for detecting intrusion in fixed networks. They concluded that the results demonstrate

*Author for correspondence

promising performance in terms of high detection and low false alarm rate.

Quinlan[6] performs inference of decision trees using a set of conditions over the attributes. Classification of new examples is carried out by applying the inferred rules.

A hybrid technique using unsupervised and supervised learning algorithm has also been studied in[7]. The similar kinds of data are grouped at an instance based on their behavior by using K-Means clustering.

Aikaterini Mitrokotsa[8] present the design and evaluation of intrusion detection models for MANETs using supervised classification algorithms. Specifically, we evaluate the performance of the MultiLayer Perceptron (MLP), the linear classifier, the Gaussian Mixture Model (GMM), the Naıve Bayes classifier and the Support Vector Machine (SVM).

The different form of classification of IDPS and classifying them in certain groups are contended[10]. Comparative analyses of different intrusion detection and prevention tools are calculated.

Görnitz, Nico[11] proposed the active learning strategy which automatically filter candidates for labeling and requires much less labeled data than state-of-art, while achieving higher detection.

2. Intrusion Detection using Supervised Model

The Supervised Model uses classification algorithms to perform the detection. The supervised Intrusion detection system uses labeled data for training. The labeling of data is difficult, time consuming and requires human intervention.

The network traffic load is used to perform detection in MANET. MANET frequently change their environment, hence researchers need to develop new IDs for changing and developing MANET. These algorithms are automated and accurate since they use statistical data. Once the training is over with the training set, they can be used for detection with any arbitrary cost matrices. The various issues should be taken into account, when a new IDS is being designed for MANETs.

The intrusion detection system has three major components. They are data preprocessing stage, building intrusion detection and evaluation of results. In stage1 data preprocessing, attack data is split into partition containing only one attack type. The samples are drawn randomly and preserving the balanced attack type distribution. In stage 2 we pass the data to proposed algorithm and in stage 3 we have evaluate the results.

The Figure 1. shows general supervised prediction learning method. We have used KDD 99 dataset to work with training set with known attacks generated. Each record in the data represents source and destination IP address. The records are independent of each other. The training data is labeled with different types of attacks or labeled as normal. The attacks categorized into four. They are Denial of Service (DoS), Remote to local(R2L), User to root(U2R) and Probe. The DoS is to prevent the legal users from getting the access. R2L is unauthorized access to remote machine. U2R is unauthorized access to local root privileges. Probe is attacker trying to get information of target host.

When checking the training set of KDD cup 1999 dataset, the DoS attack is 19.762066%, PRB is 0.83123%, R2L is 0.23869% and U2R are 0.010535%. The graphical representation of attack in the dataset is represented in Figure 2. The majority attack is DoS, followed by normal connection and rest of the categories represents less than 1% of training dataset. When data set is to be sampled these things taken into consideration. Only those values

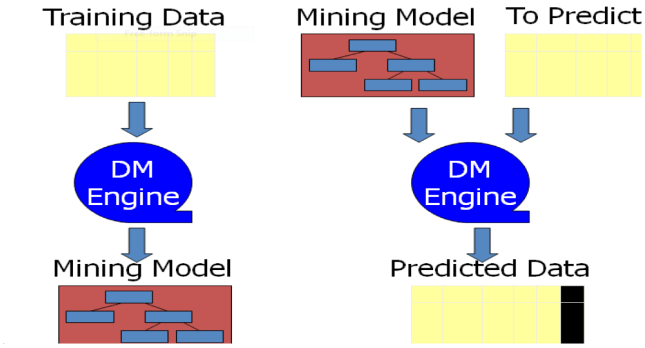

Figure 1. Prediction based on training data.

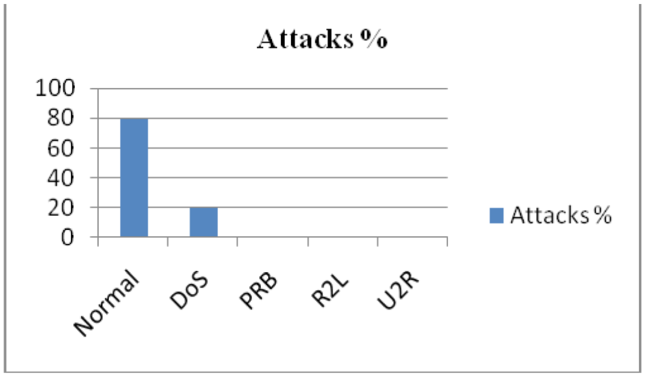

Figure 2. Training Set Attacks.

should be decreased. The proposed system is trained with both original and sampled dataset. The most confirmed data with predicted labels are selected and added. This helps us to remove the redundancy and control the size of labeled data.

2.1 Proposed Algorithm

The algorithm applied to KDD cup database is summarized as follows:

1. Load Testing data (KDD 99)
2. Search for the top 4 major values and made them as set S1
3. All other values are made as set S2.
4. The maximum value of connection C1 in S2 is selected.
5. The S1 values are compared to C1 and their order is also maintained
6. Repeat the steps 4 and 5 until optimum results obtained.
7. The system performance is tested with new set of values.
8. Check for evidence of attack and calculate false positive.
9. Train S2 from the label data.

3. Results and Analysis

KDD CUP99 data set[9] used in this analysis is from MIT Lincoln Lab used for intrusion detection in DARPA. This labeled data from controlled setting is not for real environment. In spite of its certain drawbacks it is the benchmark for most of research in IDS algorithms. This is time saving because Feature extraction need not be done and the Data labeling is already available. The focus is only on searching pattern and categorize from the training examples.

The network traffic load data is passed to proposed algorithm. The Table 1 shows the number of samples in the training dataset of KDD, when the proposed algorithm is applied on training set, the results obtained as shown in Table 2.

Table 1. Standard training data set

Normal	Attack				Total
	DOS	R2L	U2R	PROBE	
	391468	2903	53	6937	
108227	401361				509588

Table 2. Performance the proposed algorithm in the above training data set

	Normal	Dos	R2L	U2R	Probe
Normal	108227	20	8	0	32
Dos	21	391468	4	5	34
R2L	15	1	2903	6	5
U2R	2	2	3	53	7
Probe	39	43	9	1	6937

We can find the performance of above method is better than existing of supervised learning method learnt. False positive is important metrics to be considered. False positive is an event in the network, which reports a node as malicious accidently. A good IDS will have 0% false positive. The results calculated produced the accuracy about 98.57% and false ratio is minimized and limited to 1.2 and 1.3. It is experimented with different dataset, the number of normal and abnormal packets are monitored. The experiment result shows the detection methods provides high detection and reasonable low false rate.

4. Conclusion and Future Enhancement

In this paper, a supervised learning approach is investigated and experimented with International Knowledge Discovery and Data Mining Tools Competition intrusion detection benchmark (the KDDCUP 99 dataset). The supervised algorithms generally produce good accuracy for the known attacks. In this paper we have given heuristic method to existing classification in intrusion detection. The proposed algorithm is not only produced low false positive but also gives high accuracy for the four attacks discussed. We conclude the proposed algorithm gives better performance than standard algorithms for same data set.

The MANETs are highly vulnerable to attacks and new attacks are generated regularly. The KDDCUP is highly dominated by DOS and Probes. The many attacks like phf, imap etc. are under presented in the dataset. The new dataset under current scenarios mobility pattern and traffic conditions to be generated, such that the under rated attacks percentage has to be increased. The algorithm can be modified to current approach as future enhancement. The unsupervised method can be added to proposed algorithm on pattern based method for high network security.

5. References

1. Y. Li and J. Wei., "Guidelines on selecting intrusion detection methods in MANET", In Proceedings of the Information Systems Educators Conference, 2004.
2. Monita Wahengbam and Ningrinla Marchang, " Intrusion Detection in MANET using Fuzzy Logic", In IEEE 3rd National Conference, 2012; 189-192 .
3. Zhang Y., Lee W., Huang Y.: Intrusion Detection Techniques for Mobile Wireless Networks. In: Wireless Networks, 2003; 9(5), 545-556.
4. Huang Y., Lee W.: A Cooperative Intrusion Detection System for Ad Hoc Networks. In: Proceedings of the 1st ACM Workshop on Security of Ad Hoc and Sensor Networks (SASN03), 2003; 135-147. Fairfax, VA, USA .
5. N.Ye and Q.Chen," An Anomaly Detection Techniques based on a CHI-SQUARE Statistics for Detecting Intrusion into Information System" Quality and Reliability Engineering International, 2001.
6. Quinlan, J.: C4.5: Programs for Machine Learning. Morgan Kaufmann 1992
7. G. Pannell, and H. Ashman, "Anomaly Detection over User Profiles for Intrusion Detection," Information Security Management Conference, 2010
8. Mitrokotsa, Aikaterini, and Christos Dimitrakakis. "Intrusion detection in MANET using classification algorithms: The effects of cost and model selection." Ad Hoc Networks 11.1 2013; 226-237.
9. KDD data set, 1999, can be accessed at http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html.
10. Beigh, Bilal Maqbool, and M. A. Peer. "Intrusion Detection and Prevention System: Classification and Quick." 2011.
11. Görnitz, Nico, et al. "Toward supervised anomaly detection." arXiv preprint rXiv:1401.6424 2014.