International Journal of Engineering Research and Advanced Technology (IJERAT)

E-ISSN : 2454-6135

Volume.1, Issue 3
December -2015

Enhance Intrusion Detection Systems Based On SVM and Feed Forward Neural Network

Tanvi Khandelwal[1*] and Dr. Amit Shrivastav[2]

[1]M.Tech Scholar, PG Dept. of Computer Science & Engineering

[2] Head of Department, PG Dept. of Computer Science & Engineering

SAGAR Institute of Research and Technology, Bhopal, Madhya Pradesh

India

ABSTRACT

With large increase of internet users network security is important issues in today era. As variety of users have different requirement, so proper identification of safe network is required. This paper focuses on the network intrusion detection using SVM and neural network. Here SVM classify network behavior into two class first is safe and other is unsafe. Once unsafe network is identified then trained neural network identified attack type of the input sessions. Experiment is done on real dataset and obtained results are better than previous works on different parameters.

***Keywords**: Anomaly Detection, Computer Networks, Intrusion Detection, Network Security.*

1. INTRODUCTION

As network security plays important role in day to day life, so many measures are taken for the same such as encryption, firewall, VPN, etc. So one of the major role for the same is taken by intrusion detection system. As this is done by collection of different data information from various private and public network. Here software are update by the valuable information obtained from the intrusion detection system. It is known that attacker may harm or steal data in network, where attacker is internal or external from the network. But some unauthorized users or activities from different types of attackers which may internal attackers or external attackers in order to harm the running system, which are known as hackers or intruders, come into existence.

1 * Correspondence author

The main motive of such kind of hacker and intruders is to bring down bulky networks and web services. Due to increase in interest of network security of different types of attacks, many researchers has involved their interest in their field and wide variety of protocols as well as algorithm has been developed by them, In order to provide secure services to the end users. Among different type of attack intrusions is a type of attack that develop a commercial interest. Intrusion detection system is introduced for the protection from intrusion attacks.

From the above discussion we can conclude the main aim of the network Intrusion detection system is to detect all possible intrusion which perform malicious activity, computer attack, spread of viruses, computer misuse, etc. so a network intrusion detection system not only analyses

different data packets but also monitor them that travel over the internet for such kind of malicious activity. So the smooth running of overall network different server has to settle on the whole network which act as network intrusion detection system that monitor all the packets movements and identify their behaviour with the malicious activities. One more kind of network Intrusion detection system is developed that can be installed in a centralized server which also work in the similar fashion of analyzing and monitoring the different packet data units for their network intrusion behaviour. Network Intrusion detection system can be developed by two different approaches which can be named as signature based and anomaly based. In case of signature based Network Intrusion detection system it develops a collection of security threat signature. So according to the profile of each threat the data stream of different packets in the network are identified and the most matching profile is assigned to that particular packets. If the profile is malicious then that data packet comes under intrusion and it has to remove from the network in order to stop his unfair activities.

2, LITERATURE SURVEY

The KDD'99 has been the most wildly used data set for the evaluation of anomaly detection methods is prepared by Stolfo et al, based on the data captured in DARPA'98 IDS done in [1].

Agarwal and Joshi [2] proposed a two stage general to specific framework for learning a rule based model (PNrule) to learn classifier models on a data set that has widely different class distributions in the training data. The proposed PN rule evaluated on KDD dataset reports high detection rate.

Yeung and Chow [3] proposed a novelty detection approach using no parametric density estimation based on parzen window estimators with Gaussian kernels to build an intrusion detection system using normal data. This novelty detection approach was employed to detect attack categories in the KDD dataset.

In 2006, XinXu et al. [4] presented a framework for adaptive intrusion detection based on machine learning.

Lee et al. [5], introduced data mining approaches for detecting intrusions. Data mining approaches for intrusion detection include association rules that based on discovering relevant patterns of program and user behaviour.

Association rules [6], are used to learn the record patterns that describe user behavior. These methods can deal with symbolic data and the features can be defined in the form of packet and connection record details. However, mining of features is limited to entry level of the packet and requires the number of records to be large and low diversity in data; otherwise they tend to produce a large number of rules which increases the complexity of the system [7]. Data clustering methods such as the k-means and the fuzzy c-means have also been applied extensively for intrusion detection. One of the main drawbacks of clustering technique is that it is based on calculating numeric distance between the observations and hence the observations must be numeric.

In [11] a detection method that detects and diagnoses anomalies in large scale networks. First, their approach monitors the traffic using a matrix in which each cell represents the traffic volume of a link of the network at a certain time interval. Second, the main behaviour of the traffic is extracted from the matrix with the principal component analysis (PCA) and anomalies are detected in residual traffic. Finally, the origin and destination nodes of the network that are affected by the anomalous traffic are identified and reported.

3. METHODOLOGY

Whole work is divide into two module, first is separation of safe and unsafe session from the dataset using SVM and unsafe dataset passed in to feed forward neural network. Then in second module identification of type of intrusion is done in unsafe network.

3.1. Module 1.SVM

As dataset consist of sessions which has normal behavior of network as well as abnormal behavior of network. Separation of session in these two categories is done by SVM. Here pre-processing is done for the sessions where unnecessary information is removing and data is arranged for training.

Pre-Processing: Here information like type of protocol, socket type, etc. are removed. As presence of these information increases confusion for the SVM training. This can be understand as let raw data session is

{0,tcp,ftp_data,SF,491,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,2,2,0.00,0.00,0.00,0.00,1.00,0.00,0.00,150,25,0.17,0.03,0.17,0.00,0.00,0.00,0.05,0.00,normal,20}

After applying pre-processing session will be:

{0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,2,2,0.00,0.00,0.00,0.00,1.00,0.00,0.00,150,25,0.17,0.03,0.17,0.00,0.00,0.00,0.05,0.00,normal}

3.1.1. Feature Selection:

In this step of first module features present in the session are identified and store in a matrix. Separate matrix is prepare for safe and unsafe mode. This can understand as the feature matrix of safe and unsafe have two features from each session:

F11:{0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,2,2,0.00,0.00,0.00,0.00,1.00,0.00,0.00,150,25,0.17,0.03,0.17,0.00,0.00,0.00,0.05,0.00}

F12: {normal}

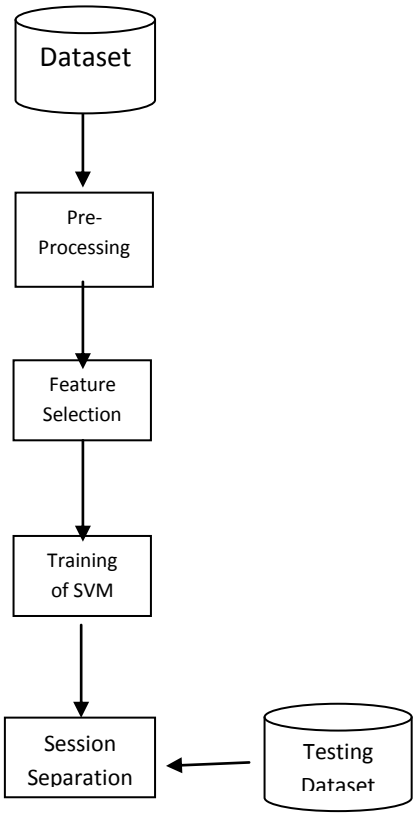

Figure 1.Proposed work first module.

So each feature matrix has two columns, where first present numeric values and other present its type.

3.1.2. Training of SVM:

In this phase both feature vector of safe and unsafe mode is pass in the SVM, here the use of linear separation training of the SVM is done. The goal of linear classification problem is to obtain two parallel hyper planes as shown in Fig. 1, wx + b=1 and wx' + b=-1, where w and b are the

classification parameters obtained during the training process. Both
hyper planes separate the training data of the two classes such that the distance between those hyper planes is maximized.

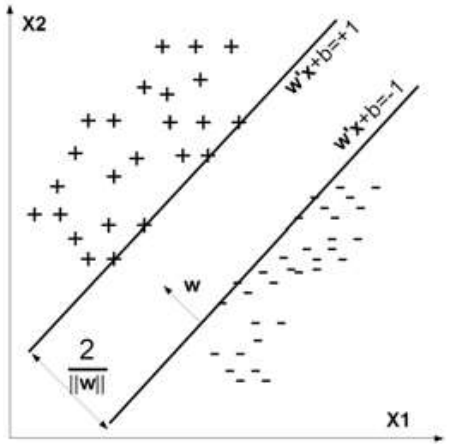

Figure2. Training data samples for two different classes are denoted by + and - signs.

3.4 Session Separation:

In this step of first module trained SVM is use for the separation of session in two categories first is safe mode and other is unsafe mode. So session from the testing dataset is pass one by one in the trained SVM.
Output of the SVM is two class + or -. So, first evaluation is required for the correct identification of session of their respected category.
Input: KDD
Output: SVM (Trained Support Vector Machine)

1. DS←Pre-Process(KDD) // DS (Filter Dataset
2. F[n, c]←Feature(DS) // F (feature), n (Total Session), c (safe and unsafe state of session)
3. Loop 1:n
4. SVM←Train_SVM(F[n, c])
5. EndLoop

3.2. Module 2: Feed Forward Neural Network

In this module dataset of attacked sessions filter by SVM are utilize to train the FFNN. Then trained dataset is utilized for the testing of unknown attack session.
Pre-Processing: Here removal of dataset unnecessary information is done in this step. This is same as filtering of pre-processing done in first module.

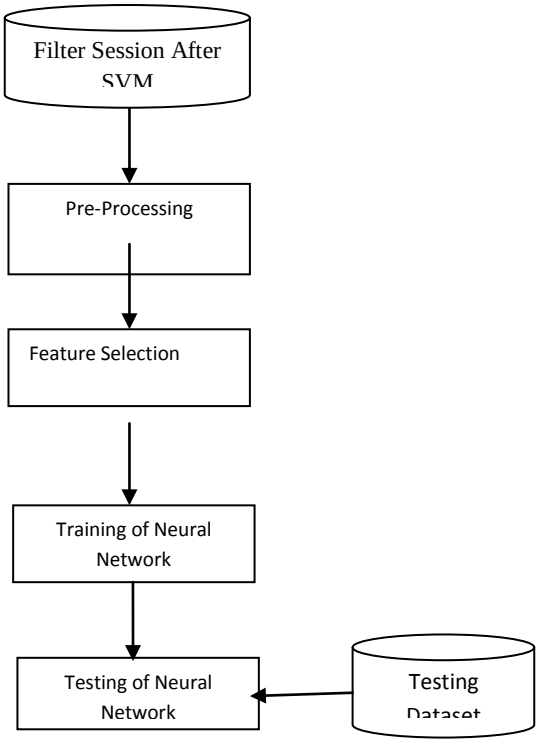

Figure 3 Proposed work second module.

Feature Collection: Here as dataset contain only attack sessions, but each contain variety of attacks so separation of each type of attack session in there group is done in this step.

3.2.1 Training of Neural Network:

In this step features obtained after pre-processing is put in feed forward neural network.

3.3. Proposed Algorithm

Input: KDD

Output: FFNN (Trained Feed Forward Neural Network)

6. DS←Pre-Process(KDD) // DS (Filter Dataset
7. F[n]←Feature(DS) // F (feature), n (Total Session)
8. Loop 1:n
9. FFNN←Train_FFNN(F[n])
10. EndLoop

4. EXPERIMENTAL DETAILS

Dataset: Experiment is done on Real KDD99 Dataset. In this dataset five kind of broad attack category is present while total session is 126000. These contain 41 feature vector in each session. Whole work is done on MATLAB 2012a.

4.1 Evaluation Parameter

Accuracy: It is the ratio of total number of true alarm divide by total number of true and false alarm.

Precision = True Positives / (True Positives+ False Positives)

Recall = True Positives / (True Positives +False Negatives)

F-Score = 2 * Precision * Recall / (Precision + Recall)

5. RESULTS AND DISCUSSION

Data-Set Size	SVM
	Accuracy
9000	0.99567
8000	0.9866
4000	0.993

Table.1 SVM separation values of different size data.

From above table 1 it is obtained that with the increase in dataset size accuracy rate decreases. But this is done as types of data increases so confusion for SVM also increases. As number of patterns are more in the dataset so results are more accurate after a consistent data type.

Data-Set Size	SVM		
	Precision	Recall	F-Measure
9000	0.99316	0.99849	0.99582
8000	0.9759	0.998	0.986
4000	0.975	1	0.992

Table2.SVM Precision, Recall, F-measure values at different Dataset Sizes.

Dataset	Normal	Attack
9000	4675	4325
8000	4120	3880
4000	1973	2027

Table3. SVM Normal and attack session values at different Dataset Sizes.

From above table 2 and 3 it is obtained that with the increase in dataset size precision value rate increase. As numbers of patterns are more in the dataset so results are more accurate.

9000	Actual	Proposed Work
DOS	3159	3159
Probe	890	890
R2L	222	218
U2R	53	53
8000	Actual	Proposed Work
DOS	2826	2826
Probe	786	786
R2L	214	210
U2R	53	53
4000	Actual	Proposed Work
DOS	1399	1399
Probe	407	407
R2L	180	175
U2R	40	40

Table 4. Represent detection of different attack from the proposed work.

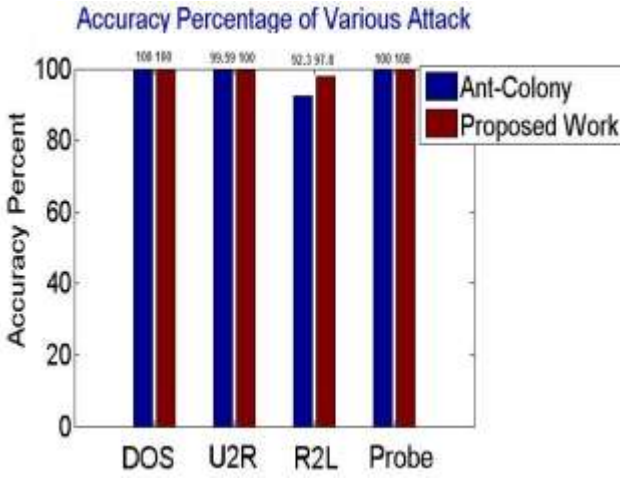

Figure5. Accuracy comparison of FFNN for different attacks.

From above fig. 3 and table 4 it is seen that proposed work has achieved same accuracy as done by [8] for DOS and Probe attack, while proposed work well in case of R2L and U2R attack. So overall accuracy for identification of type of attack is high and appreciable.

6. CONCLUSION

As different approaches of network security is done by various researchers. In this paper a combination of SVM and Feed Forward neural network is done for the detection of intrusion in network. Results obtain are highly appreciable as trained networks identify all type of intrusions more than 99% of accuracy. As in future it needs to be improved by putting data on the unsupervised network, so it automatically updates the new behaviour of the intruder.

REFERENCES

1.]Leonid Portnoy ,EleazarEskin and Stolfo, "Intrusion Detection with Unlabeled Data Using Clustering" Department of Computer Science, Columbia University, Newyork, NY 10027

2. R. Agarwal, and M. V. Joshi, "PNrule: A New Framework for Learning Classifier Models in Data Mining", Technical Report TR 00-015, Department of Computer Science, University of Minnesota, 2000.

3. Dit-Yan Yeung, Calvin Chow, "Parzen-Window Network Intrusion Detectors," icpr, vol. 4, pp.40385, 16th International Conference on Pattern Recognition (ICPR'02) - Volume 4, 2002

4. XinXu, Adaptive Intrusion Detection Based on Machine Learning: Feature Extraction, Classifier Construction and Sequential Pattern Prediction, Institute of Automation, College of Mechantronics Engineering and Automation, National University of Defence Technology, Changsha, 410073, P.R.China, International Journal of Web Services Practices, Vol.2, No.1-2 (2006), pp. 49-58

5. Lee W., Stolfo S., and Mok K., "A Data Mining Framework for Building Intrusion Detection Model," in Proceedings of IEEE Symposium on Security and Privacy Oakland, pp. 120132, 1999.

6. Agrawal R., Imielinski T., and Swami A., "Mining Association Rules between Sets of Items in Large Databases," in Proceedings of the International Conference on Management of Data , USA, vol. 22, pp. 207216, 1993.

7. Abraham T., "IDDM: Intrusion Detection using Data Mining Techniques," available at: http://www.dsto.defence.gov.au/publications/2345/DSTOGD0286.pdf, last visited 2008.

8. Chetan Gupta 1, Amit Sinhal2 and Rachana Kamble3 " An Enhanced Associative Ant Colony Optimization technique based Intrusion Detection System". 2012 IEEE Conference.

9. Sen J., "An AgentBased Intrusion Detection System for Local Area Networks," International Journal of Communication Networks and Information Security , vol. 2, no. 2, pp. 128140, 2010.

10. KDDCUP 1999 Data, available at: http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html, last visited 2013.

11. Vidarshana W. Bandara and Anura P. Jayasumana Extracting Baseline Patterns in Internet Traffic Using Robust Principal Components. 36th annual conference on Local Computer Network.