Hybrid Model For Intrusion Detection Using Naive Bayesian And Support Vector Machine

Amit D. Sagale, S.G. Kale
Department of Information Technology
Yeshwantrao Chavan College of Engineering, Nagpur-10, Maharashtra, India
Email: amitkayas@gmail.com, s12_kale@gmail.com

Abstract—Computer networks are nowadays subject to an increase number of attacks. Intrusion Detection Systems (IDS) are designed to protect them by identifying anomaly behaviours or improper uses. Since the scope is different in every case (register already-known menace to later recognize them or model legal uses to trigger when a variation is detected), so far to respond against both kind of attacks IDS have failed. System developed with the single algorithms like classification, neural networks, clustering etc. gives better detection rate and less false alarm rate. Recent papers show that the cascading of two algorithm yields much better performance than the system developed with the single algorithm. Intrusion detection systems that uses solo algorithm, the accuracy and detection rate were not up to mark. Increase in the false alarm rate was also encountered. Cascading of algorithm (Hybrid model) is performed to solve this problem. This paper represents two algorithms for developing the intrusion detection system. Naïve Bayesian (NB) and Support Vector Machine (SVM) are combined to maximize the accuracy, which is the advantage of NB and diminish the wrong alarm rate which is the advantage of SVM. In this paper we proposed hybrid model which give higher detection rate and low false positive rate for IDS.

Keywords : NB, IDS, SVM, Hybrid

I. INTRODUCTION

Nowadays, many organizations use Internet services as their communication and marketplace to do business. Together with the growth of computer networks uses, the growing rate of network attacks has been advancing, impacting to the availability, confidentiality, and integrity of critical information data. Therefore a network administrator must use one or more security tools such as firewall, antivirus, IDS and Honey Pot to prevent important data from criminal enterprises. A network system using a firewall only is not sufficient to prevent networks from all attack types. The firewall cannot protect the network against intrusion attempts during the opening port. Hence a Real-Time Intrusion Detection System is a prevention tool that gives an alarm signal to the computer user or network administrator for opposed activity on the opening session, by inspecting hazardous network activities.

There are two general categories of attacks which intrusion detection technologies attempt to introduce anomaly detection and misuse detection. Anomaly detection identifies all activities that different from established patterns for users, or groups of users. Anomaly detection usually involves the creation of knowledge bases that contain the profiles of the monitored activities. The second general categories to intrusion detection are misuse detection. These techniques involve the comparison of a user’s activities with the known behaviors of attackers attempting to penetrate a system. While anomaly detection typically utilizes threshold monitoring to indicate when a certain well-known metric has been reached, misuse detection techniques frequently use rule-based approaches. When applied to misuse detection, the rules become follows for network attacks. The intrusion detection mechanism identifies a potential attack if a user’s activities are found to be consistent with the established rules. The use of full rules is critical in the application of expert systems for intrusion detection.

We present hybrid approaches for modeling IDS. Naive Bayesian (NB) and Support Vector machines (SVM) are combined as hierarchical hybrid intelligent system model (NB-SVM) and an ensemble approach combining the base classifiers. The hybrid intrusion detection model combined the individual base classifiers and other hybrid machine learning paradigms to maximize detection accuracy and minimize computational complexity.

II. INTRUSION DETECTION SYSTEMS

Intrusion detection systems are used to identify, classify and possibly, to respond to benign activities. Also,
Intrusion Detection System (IDS) is used to monitor all or traffic, detect malicious behaviour activities, and respond to the activities. Network intrusion detection system was establish for the purpose of malicious activities detection to strengthen the security, confidentiality, and integrity of critical information systems. These systems can be network-based or host-based.

HIDS(Hybrid intrusion detection System) is used to analyze the internal event such as process identifier while NIDS is to analyze the external event such as traffic volume, IP address, service port and others.

The challenge of the study is how can we have an IDS with higher detection and low false positive rate? [4]

Intrusion detection has two main techniques which are misuse-based intrusion detection and anomaly based intrusion detection. Misuse-based intrusion detection IDSs that employ misuse detection approach detect attacks by comparing the predefine signatures against the network traffics captured by the IDSs. When a match is found, the IDSs will take action as the traffics are considered unsafe to computer systems or computer networks.

Actions taken by the IDSs will normally include sending alerts to network administrator.

IDSs that implement misuse detection approach are, however, incapable of detecting novel attacks. The network administrator will need to update the stored signatures frequently to make sure that the IDSs perform well in detecting intrusions.

[5]Anomaly based intrusion detection IDSs that employ anomaly detection are capable of identifying new attacks, that contain activities deviate from the normal.

Such IDSs utilize the build profiles that are learned based on normal activities in computer networks. This system has two stapes:

1) Learning: It works on profiles. The profiles represent the normal activities of the users, systems, or network connections, applications. Great care should be taken while defining profiles because currently there is no effective way to define normal profiles that can achieve high detection rate and low false positives at the same time.

2) Detection: The profile is used to detect any deviance in user normal behaviour. [7]

Different Types of Attacks

- Denial of Service (DOS): Making some machine resources too busy to answer to legitimate users requests.
- User to Root (U2R): Exploiting vulnerability on a system to obtain a root access.
- Remote To Local (R2L): Using vulnerability in order to obtain a local access like a machine user.
- Probing: Collecting useful information or known vulnerabilities about a network or a system. [8]

III. MACHINE LEARNING ALGORITHM:-

Machine learning studies how to automatically discover to make accurate predictions based on past observations. This type of algorithm we provided Knowledge with result for learning purpose and then we providing knowledge algorithm give result on past observation. At the time of learning algorithm makes some rules or threshold value for each classifier. There are many algorithm are used as machine like Decision tree, Naive Bayesian, Support Vector Machine etc.

1. Naive Bayesian

Naive Bayesian classifier is a simple classification scheme, which estimates the class-conditional probability by assuming that the attributes are conditionally independent, given the class label c. The conditional independence assumption can be formally stated as follows:

\[ P(A | C = c) = \prod_{i=1}^{n} P(A_i | C = c) \]  

(1)

Where each attribute set A = {A1,A2,…….An} consists of n attribute values. With the conditional independence assumption, instead of computing the class conditional probability for every grouping of A, only estimate the conditional probability of each Ai, given C. The latter approach is more practical because it does not require a very large training set to obtain a good estimate of the probability. To classify a test example, the naive Bayesian classifier computes the posterior probability for each class C.

\[ P(C | A) = \frac{P(C)\prod_{i=1}^{n} P(A_i | C)}{P(A)} \]  

(2)

Since P(A) is fixed for every A, it is sufficient to choose the class that maximizes the numerator term.
The naïve Bayesian classifier has several advantages. It is easy to use, and unlike other classification approaches, only one time scan of the training data is required. The naïve Bayesian classifier can easily handle missing attribute values by simply omitting the probability when calculating the likelihoods of membership in each class.

2. Support Vector Machines

Support Vector Machines have been proposed as a novel technique for intrusion detection. A Support Vector Machine (SVM) maps input (real-valued) feature vectors into a higher dimensional feature space through some nonlinear mapping. SVMs are powerful tools for providing solutions to classification, regression and density estimation types of problems. These are developing on the principle of structural risk minimization. Structural risk minimization seeks to find a hypothesis for which one can find lowest probability of error. The structural risk minimization can be achieved by finding the hyper plane with maximum separable margin for the data.

Computing the hyper plane to separate the data points i.e. training a SVM leads to quadratic optimization problems. SVM uses a feature called kernel to solve this problems. Kernel transforms linear algorithms into nonlinear ones via a map into feature spaces. There are many kernel functions; some of them are Polynomial, radial basis function, two layer sigmoid neural nets etc. The user may provide one of these functions at the time of training classifier, which selects support vectors along the surface of this function. SVMs classify data by via these support vectors, which are member of the set of training inputs that outline a hyper plane in feature space. The implementation of SVM intrusion detection system has two phases: training and testing.

The main advantage of this method is speed of the SVMs, as the capability of detecting intrusions in real-time is very important. SVMs can learn a larger set of patterns and be able to better scale, because the classification complexity does not depend on the dimensionality of the feature space. SVMs also have the ability to update the training patterns dynamically whenever there is a new pattern during classification. The main disadvantage is SVM can only handle binary-class classification whereas intrusion detection requires multi-class classification.

IV. PROPOSED ALGORITHM:-

In this hybrid model we cascading naïve Bayesian and Support Vector Machine. First we implement Naïve. Bayesian which can divide dataset in two classes i.e. normal and anomaly. Then this anomaly dataset is provided to Support Vector Machine. As SVMs are able to handle only binary class classification problems, we need to employ four SVMs, for the 4-class classification intrusion detection. We divided the anomaly dataset into the four classes of attacks (Probe, DOS, U2R, and R2L).

Following are step included:-

1) The proposed algorithm first initializes the weight of training examples to 1/n, where n is the total number of examples in training dataset, and then creates a new dataset from training dataset using selection with replacement technique.

2) After that it calculates the prior and conditional probabilities of new dataset, and classifies the training examples with these probabilities value.

3) The weights of the training examples updated according to how they were classified. If a training example is misclassified then its weight is increased, or if correctly classified then its weight is decreased.

4) Then the algorithm creates another new data set with the misclassification error produced by each training example from training dataset, and continues the process until all the training examples are correctly classified.

5) To classify a new example use all the probabilities in each round and consider the class of new example with highest classifier's vote.

6) Then anomaly dataset provided to SVM for further classification.

V. CONCLUSION

In this paper we proposed hybrid model for intrusion detection. we cascading Naïve Bayesian and Support Vector Machine and evaluated their performance on the benchmark KDD Cup 99 Intrusion data. We have first explored a Naïve Bayesian as an intrusion detection model. We also conducted experiments with support vector machines (SVM) As the Naïve Bayesian was used as a binary classifier , then we provided dataset SVM to employed four classifiers for 4-class classification. The empirical results Naïve Bayesian indicate that Naïve Bayesian gives better accuracy for two class classification.

REFERENCES:


Modeling Intrusion Detection Systems Using Hybrid Intelligent Systems Sandya Peddabachigari, IEEE-2005


Feature selection and design of Intrusion Detection System based on k-means and triangle area support vector machine, Pingjie Tang, Ranq-an Jiang, Mingwei Zhao. IEEE-2010.

A Hybrid And Hierarchicanli Ds Paradigm Utilizing Naïve Bayes Classifier Qin ZHAO, Jizhou SUN, Song ZHANG IEEE-2004.


A New Data Mining Based Hybrid Network Intrusion Detection Model Virendra Barot, Durga Toshniwal IEEE-2012