ISSN 2395-1621

Study of Machine Learning Based Intrusion Detection System

[#1]Prashant Wakhare, [#2]Dr S.T.Singh

[1]Prashant_mitr@rediffmail.com
[2]stsingh47@gmail.com

Computer Engineering,
Savitribai Phule Pune University
P K Technical Campus, Pune

ABSTRACT

The network security has the key of financial and business web application. Intrusion detection is one of the looms to resolve the problem of network security. One of the main challenge is detection of intrusion. Diffrent Intrusion Detection method have been proposed Out of this method machine learning based approch are observed to be efficient in terms of detection accuracy and alert genration of system to be act immediately. A study on differnt intrusion in terms of supervised, un-supervised based anomaly detection as well as misuse detection are reviewd in this work. The study also classifies the machine learning algorithms are used. The performance of algorithms are compared and efficient method are identified.

Keywords— Intrusion Detection System, Machine learning algorithms.

ARTICLE INFO

Article History

Received : 15th June 2015

Received in revised form :

17th June , 2015

Accepted : 21th June, 2015

Published online :

24th June 2015

I. INTRODUCTION

In the networks are composed of a number of limited, battery-powered and multi-utilitarian devices called networks which are obtusely arranged to collect data from untended environments. It offers numerous advantages over conventional networking in terms of lower cost, scalability, flexibility, reliability and ease of deployment. The constrained network resources in terms of memory, processing, transmission and power make IDS. vulnerable to a variety of malignant attacks. This provides the way for the attackers to launch attacks in these protocols. Modeling a stable security protocol is a difficult task and more expensive that leads to enforcement degradation. Intrusions are the set of actions by the intruders that attempt to compromise the principles of security. An Intruder attempts to gain illegitimate entry to a system/network . It is built on many patterns such as refusing the services by inundating system resources, speedily inseminating a virus or worm, and gaining authorization of root users to perform malignant behavior. Network intruders enter the hosts through Enumeration, Viruses, Trojan horse, E-mail infection, password cracking, and router attacks [8]. Intrusions are caused by insiders an authorized users also while they attempt to gain and try to misuse unauthorized privileges. The various causes of intrusions include erroneous innovation, planning, hardware/software model and susceptibility, applications, component, operational defects and external disturbances.

II. LITERATURE SURVEY

In this section, related literature about support vector machine approach and preparation of datasets for data mining activity will be reviewed and discussed.

Annie George [4], Anomaly detection has emerged as an important technique in many application areas mainly for network security. Anomaly detection based on support vector machine learning algorithms considered as the classification problem on the network data has given that. Dimensionality reduction
and classification algorithms are explored and evaluated using KDD 1999 dataset for Intrusion Detection System.
The result shows the decrease in execution time for the classification as they reduce the dimension of the input data and also the precision and recall parameter values of the

classification algorithm shows that the Support Vector Machine with Principle Component Analysis method is more accurate as the number of misclassification decreases.

Kui W. Mok [5], there is produced the need to update an installed intrusion detection system (IDS) due to new attack methods or upgraded computing environments. This paper describes a data mining technique for adaptively building Intrusion Detection System (IDS) models. The central idea is to utilize auditing programs to extract an extensive set of features that describe each network connection or host session, and apply data mining programs to learn rules that accurately capture the behavior of intrusions and normal activities. These rules can then be used for signature based detection and novel detection. We discuss the idea of our data mining programs, namely, classification, meta-learning, association rules, and frequent episodes. We analyze the results of applying this technique to the extensively gathered network audit data for the 1998 DARPA Intrusion Detection Evaluation Program.

V. Jyothsna, V. V. Rama Prasad, K. Munivara Prasad [6], With the advent of anomaly-based intrusion detection systems, many approaches and techniques have been developed to track novel attacks on the systems. High detection rate of 98 percent at a low alarm rate of 1 percent can be achieved by using these techniques. Though anomaly-based approaches are efficient, signature-based detection is preferred for mainstream implementation of intrusion detection systems. As a variety of anomaly detection techniques were suggested, it is difficult to compare the strengths, weaknesses of these methods. The reason why industries do not favor the anomaly-based intrusion detection methods can be well understood by validating the efficiencies of the all the methods. To investigate this issue, the current state of the experiment practice in the field of anomaly-based intrusion detection is reviewed and survey recent studies in this. This paper contains summarization study and identification of the drawbacks of formerly surveyed works.

CHEN Bo, Ma Wu [7], the effective way of improving the efficiency of intrusion detection is to reduce the heavy data process workload. In this paper, the dimensionality reduction use of technology in the classic dimensionality reduction algorithm principal component to analysis large scale data source for reduced-made features of the original data be retained and improved the efficiency of intrusion detection. And use BP neural network training the data after dimensionality reduction, will be effective in normal and abnormal data distinction, and achieved good results.

Paul Dokas, Vipin kumar [8], in which they gives an overview of our research in building rare class prediction models for identifying known intrusions and their variations and anomaly detection schemes for detecting novel attacks whose nature is unknown. Disadvantage of this paper is that due to the fact that the number of instances of U2R and R2L attacks in the training data set is very low, these numbers are not adequate as a standard performance measure. It could be biased if we use these numbers as a measure for performance of the system.

III. INTRUSION DETECTION SYSTEM

An intrusion detection system (IDS) is a frame of reference used to detect illegitimate intrusions in a computer system. There are three types of Intrusion detection system. They are

1. Misuse detection,
2. Novel detection and
3. Specification detection.

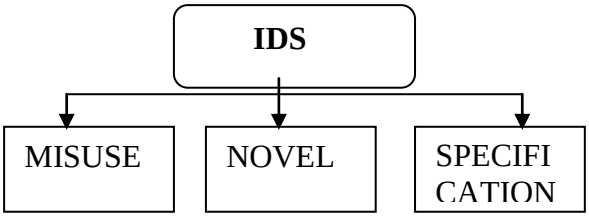

Fig.2 shows the classification of intrusion detection systems**.**

3.1. Misuse Detection

In misuse detection, the attitude of Misuse detection is correlated with infamous attack patterns. This technique requires knowledge to build attack patterns and difficult to detect unknown attacks.

3.1.1Limitation

This technique requires familiarity to build attack models and fail to detect novel attacks.

3.2. Anomaly Detection

Anomaly detection analyses and classifies the behavior of normal or abnormal according to certain metrics such as changes made in payload of packet and retransmission of packet in specified threshold. This technique is used to detect unknown attacks.

3.2.1 Limitation

This leads to an increase in substantial amount of false Alarm rate.

3.3. Specification Detection

Specification detection incorporates misuse and anomaly detection and it targets on detecting the variations from known.

3.3.1 Limitation

It requires manual development of all specifications. Out of three detection methods, the anomaly based detection techniques are studied in this paper. The next section presents the various anomaly based detections in WSN.

IV. ANOMALY-BASED INTRUSION DETECTION SYSTEM

An anomaly detection system is one of the intrusion detection systems available to model the normal system behavior which is effective in identifying both known as well as unknown attacks. It is built on a normal system that studies the network or program activity. There are a number of different architectures and methods used for anomaly detection. They are statistical approach, clustering approach, centralized approach, artificial immune system, isolation table, machine learning approach and game-theory approach. Table 1 shows the overall comparison of IDS in terms of accuracy, energy efficiency, memory requirements and network structure. The classification of anomaly-based Intrusion detection System is shown in Fig.3.

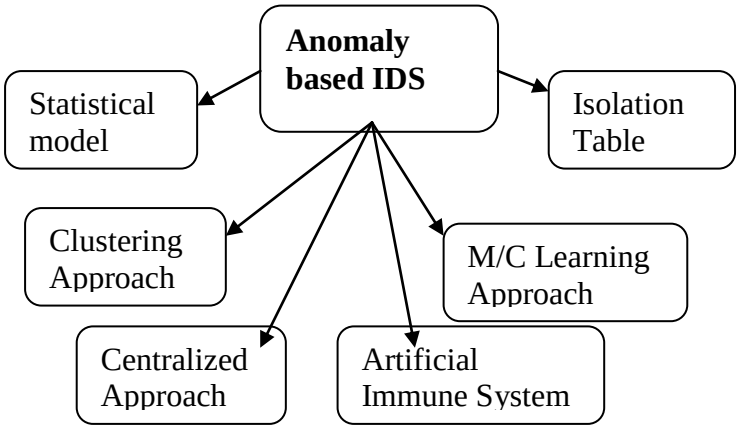

Fig 3. Shows the classification Anomaly based IDS

4.1 Statistical Model-based approach

These approaches are based on statistical models built on each sensor node to classify the packets as normal or abnormal. At every node, end packets from each neighboring nodes are used to calculate the statistical metrics. Each arriving packet is compared with statistical model to classify it as normal or anomaly.

4.2 Clustering Model-based approach

This model uses unsupervised learning algorithms for routing attacks to build a model of normal behavior. The traffic samples are considered as a group of clusters. The clusters that contain less training traffic samples than a certain threshold are considered as anomaly. This model is used to detect anomaly in the traffic patterns.

4.3 Centralized approach

A centralized anomaly detection system known as ANDES is available in the literature. In this approach, A detection agent is positioned in the base station and information are cascaded and scrutinized to identify network anomaly. It collects application data, management information and the node status information to detect intrusions.

4.4 Artificial Immune System-based approach

Artificial Immune System is very efficient and effective to detect node misbehaviors in WSN. Dendritic Cell algorithm is modeled to detect the cache poisoning attacks. Sensor nodes build two tables namely, interest cache table and data cache table. When a node receives a packet, necessary updates are performed in two cache tables. The signals and antigens are excerpted from each packet and moved to dendritic cell where the antigens are classified as benign or malicious.

4.5 Machine learning-based approach

Machine learning and automaton-based learning approaches are used for anomaly detection in WSN. These approaches are based on packet sampling, where a capacity of the packets roam in the network is fragmented to identify the malicious nodes. It makes use of hidden markov model to detect the affair to be out of range and raises a horror.

4.6 Isolation Table

Isolation table registers the anomaly information and detection table isolates the nodes in the network. The tables are generated by cluster head and forwarded to the base station. Out of the different anomaly based IDS, machine learning based methods are discussed in the next section.

V.MACHINES LEARNING-BASED ANOMALY DETECTION

Different techniques have been used for anomaly detection. Existing intrusion detections are not sufficient in detecting the novel attacks. Therefore, some anomaly detection approaches work on the available normal data and model them to identify the deviations. To solve the detection systems that rely on human intervention, machine learning based anomaly detections are discussed. Machine learning deals with ability of a program to train and enhance the performance on a certain task [3]. The classification of Machine-learning based anomaly detection is as follows:

5.1 Supervised-learning based anomaly detection
5.2 Unsupervised-learning based anomaly detection

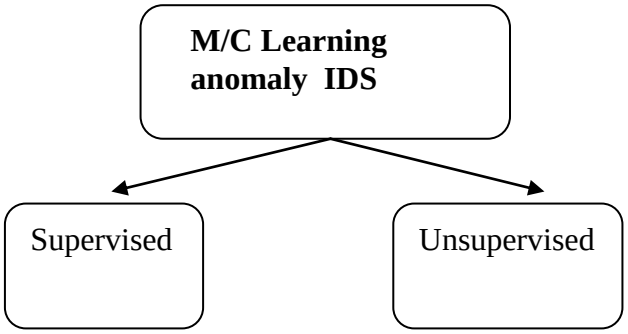

Fig 3 M/C learning anomaly IDS

5.1 Supervised Learning-based Anomaly Detection

In the supervised anomaly detection, the learning comes from the labeled examples in the training data set and mostly used for classification. Some of the supervised learning algorithms that are used for anomaly detection are designed based on support vector machine (SVM). They are also combined with principal component analysis (PCA), particle swarm optimization (PSO), AdaBoost-based classifier, and K-Means algorithm with C4.5 Decision trees to classify the network behaviors as either normal or abnormal behavior. Some examples of supervised machine learning algorithms are shown in Fig.5.

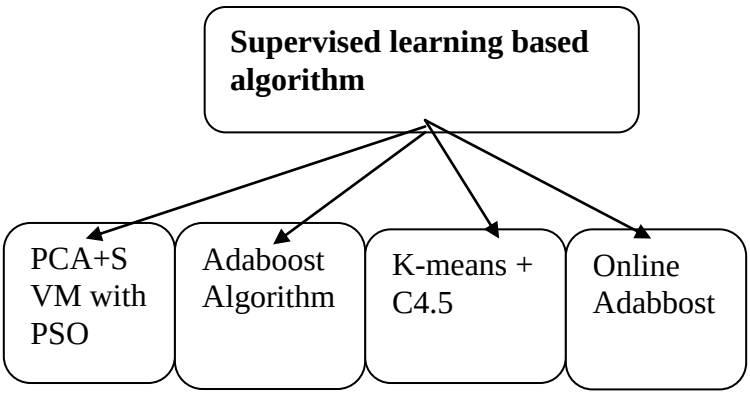

Fig 4 Classification Supervised learning based algorithm

5.1.1 PCA+SVM with PSO Enhancement

The detection model based on SVM combines PCA and PSO for anomaly detection. Principal Component Analysis (PCA) is used to reduce the dimensions of data. Particle Swarm Optimization algorithm is used to optimize the factors in SVM. PSO performs searches using a population called a swarm. Each particle moves in the direction of the previously best position and the best global position is used to find the optimal solution. The fitness of each particle is compared with its fitness value to update the particles best value.

5.1.2 AdaBoost algorithm-based approach

The detection model based on AdaBoost algorithm is used for anomaly detection. The decision stumps are used as weak classifier, and rules are provided for both categorical and continuous features. The data are classified for training that contains both normal data and attacks. The normal data is classified as "+1" and attack data is classified as "-1". A decision stump is built for each feature of the data. Strong classifier is attained by cascading the weak classifiers for the classification of network attacks.

5.1.3 K-Means + C4.5 Decision Tree-based Detection

The detection model using K-Means and C4.5 is used for distinguishing the normal and anomalous activities in a computer network. Network intrusion detection system (NIDS) allows detecting the security policy violations. The process of cascading the K-Means and C4.5 method includes two phases:

i) Selection phase

ii) Classification phase.

In the selection phase, Euclidean distance is measured to identify the closest cluster and C4.5 decision tree is employed to handle the neighbor cluster, whereas in the classification phase, C4.5 decision tree are computed on the test instance to classify it as normal one or anomalous one.

5.1.4 Online AdaBoost-based Intrusion Detection Approach

In the online AdaBoost classifier, for each network connection, weak classifiers are constructed for both the continuous and categorical features . Local intrusion detection models are designed using two algorithms:

i) AdaBoost algorithm and decision stumps
ii) Online AdaBoost classifier and
iii) Online Gaussian Mixture Models (GMM).

These are considered as weak classifiers. The local parametric models for intrusion detection are shared between the nodes of the network. Particle swarm optimization and support vector machine are used to cascade the local detection models into a global detection model.

Table 1. Comparison of different algorithm.

Sr. No	IDS	Detection Rate (%)	False Alarm rate (%)
1	PCA+SVM With PSO	99.75	Not discussed
2	Adaboost Algo	90.88	1.7
3	K-means + C4.5	99.6	0.1
4	Online Adaboost	99.99	0.39

5.2 Unsupervised Learning-based Anomaly Detection

In the unsupervised learning algorithm, the learning process is unsupervised. The input data are not class labeled and mostly clustering algorithms are used to discover classes within the data. Some of the unsupervised-learning based algorithms are hyper spherical cluster-based approach and principal component classifier based detection approach. They operate in both the centralized and distributed manner.

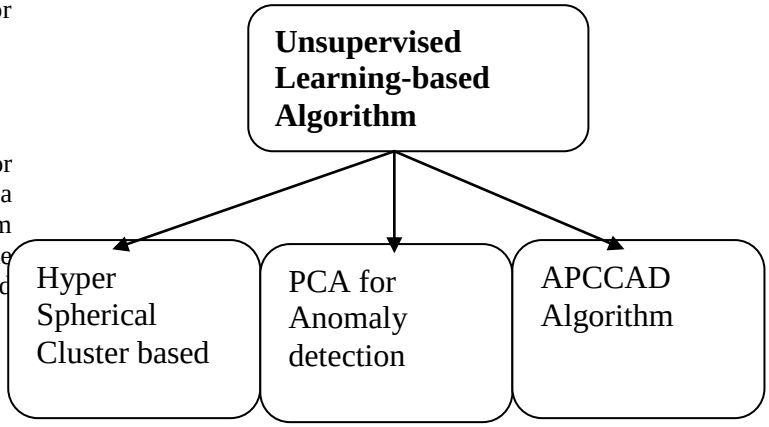

Fig 5 classification of unsupervised learning based algorithm

5.2.1 Hyper Spherical Cluster-based Anomaly Detection

A hyper spherical cluster based detection algorithm is used for identifying anomaly in WSN. In the centralized approach, each sensor node sends all its data to the gateway node and combines all data to form a combined dataset. The fixed-width clustering algorithms are used for anomaly detection. They randomly choose data points as centroid and Euclidean distance is measured between the centroid and next remaining data vector. If the distance to the closest centroid from a data is less than radius, then the data is

added to that cluster, otherwise a new cluster is formed. In the distributed approach, the model is scattered to all sensor nodes. The local anomalies are detected and clusters are classified as normal or anomalous using K-NN classifier. Fig.5 shows centralized and distributed approach.

5.2.2 Anomaly Detection using Principal Component Analysis

A new distributed online anomaly detection model is designed to the dimension reduction. The candid-covariance free principal component analysis is utilized for data reduction in IDS. It includes two main stages. They are:

i) Training stage
ii) Detection stage.

In the training stage, the observations are gathered at every sensor node to find the local normal model and sent it to the cluster head to create a global normal model. The detection threshold that represents the local normal model is chosen as the maximum and minimum range, whereas in the detection stage, every sensor observation is classified as normal or anomalous by analyzing the detection threshold from global normal model.

5.2.3 Adaptive Principal Component Classifier-based Anomaly Detection

The principal component classifier-based anomaly detection model is designed to detect anomalous sensor measurements to track dynamic changes. The model has three phases:

i) Training phase,
ii) online detection phase and
iii) Update phase.

In the training phase, the standard data analysis is collected at each sensor node to frame normal model. The detection phase compares each data with normal model framed in the training phase to classify the data as normal or anomalous. In the update phase, the normal model is retrained to produce a new normal reference model. The performance comparison of unsupervised anomaly detection algorithms in terms of detection rate (DR) and false alarm rate (FAR) are shown in Table.2

.

Table 2 Comparison of unsupervised algorithm

Sr. No	IDS	Detection Rate (%)	False alarm rate (%)
1	Hyper Spherical cluster based	85.47	1.48
2	PCA	82.86	13.3
3	APCCAD	97.84	1.10

VI. CONCLUSION

The network security is progressively used as an important terrace to aggregate and observe data from untended environments. The deployment of constrained sensor resources in such environment is susceptible to a variety of potential attacks. There are various intrusion detection mechanisms that are used to identify the attacks in a network with high detection rate. Anomaly detection approaches are modeled to identify the deviations in the system due to unknown attacks. To handle the detection systems that rely on human intervention, machine-learning based anomaly detections are designed that are capable of detecting novel attacks. Compared with classification of machine-learning-based anomaly detections, it is observed that unsupervised-learning based anomaly detection is efficient in unknown attack detection.

REFRENCES

[1] Adriana-Cristina Enache Intrusion Detection Based On Support Vector Machine Optimized With Swarm Intelligence, presented at 9th IEEE International on Applied Computational Intelligence and Informatics (2014), Timisoara, Romania.

[2] A.M.Chandrasekhar and K.Raghuveer, Intrusion Detection Technique by using K-means, Fuzzy Neural Network and SVM classifiers, presented at International Conference on Computer Communication and Informatics (ICCCI-2013), Coimbatore, INDIA.

[3] Sandip Ashok Shivarkar, and Mininath Raosaheb Bendre, Hybrid Approach for Intrusion Detection Using Conditional Random Fields, International Journal of Computer Technology and Electronics Engineering (IJCTEE) Volume 1, Issue 3.

[4] Annie George, Anomaly Detection based on Machine Learning: Dimensionality Reduction using PCA and Classification using SVM, International Journal of Computer Applications (0975 to 8887) Volume 47 and No.21, June 2012.

[5] Kui W. Mok, A data mining framework for building intrusion detection model, In: Gong L., Reiter M.K. (eds.): Proceedings of the IEEE Symposium on Security and Privacy. Oakland, CA: IEEE Computer Society Press, pp.120 - 132, 1999.

[6] V. Jyothsna, V. V. Rama Prasad, K. Munivara Prasad, A Review of Anomaly based Intrusion Detection Systems, International Journal of Computer Applications (0975 to 8887) Volume 28 No.7, August 2011.

[7] CHEN Bo, Ma Wu, Research of Intrusion Detection based on Principal Components Analysis, Information Engineering Institute, Dalian University, China, Second International Conference on Information and Computing Science, 2009.