Novel Approach of Intrusion Classification by Support Vector Machine with Flower Pollination Algorithm

Parshant Gupta[1], Raman Goyal[2]
M-Tech (CSE Department), LLRIET MOGA (PUNJAB) [1]
Assistant Professor (CSE Department), LLRIET MOGA (PUNJAB) [2]

Abstract **-** Intrusion Detection Systems (IDS) monitor a secured network for the evidence of malicious activities originating either inside or outside. Upon identifying a suspicious traffic, IDS generates and logs an alert. Unfortunately, most of the alerts generated are either false positive, i.e. benign traffic that has been classified as intrusions, or irrelevant, i.e. attacks that are not successful. The abundance of false positive alerts makes it difficult for the security analyst to find successful attacks and take remedial action. IN this paper proposed approach working on optimization feature by give relative weight this is done by flower pollination approach, which given weight to feature and then learn by SGD and adaptive boost.

Keywords **-** *SGD, ADAboost, FPA, optimization, IDS*

I. INTRODUCTION

Intrusion is defined as an unwanted action that pursuit unauthorized access, variation in the data or assimilate the system unstable by susceptibility the existing activities in the system [1]. Intrusion is a process to break into or misuse the system of other and violate the security policies. In security system there are two threats intruder and other one is malware. Intruder is the major threat in the internet and it's very important to overcome by the effect of this attack [1]. There are several techniques used to protect the system from the intrusion. In this paper we only use to discuss about the IDS. IDS stand for intrusion detection system which is used to automate the detection process. Network based IDS monitors the data packets and drops the unauthorized data packets [2]. There are one another method that is used to prevent the system from the intrusion termed as IPS which stands for intrusion prevention system or the combination of both the system i.e. intrusion detection system and intrusion prevention system prevention. IDS used to control the network surroundings and alert a human about the presence of an attack [IDS]. It is also utilizes to identify the attacks that are difficult and resource intensive to prevent but can be mitigate once they started [3].

Types of IDS

There are three types of intrusion detection system:

1) *Network Intrusion Detection System:* Snort is the best example of network intrusion detection system. This system used to observing network traffic and to control multiple hosts. There is a development in the execution of this system by using a technique where traffic networks are connected to the hub, network tap. [4][5]
2) *Host-based Intrusion Detection System:* This system detects the intrusion with the help of analyses of system calls, application logs, modification in file-system (binaries, a file having passwords etc.). This all detection is possible due to an agent in a host which is the key component of the host-based intrusion detection system. As the name indicates this system is really based on the host for the purpose of detection [4] [5]
3) *Hybrid Intrusion Detection System:* as the name indicates it a hybrid system which means **it** is a combination of two or more systems. Here it is a combination of above two systems. We can say that it is a combination approach. To frame a comprehensive view of the network there is a combination of agent attached to the host with network information. The common example of hybrid IDS is Prelude. [4][5]

There are various techniques uses and represented by researchers to identify intrusion in the system like Liyuan Xian [6] uses Bayesian network classifier to detect the intrusion. Tianyi Xing [7] uses snort flow model to identify the intrusion and its solution system by reconfiguring the cloud networks. There are number of researchers and application used for intrusion detection.

II. LITERATURE REVIEW

Ravale et al. [8**]** in this paper, the author proposed a hybrid approach of data mining in which K-mean clustering algorithm is combined with the RBF kernel function which belongs to the SVM classification method. The main purpose to develop this technique is to reduce the number of the related attribute with each data. The result of this technique shows that this technique performs better and provide good accuracy rate and decision tree.

Parvat et al. [9] presented a new method for intrusion detection system by using the Deep Packet Inspection method and algorithm to detect signatures. It measures the time, space and accuracy of the system. DPI method is used to identify the payload traffic, network traffic and quality of service. The

main function of DPI is a detection of antivirus, protocol, and IDS. The detection engine is supported by signatures. This model improves the performance of DPI for intrusion detection.

Karkouch et al. [10**]** in this paper the main aim is to enhance the Data quality in IoT by providing the state of art. The author discussed the properties of data in IoT. Data quality concept is defined and a set of generic and domain-specific dimensions were defined. IoT related factors and their effects on data quality were also analyzed. Data cleaning techniques are used to enhance the Data quality.

Guo, Chun, et al. [11] the author studied on the intrusion detection system with a hybrid approach. Hybrid classifiers were used for proving the accuracy in intrusion detection. Distance Sum Based Support Vector Machine (DSSVM) model for effective intrusion detection. In this technique they used sum of distance, correlation between the sample and clusters centers. The author considers a data set represented by n-dimensional features vectors. Center of the cluster is measured by using the clustering algorithm and classification is performed with SVM classifier.

Ghanem et al. [12] in this paper author discussed on the network anomalies and their detection methods by using the hybrid approach. These methods are based on the meta-heuristic approach and genetic algorithm. This approach is based on the negative selection based detector generation. All the experiments are performed on the KDD data set. The result of this paper shows effectiveness in generating a number of detectors.

Lee et al. [13] author proposed a novel approach in data mining for identifying the uncertain data patterns. This approach provides the accuracy of the data from the mining process with high correctness. Results show that the performance on various types of data set and runtime on these data sets is effective. It increased the memory usage and scalability of the large data set.

Aljawarneh et al. [14] in this paper, the author presented a Simplified Regular Expression (SRE) method which is based on the network-based signature detection and this is read by IDS and firewalls. In this approach, Enhanced Contiguous Substring Rewarded (ECSR) algorithm is developed to improve the results. The signature generated by the SRE is found more effective. The result after the evaluation shows that signatures provided accurate results.

Airehrour et al. [15] in this paper, the author survey on the IoT and the methodologies related to it. In this survey, the author analyzed the routing protocol their working for providing reliable communication in the network. The author also discussed the weakness, challenges, and strategies of the techniques.

III. PROPOSED WORK

In this paper following methodology is proposed into three phases Feature Selection, Feature Extraction and Normal.

A. Feature Selection: In the feature selection phase the following steps takes place as shown in Figure 1.1

Step1: KDD-99 dataset with 41 features

Step2: Feature Selection by information gain & correlation method.

Step3: Input the feature & labels into SVC, SGD & adaptive boost &make three models.

Step4: Perform the test on these models & calculate the precision, recall, and accuracy.

B. Feature Extraction: In this feature extraction phase, the following steps take place as shown in the Figure 1.1

Step1: KDD-99 dataset with 41 features

Step2: Feature extraction with FPA

Step3: Input the feature & labels into SVC, SGD & adaptive boost & make three models.

Step4: Perform the test on these models and calculate the precision, recall & accuracy.

C. Normal: In this phase the following steps takes place as shown in Figure 1.1

Step1: KDD-99 dataset with 41 features

Step2: Input the feature & labels into SVC, SGD & adaptive boost &make three models.

Step3: Perform the test on these models and calculate the precision, recall accuracy.

SGD: Stochastic Gradient Descent (SGD) is a straightforward yet extremely productive way to deal with the discriminative learning of direct classifiers under convex loss functions such as (linear), for example, (direct) Support Vector Machines and Logistic Regression. Despite the fact that SGD has been around in the machine learning group for quite a while, it has

gotten a lot of consideration only as of late with regards to huge scale learning.

SGD has been effectively connected to the huge scale and inadequate machine learning issues regularly experienced in content characterization and characteristic dialect handling. Given that the information is scanty, the classifiers in this module effectively scale to issues with more than 10^5 preparing illustrations and more than 10^5 highlights.

SGD Classifier supports multi-class classification by combining multiple binary classifiers in a "one versus all" (OVA) scheme. For each of the K classes, a binary classifier is learned that discriminates between that and all other $K-1$ classes. At testing time, we compute the confidence score (i.e. the signed distances to the hyper-plane) for each classifier and choose the class with the highest confidence. The Figure below illustrates the OVA approach on the iris dataset. The dashed lines represent the three OVA classifiers; the background colors show the decision surface induced by the three classifiers.

SVC: SVC is another implementation of Support Vector Classification for the case of a linear kernel. Note that Linear SVC does not accept keyword kernel, as this is assumed to be linear. It also lacks some of the members of SVC and NuSVC, like Support.

FLOW CHART OF PRESENT WORK

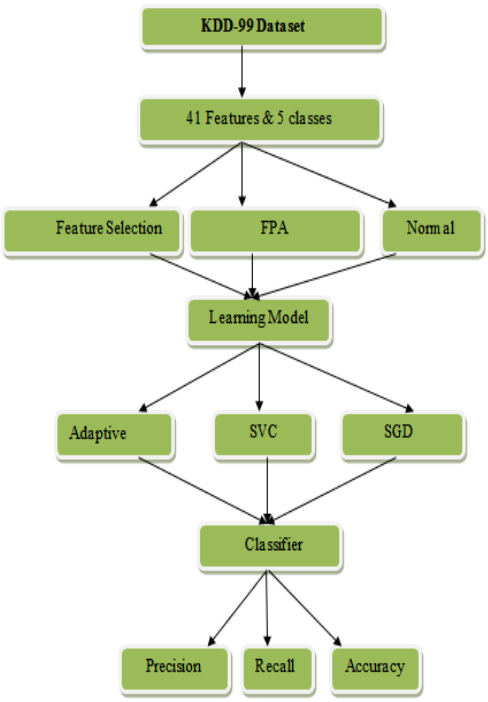

Figure 1.1 Flow chart of proposed methodology

IV. RESULTS

In in this section author explain the results of the classifier and graphs show the variations in it.

Table1.1 Result table of training and testing data

Algorithms	Training Accuracy	Testing
FPA Adaptive Boost	96.00%	99.68%
FPA+ SGD	94.44%	99.85%
SVM	75.62%	60.42%

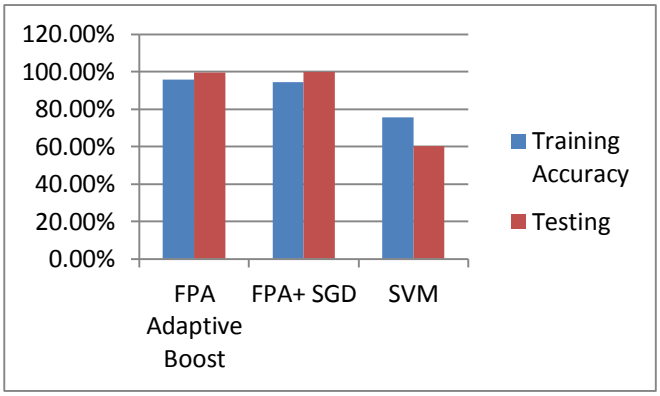

Figure 1.2 Result of training and testing data

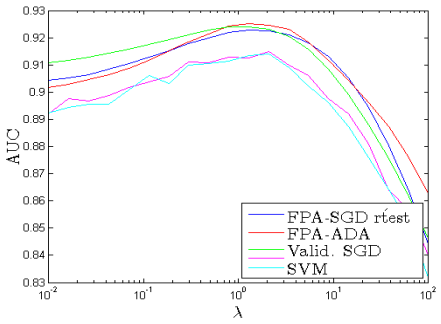

Figure 1.3 ROC curve comparison of different approaches

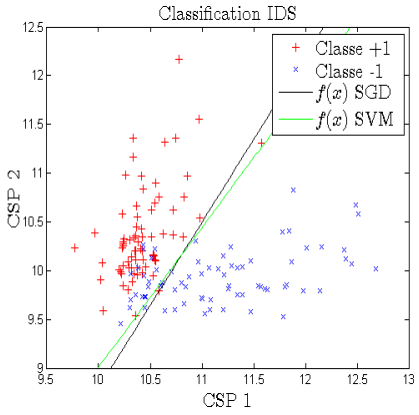

Figure 1.2 Classification visualization by SVM and SGD

V. CONCLUSION

In this research, we evaluated various machine learning classifiers to enhance the malware detection outcome for a large collection of file samples and obtain the optimum classifier able to detect mobile malware. The classifiers selected were SGD, SVC, and Adaptive boost with FPA. Support vector machine is good in time series and mathematical formulations. Performance metrics train with labeling in data set and Adaptive boost with FPA classify them in two linear and non-linear class malicious or normal nodes. After this experiment, we can conclude that adaptive boost with FPA better than SVM is much powerful in packet dropping attacks because it is able to detect 93-96% accurate compare to other machine learning techniques.

VI. REFERENCES

[1]. Al-Jarrah, Omar, and Ahmad Arafat. "Network Intrusion Detection System using attack behavior classification." *Information and Communication Systems (ICICS), 2014 5th International Conference on*. IEEE, 2014.

[2]. Kenkre, Poonam Sinai, Anusha Pai, and Louella Colaco. "Real time intrusion detection and prevention system." *Proceedings of the 3rd International Conference on Frontiers of Intelligent Computing: Theory and Applications (FICTA) 2014*. Springer, Cham, 2015.

[3]. Pathan, Al-Sakib Khan, ed. *The state of the art in intrusion prevention and detection*. CRC press, 2014.

[4]. Kumar, B. Santos, T. Ch, Ra Sekhara Phani Raju, M. Ratnakar, Sk Dawood Baba, and N. Sudhakar. "Intrusion Detection System-Types and Prevention." (2013).

[5]. Vishakhapatnam, A. P. "Intrusion Detection System-Types and Prevention."

[6]. Xiao, Liyuan, Yetian Chen, and Carl K. Chang. "Bayesian model averaging of Bayesian network classifiers for intrusion detection." *Computer Software and Applications Conference Workshops (COMPSACW), 2014 IEEE 38th International*. IEEE, 2014.

[7]. Xing, Tianyi, et al. "Snortflow: A openflow-based intrusion prevention system in cloud environment." *Research and Educational Experiment Workshop (GREE), 2013 Second GENI*. IEEE, 2013.

[8]. Ravale, Ujwala, Nilesh Marathe, and Puja Padiya. "Feature selection based hybrid anomaly intrusion detection system using K means and RBF kernel function." *Procedia Computer Science* 45 (2015): 428-435.

[9]. Parvat, Thaksen J., and Pravin Chandra. "A Novel approach to deep packet inspection for intrusion detection." *Procedia Computer Science* 45 (2015): 506-513.

[10]. Karkouch, Aimad, et al. "Data quality in internet of things: A state-of-the-art survey." *Journal of Network and Computer Applications* 73 (2016): 57-81.

[11]. Guo, Chun, et al. "A distance sum-based hybrid method for intrusion detection." *Applied intelligence* 40.1 (2014): 178-188.

[12]. Ghanem, Tamer F., Wail S. Elkilani, and Hatem M. Abdul-Kader. "A hybrid approach for efficient anomaly detection using metaheuristic methods." *Journal of advanced research* 6.4 (2015): 609-619.

[13]. Lee, Gangin, and Unil Yun. "A new efficient approach for mining uncertain frequent patterns using minimum data structure without false positives." *Future Generation Computer Systems* 68 (2017): 89-110.

[14]. Aljawarneh, Shadi A., Raja A. Moftah, and Abdelsalam M. Maatuk. "Investigations of automatic methods for detecting the polymorphic worms signatures." *Future Generation Computer Systems* 60 (2016): 67-77.

[15]. Airehrour, David, Jairo Gutierrez, and Sayan Kumar Ray. "Secure routing for internet of things: A survey." *Journal of Network and Computer Applications*66 (2016): 198-213.