Intrusion Detection Based on Information Entropy of Multiple Support Vector Machine

Li Wei
School of Mathematics and Computer Science, Yunnan University of Nationalities
Kunming 650031, China
Email: weili19891 {at} 163.com

Abstract—**Intrusion detection is essentially a classification problem. It is very important to increase the classification accuracy. Support vector machine (SVM) is a kind of very effective tool to solve classification problems, accuracy of SVM intrusion detection Based on information entropy is relatively high, but how to achieve higher accuracy is still a new issue. In this paper, we use multiple SVM intrusion detection algorithm based on information entropy to solve the problem, first, we use information entropy theory to optimize the SVM kernel function, then reduce and reclassify the Characteristic attributes of each Network connection, finally, training each Reclassification set of attributes with two Support vector machine, the experimental results show that the proposed method is an efficient network intrusion detection method.**

Keywords*—*Intrusion detection*; *information entropy*; *Support vector machine

I. INTRODUCTION

With the development of network and information technology, application of network in the political, economic, and daily life is more widely. However, the ensuing increasing number of cyber attacks must arouse our attention, therefore, we introduction the network intrusion detection to protect the security of information. Intrusion detection system as a powerful complement of static firewall technology, has became the most important part of the deep defense system of network security. It is a network security technology that is used to prevent tampering, deleting, and plagiarism. According to the detection method, intrusion detection technology can be divided into two types traditionally. That are: Misuse intrusion detection and anomaly intrusion detection[1].The foundation of Misuse intrusion detection is to establish hacking feature library, use feature matching method to determine attacks, this method can detect known attacks accurately, but is powerless against unknown attacks. Anomaly intrusion detection need to establish a normal user behavior patterns, whether to deviate significantly from the normal behavior patterns as a basis for testing, because it can detect unknown and potential attacks, it has become a research hotspot.

II. SVM INTRUSION DETECTION MODEL BASED ON INFORMATION ENTROPY

Research data packets and their characteristics statistical model during the transmission of the network, its essence is research the uncertainty of information, the relation between the information entropy and its degree of uncertainty is equivalence,to be able to carry out a unified quantitative calculation of data packets and its characteristics, we selected entropy as a standard to measure their importance. Fuse the information entropy theory and support vector machine algorithm, on the one hand make up the deficiencies of support vector machine algorithm, on the other hand, can take advantage of the high accuracy of information entropy on intrusion feature discriminant[3][4].

There is always an information entropy difference between the monitored Computer system or network and the open network, the direction of information flow is always from the large entropy value to the small entropy value; on the contrary, the penetration message from the direction entropy value is low to the direction entropy value is high is not only pure but also small. Facilitate the study of the contour of their state. Based on this theory, SVM intrusion detection model based on information entropy can contain the following aspects: SVM algorithm combined with information entropy theory, to reduce feature vector and find the optimal parameters of SVM kernel function.

Model-based

Here we assume that the monitored computer system is known as the starting point, select suitable deployment location for the intrusion detection system,obtain its network data to conduct research.

Suppose one Computer system or network to be monitored are known.

As shown below, M_i is the internal computer system or network to be monitored, M_0 is the external open network，IDS_1 is the intrusion detection system deployed between them, Since M_i is known, we can not only determine its users, computer systems and network configurations, but also can obtain the sample data of M_i or all data[5]. Through observation and research on M_i, you can determine there exist a group of metrics which can be expected to operate that describe the users or processes. By contrast, M_0 is the external open network, even there exist a predictive model of M_0, adequate sampling samples still can not be guaranteed, it can be concluded that: research of intrusion

detection technology must be based on the condition that the monitored computer system or network is already known.

Since M_i is known, we can get a network connection set Y between M_i and M_0, IDS_1 is the intrusion detection system deployed between them, the input of IDS_1 is Y, the output of it is Z, Where +1 represents the normal network connection, -1 indicates abnormal network connection..

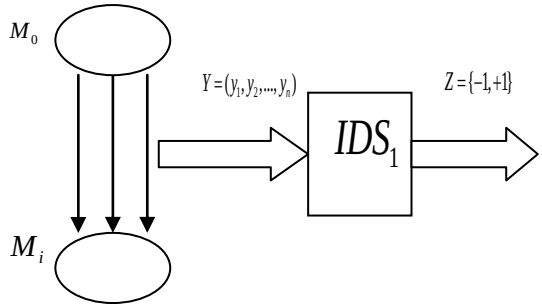

Figure1. IDS network structure

Optimize the SVM kernel function

Each network connection between M_i and M_0 can be considered as a intrusion event of random sample, through the observation and statistical analysis of test results, we can get an overall understanding of the Intrusion event.

Suppose two when the Monitored computer system or network is not under attack, the operating of Users and processes comply with a series of provisions, does not contain the sequence of commands which damage system security policy, generally conformity with statistical forecasting model.

If the users and processes of M_i meet statistical forecasting model, you can use some of the Characteristic parameters of M_i and closed-value to define the behavior of normal users and the normal contour of the system, then compare the transient contour with the normal contour, if the difference exceeds the tolerance threshold, it is determined to be abnormal. This paper introduces the information entropy to improve the process of building normal contour of M_i.

Based on the above assumptions, the process is as follows:

Step one: Extracted l kind of feature from each connection to represents a connection, namely $y_i = (c_1, c_2, ..., c_l)$.

Step two: Standardize the raw data. Since the original data has 41 feature properties, and the range for each attribute is not the same, so we need to do canonicalized processing for the data, normalized the values for each attribute to $\{-1, +1\}$. In this paper, we use the amount of information to replace features, obtained the amount of information of $c_i : c_i^{'} = -1bp_{c_i}$, measurement problems of unification of feature information is solved through using the amount of information, eliminates the uncertainties of feature information, Get $y_i^{'} = (c_1^{'}, c_2^{'}, ..., c_l^{'})$ indicates the connection y_i, using the training sample set for learning to get entropy $E_{M_i} = (E_{c_1^{'}}, E_{c_2^{'}}, ..., E_{c_l^{'}})$ and variance $D_{M_i} = (D_{c_1^{'}}, D_{c_2^{'}}, ..., D_{c_l^{'}})$, through using E_{M_i} and D_{M_i} ,we can describe Contour value of normal mode of M_i.

Step three: Set confidence interval based on the entropy and variance of the amount of feature, if $c_i^{'}$ is in this range is normal, Denoted +1, if $c_i^{'}$ is beyond this range is abnormal, Denoted -1.The input is: $y_i^{'} = (c_1^{'}, c_2^{'}, ..., c_l^{'})$, the output is: $z_i^{'} = (I_{c_1^{'}}, I_{c_2^{'}}, ..., I_{c_l^{'}})$.

Step four: Find an optimal classification hyperplane through SVM, divide the given input sample into two categories: normal and abnormal, make the classification Interval between the two types of data as large as possible. The problem is transformed into finding a mapping: $g : Z^{'} \to Z$, $z_i^{'} \in Z^{'}$, $Z = \{-1, +1\}$; minimize the risk of misclassification, this transformation may be more complex, this idea is not easy to achieve under normal circumstances. However, note that the optimization function involves only the inner product calculation between the training samples, in fact, in high dimensional space we actually only need Conduct inner product operation, this inner product operation can be alized. According to the theory of Functional Analysis, as long as the kernel function satisfies Mercer condition, it corresponds to the inner product of a particulartransformation space.The corresponding classification function is:

$$g(y^{'}) = \mathrm{sgn}\{\sum_{i=1}^{n} a_i^{*} z_i^{'} K(z_i^{'} \bullet z^{'}) + b^{*}\} \qquad (1)$$

In a nutshell, introduce information entropy into the SVM detection modeling, base feature statistical laws, use the amount of information of feature to represent the feature itself, standardize the metrics of the characteristics of the original data [6][7].

The reduction of the characteristic attribute

Using support vector machine detection technique transforms the input space into a high dimensional space, to find the optimal classification surface in this space. However, in practice the training speed and the detection efficiency both need to improve, the number of features of training samples decide the dimension of the matrix in the objective function of the quadratic programming problems, each added characteristic may increase the cost and the run time of the system, make the speed and dimension appear exponential growth in solving linear programming problems, Therefore, it

difficult to apply to practical intrusion detection.

In order to solve the above problems, this paper re-examine the original data, According to the distribution characteristics of characteristic attributes and the association between the characteristics, the basic attribute set, the traffic attribute set, the content attribute set and the host attribute set, make a redrawing category for the original basic attribute set, the traffic attribute set, the content attribute set and the host traffic attribute set [8].

Suppose three The uplink data between the monitored computer system or network and the open network, contains all data and command sequences of users and process operations.

From the above assumption, divide the 41 characteristic attributes of each connection into three categories: Uplink attribute set、the downlink attribute set and the status attribute set. First, stripped the downlink attribute set, selected information entropy as a standard for measuring the importance of the characteristic, through observing the degree of change of the detection correct rate after delete a characteristic, sort characteristics based on their importance [9]. This approach not only ensure the selected characteristic subset without losing the information the original input space contained in, but also reduces the characteristic dimension of the sample, effectively reducing the calculation scale of real-time detection.

III. INTRUSION DETECTION METHOD BASED ON THE INTEGRATION OF MULTIPLE SUPPORT VECTOR MACHINES

Single SVM applied to intrusion detection, although can get a higher intrusion detection rate, but it also brings a relatively large false rate at the same time, Therefore, put multiple support vector machine classifiers for decision-making combination using $D-S$ evidence theory, then use two SVM to train the separated uplink attribute set and the status attribute set, in order to improve the detection rate of intrusion detection, and reduce the false detection rate at the same time[10].

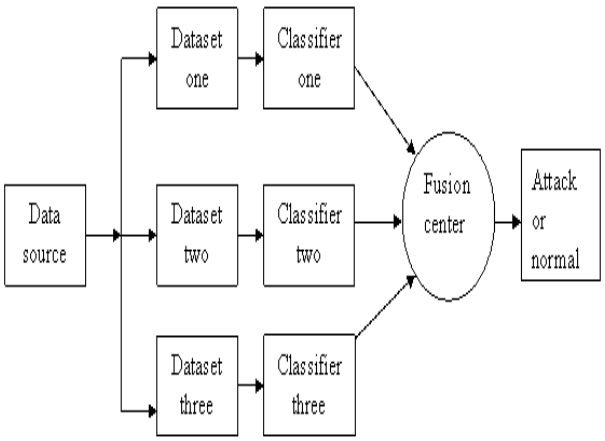

Figure2. Classifier Module based on the fusion of multi-SVM

Based on the above multi-SVM intrusion detection model, based on using the KDD99 data set, proposed a new fusion strategy namely $D-S$ evidence theory, this theory is the expansion of the classical probability model, and it has a good theoretical basis, its main advantage is that when the conflict between the evidence is smaller, degree of confidence for evidence focus on a smaller uncertainty, the advantage of the new integration strategy is that two classifiers are used for each characteristic attribute set, so this method can ensure a better classification accuracy and a relatively low false alarm rate, meanwhile the execution efficiency of a single classifier is high, the entire integration PROCESS IS SIMPLE AND EASY TO IMPLEMENT.

IV. SIMULATION EXPERIMENT AND RESULT ANALYSIS

The experiment data of this paper comes from KDD99 Dataset, simulate the Intrusion detection test environment of U.S. military LAN, provided by the Lincoln Laboratory of MIT. In this experiment, we select two hundred, five hundred, one thousand training data as candidate sample set from the training set of KDD99 dataset.

In order to compare the experimental results with multiple entropy-based SVM intrusion detection Algorithm, for each set of training data sets, we also use the traditional entropy-based single SVM intrusion detection algorithm for experiments, the experimental results are shown below.

TABLEI SVM ALGORITHM PERFORMANCE ANALYSIS

Total number of samples	**single SVM intrusion detection algorithm based on information entropy**			**intrusion detection algorithm of the integration of multiple SVM based on information entropy**		
	Detection rate	False alarm rate	False negative rate	Detection rate	False alarm rate	False negative rate
200	95.41%	0.77 %	1.23%	96.21%	0.70 %	1.17%
500	95.93%	0.63 %	1.07%	96.64%	0.58 %	0.98%
1000	96.72%	0.50 %	0.83%	97.35%	0.45 %	0.73%

By comparison the data in the table above, we find under the condition that the total number of samples are the same, the detection rate of this method is higher than the entropy-based single SVM intrusion detection algorithm, both the false alarm rate and the false negative rate are lower than the entropy-based single SVM intrusion detection algorithm, therefore, the method of this paper is effective.

ACKNOWLEDGMENT

This research is supported by Foundation of Educational Committee of Yunnan Province (Grant No. 2013J119C).

REFERENCES

1. YangYi-xian,NiuXin-xin.Intrusion Detection theory and technique[M]Beijing：Higher Education Press,2006(in Chi-nese).

2. Tang Zhengjun,Li Jianhua. Intrusion Detection Technology[M]. Beijing: Tsinghua University Press,2004.
3. Corts C, Vapnik V. Support vector networks [J]. Machine Learning, 1995, 20: 273-297.
4. Bace R G. Intrusion Detection [M] . US: Macmillan Technical Publishing, 1999.
5. ZhuWenjie, Wang Qiang.SVM intrusion detection technology based on nformation entropy[J]. Computer Engineering and Science, 2013.
6. Denning D E.An intrusion detection model [J].IEEE Transaction on Software
7. Engineering,1987,13(2):222-223.
8. Sommer R, Paxson V.Outside the closed world : On using Machine learning for Network intrusion detection [C] // Proc of2010IEEE Symposium on Security and Privacy 2010:305-316.
9. Sommer R, Balzarotti D. Recent advances in intrusion detection[C] // Proc of RAID'11,2011:1.
10. Li Kunlun, Huang Houkuan, Tian Shengfeng. Fuzzy Multi-Class Support Vector Machine and its Application in Intrusion Detection[J]. Computer Journal, 2005,28(2):274-280.
11. Xiao Yun, Han Chongzhao, Zheng Qinghua. Network intrusion detection method based on multi-class support vector machine[J]. Journal of Xi'an Jiaotong University, 2005,39(6): 562-565.