Research on Property and Model Optimization of Multiclass SVM for NIDS

Jianhao Song, Gang Zhao, Junyi Song
School of Information Management, Beijing Information Science & Technology University
Beijing, China
zhaogang@ieee.org

***Abstract*—By investigating insufficiency of typical artificial intelligence algorithms aiming at the high rate of False-Positives and False-Negatives in the Intrusion Detection Systems (IDS), this paper presents an approach that Support Vector Machine(SVM) is embedded in Network Intrusion Detection System (NIDS). At the same time, by using online data and K-fold cross-validation method, this paper proposes a method to optimize the attributes and model of SVM respectively. Experimental results show that by using this method as the detection core of the intrusion detection system, the rate of False-Negatives in IDS can be reduced significantly.**

Keywords-IDS; SVM; online detection; rate of False-Negatives; rate of False-Positives

I. INTRODUCTION

Since the model of intrusion detection is proposed in 1987, intrusion detection has gone through a rapid development during the just 20 years. Invasion is any action can damage computers or confidentiality, integrity, availability of network. The intrusion detection system (IDS) is an important part of the deep defense system of network security. It discovers identifies intrusion actions and attempts to system and gives alarms through monitoring and analyzing network flow, auditing system log or other methods. Intrusion detection methods can generally be divided into two categories: Signature-based IDS and Statistical anomaly-based IDS. Signature-based IDS monitors packets in the Network and compares with pre-configured and pre-determined attack patterns known as signatures. A statistical anomaly-based IDS determines normal network activity like what sort of bandwidth is generally used, what protocols are used, what ports and devices generally connect to each other and alert the administrator or user when traffic is detected as anomaly.

Because of the inherent flaws existed in two methods mentioned above, rates of False Positive and False Negative are high. Thus, researchers introduce machine learning methods to solve the problems of data processing. However, most traditional machine learning algorithms are based on assumptions that amount of samples tends to infinity and required high data regularity. Until recently, various intelligent IDS cannot create ideal results.

This paper aiming at many problems encountered in IDS, put forward that apply the statistical learning theory which specifically studies the small sample learning in the field of machine learning and adopt the Support Vector Machine (SVM) [1]-the most mature method in statistical learning theory-to resolve the heavy overhead, slow detection rate and high rate of False-positives and False-negative problems in NIDS. In addition, this paper also proposes methods to select and optimize the attributes and model of SVM. Finally, this paper uses port scan and DoS attacks software to test and validate, gets rid of the offline test limitation [2] in existing studies.

II. INTRODUCTION OF SVM PRINCIPLE

SVM is the machine learning technique developed in mid-1990s and popular statistical learning method in common use. SVM method is based on VC Dimension theory and structure risk minimization principle of statistical learning theory. With limited sample information, SVM can reach the best compromise between complexity and learning ability of model, in order to obtain the best generalization ability [3]. When comparing with traditional neural networks method which based on empirical risk minimization, SVM not only has more simple structure, but also has better generalization ability for small samples.

SVM is evolved from optimum classification surface in the linear separable case. The basic idea can be explained in two dimension condition in Figure 1.

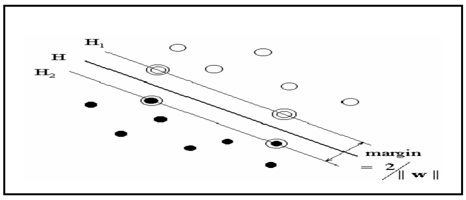

Figure 1. The optimum classification in two dimensional cases.

In the Fig.1, solid points and hollow points represent two kinds of samples. H is sorting line that can correctly separate two types of samples. H1 and H2 are straight lines parallel to H and cross all samples which nearest to H respectively. Distance between them called classification intervals or margin. The so-called optimum sorting line is required capabilities not only separate two types correctly, but also maximize the margin. Separate two types of training samples correctly means ensure empirical risk minimization, maximize the margin means minimize confidence range in promotion border, thus minimize the real risk.

Under linear separable case, it can converse issue of configure optimum hyper-plane to calculate the minimum value of $\Phi(w)= \| w \|^{2}$. Support vector interval can be calculated as $2/\| w \|$. Distance between any point x and Hyper-plane is $(w\times x+b/\| w \|)$, while the optimum hyper-plane which having a maximum interval need in such

Published by Atlantis Press, Paris, France.

condition: the number of VC dimension of regulation hyper-plane should not bigger than min([R2,A2],n)+1, n is amount of dimensions of vector space, all vectors waiting for split are located within a hypersphere have radius R, and $\|w\| \leq A$. Build the optimum hyper-plane to split two types can be transformed to quadratic programming, that is calculate the minimum value of $w^2/2$ [4] when $yi(w \times xi+b) \geq 1(i=1\ldots l)$.

III. SNORT Intrusion Detection System

Structures and principles of traditional IDS is much the same, and detection cores are all rules-match according to the existing rule.

A. Introduction of SNORT

SNORT is a powerful lightweight NIDS that has abilities to analysis real-time data, match content in network data and log. It can detect a variety of attacks and provide real-time alarm for those attacks, and runs on a host to monitor the network data. SNORT can match patterns between the network data and detection rules, thereby detecting a variety of possible intrusion attempts. Moreover, SNORT also has good scalability and portability. Figure 2 shows the workflow of SNORT.

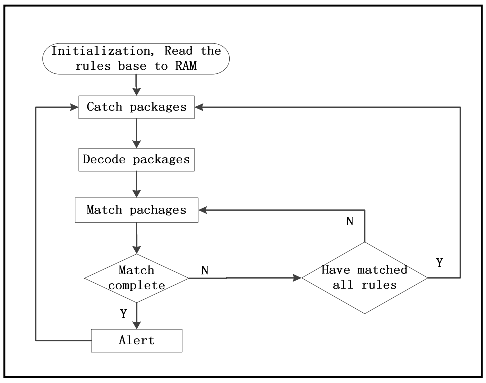

Figure 2. Workflow of SNORT.

B. Bottlenecks in SNORT Development

Similar to some other traditional IDS, the development of SNORT has encountered a bottleneck. IDS face the following major problems.

IDS have several commonly detection methods include feature detection, anomaly detection, state detection, protocol analysis, etc. These detection methods all have flaws. Such as anomaly detection commonly used statistical methods to detect, but it is difficult to effectively determine the threshold of statistical methods, small value will produce a lot of false-positives, big value will produce a large number of false-negatives. In protocol analysis detection methods, the normal IDS just simply dealt with commonly protocol such as HTTP, FTP, SMTP, etc. The large number of rest protocol packets entirely possible cause false-negatives. If consider supporting as many as possible of the protocol type analysis, the cost of the network will not be able to afford.

IDS can only identify the IP address, cannot locate the IP address and identify the data source. When IDS found the attack events, it can only close network exports and a few ports of server, but this close will also affect other normal users' use. Thus, it lacks a more effective response handling mechanism.

Existed IDS products are mostly used feature detection technology, these IDS products cannot adapted exchange technology and development of high-bandwidth environment, in the case of large flows impact and multi-IP fragmentation, IDS will paralysis or loss packages and then form DoS attacks.

C. Traditional solutions and its defects

1) Improvement of analytical techniques

The final solutions of false-positives and false-negatives in intrusion detection rely on improvements of analysis technology. Mainly analysis methods of intrusion detection include statistical analysis, protocol analysis, behavior analysis, etc. By count the occur number of relevant events in network, Statistical analysis can discriminant attacks. However, this method can only be applied in some types of attacks, such as DoS attacks, effect is not obvious for some attacks with small data scale.

Based on reorganization of network data stream, protocol analysis technique can understand application protocols then use pattern matching and statistical analysis techniques to ascertain attacks. If use protocol analysis, alarm will emerge only when event is detected coincide with the protocol like HTTP. Assuming this feature appeared in Mail, because they do not coincide with the protocol, it will not alarm.

Behavior analysis technology is not only a simple analysis for single attacks but also according to the latter and previous event to confirm whether the attacks occurred, whether the attack behavior is in force. However, because of extremely difficult of algorithm processing and rule-making, it is not mature enough yet.

2) Improved the processing method of network with large amount of data

With requirements of large amount of data processing, performance requirements for IDS are also gradually increasing, hence Gigabit IDS and other products emerged. However, if intrusion detection products not only have attack analysis, but also had the function of contents recovery and network audit, system can hardly working under a Gigabit environment completely.

3) Interacting with firewall

IDS discover attacks then sent to firewall automatically, firewall load dynamic rules to intercept intrusion. This function called firewall interacting. It is not yet coming into practicality, primary a concept. Casual use can cause a lot of problems. It will make a negative impact on firewall stability and network applications if tested inadequately.

IV. Apply SVM Methods to Solve SNORT Problems

Above methods are not good solutions to the problems which IDS faced. So people put the focus on combination of AI methods and IDS. Most existing AI methods applied IDS are in offline testing stage, due to lack of a theoretical basis and methods itself have inherently flawed, those IDS cannot achieved ideal results.

This study through compare several machine learning algorithms like Bayesian, neural networks, decision trees and SVM, found that SVM algorithm is an ideal methods for invasion judgment[5] because sample it requires is small, classification is accuracy and other features. Therefore, this study decided to use SVM algorithm as detection core.

There are two advantages by using SVM as solution:

- As a classification algorithm, SVM calculated the optimum solution of distance, which is the fairest classification. It can resolve high rate of false-positives and false-negatives problems exist in IDS.
- Volume of model that SVM used for classification is small. Because using small samples to classify is an advantage of SVM, so it greatly reducing detection time compared to traditional IDS.

Through a lot of tests, we verified that aiming at IDS, SVM model are better than neural network, Naive Bayes and other traditional classification algorithms [6] both on training and testing.

A. Program Combination

Because IDS will ultimately be applied in actual network environment, this experiment decided to use famous NIDS, SNORT, as framework and SVM algorithm as detection core, applied SVM in practical online testing. Until now, most SVM applications are testing in offline environment; it cannot be run in online environments, which are the problem our studies can solve.

According to a variety of different types of intrusion actions online, this study proposed that based on BSVM, through modify SNORT, we can implement multiple classifications of intrusion actions, system workflow shown in Figure 3 below.

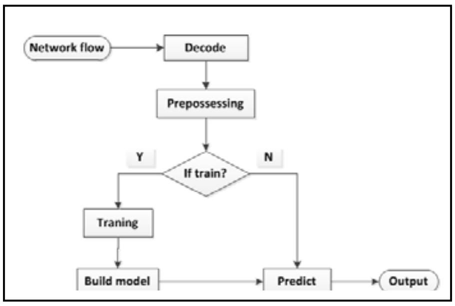

Figure 3. System workflow.

System decodes captured network flow from network card by Decode module, then transmits decoded data to Preprocessor module for preprocess, reassemble slice packets and unify format of URL string requested by HTTP. After preprocessing, data is converted into SVM acceptable format. If we want to train SVM model, system will pass data to SVM-train module, model will be generated after collect a certain amount of data. If we want to detect online data, data will be passed to SVM-predict module and predicted by BSVM algorithm will. At last system call Output module to output predicted results.

B. Attribute Optimization

The equations are an exception to the prescribed specifications of this template. You will need to determine whether or not your equation should be typed using either the Times New Roman or the Symbol font (please no other font). To create multileveled equations, it may be necessary to treat the equation as a graphic and insert it into the text after your paper is styled.

This study uses a famous port scanning software-Nmap, and DoS attack software-UDP flood to attack, uses SNORT to capture data.

Because of network data contains a large number of invalid attributes or attributes insignificant to judge incursions, so we need to filter captured data first. After considered comparison and verification, we selected 9 most characteristic attributes from 14 original attributes to consist vector, as shown in Table 1.

TABLE I. VECTOR ATTRIBUTE CONSISTENCE

Index	Feature	Content
1	ip_proto	protocol type
2	dport	destination port
3	th_flag	TCP flags
4	uh_len	packet length
5	packet_flags	packet flag
6	th_win	TCP window size
7	sport	source port
8	scan_thr	threshold of port scan
9	dos_thr	threshold of DoS attacks

C. Model Optimization

The most important part in SVM classification algorithm is model training. Model's quality determines accuracy and speed of classification. This study specializes trained and optimized the online data collected before and find the optimum model by contrast test. Another essential factor in the SVM model is selection of kernel function. Because of existing research cannot give apposite kernel function selection method theoretically, this paper only selects it through a large number data from past experiments. The most suitable kernel function used for intrusion detection classification recognized in academia is radial basis function (RBF), so we use RBF as kernel function of SVM in this experiment.

RBF kernel function parameter (gamma, g) and the penalty coefficient (cost, c) are two other significant parameters of SVM model. This experiment applies famous K-fold cross-validation method to select these two parameters. Through divided training data into appropriate number of groups, train the first and second group then compares with the third group to get accuracy. And so on, until find a group with the highest accuracy then use script tests parameters g and c in the range of [c, g] = [2 ^ -10, 2 ^ 10] * [2 ^ -10, 2 ^ 10] one by one and select a group of g and c with highest accuracy. Finally, we obtain the best SVM model from existing training data.

The experiment utilizes original data set collected in previous step, through vectorization and normalization, conversed to data that SVM can identify then generate SVM model. Offline test results of model are shown in Table 2.

TABLE II. MODEL ACCURACY

Model name	Model 1	Model 2
Model size	147KB	33.4KB
the number of samples	6078	6078
Support Vector	1861	426
Gamma	0.125	0.0078
C	1	2048
Model training time	3.9	3.6
Misclassified	31	8
Detection time	Nanosecond class	Nanosecond class
Accuracy rate	99.49%	99.89%

V. EXPERIMENTAL RESULTS AND ANALYSIS

This experiment adopts real attack data and online data for testing and verification, different with other methods which applied offline data set to test.

In part of port scan, we run Slow Comprehensive Scan of Nmap and run maximum load attacks of UDP flood in part of DoS attacks. We can calculate rate of False-negatives and False-positives of rule base based SNORT (RULE-SNORT) and SVM based SNORT (SVM-SNORT) either through collected data, as shown in Table 3.

TABLE III. FALSE NEGATIVE RATE AND FALSE ALARM RATE

Experimental conditions	RULES	SVM-original model	K-fold cross SVM model
Port scan False-negative rate	99.92%	54.34%	57.60%
DoS attacks False-negative rate	99.95%	52.08%	55.67%
Port scan False-positive rate	29.02%	0%	0%
DoS attacks False-positive rate	51.15%	0%	0%

According to table above, rate of false-negative of SVM-SNORT is less than 5% when testing by Nmap, but in RULE-SNORT the rate is as high as 99%. According to DoS attacks, rate of false-negative in SVM-SNORT is about 52%, in RULE-SNORT the rate is up to 99%. Rate of false-positive in SVM-SNORT is almost zero, much lower than in RULE-SNORT. It is proved that without rules configuration, SNORT cannot respond to popular port scan tools and DoS attacks. However, because of its statistical characteristics and effective selected vectors, SVM algorithm can make accurate judgments to intrusion.

K-fold cross-validation method can not only optimize model but also test result of attribute selection. Stronger characteristics attribute have, smaller model after optimize. Nevertheless, if characteristics of attribute are so strong that one attribute have decisive impact to entire result, it is not suitable for application of SVM algorithm. Because advantage of SVM algorithm is high-dimensional pattern recognition, if dependent on only a few attributes mean dimension is low, expert system will be more suitable in this case. So amount of attributes is not the only factor, information contained in characteristics of attribute should also similar. Only in this way could SVM algorithm most suitable.

Experimental results show that after attributes and model optimization, SVM-SNORT can reduce rate of False-negative and False-positives significantly in system. Due to small model, the detect speed reach nanosecond class, that will not occupy operating system's resources.

VI. CONCLUSIONS

High rate of False-positives and False-negatives is the biggest problem faced by existing IDS. This experiment has successfully combined multiclass SVM with SNORT and applied online data and K-fold cross-validation to optimize attributes and model. According to experimental results, the rate of False-negative and False-positive of SVM-SNORT are far less than RULE-SNORT, proved that introduce statistical learning theory to intrusion detection system, apply SVM algorithm to replace the role of existing expert system in detect intrusions is feasible.

Next step in this study is to obtain more attacks samples, achieve more attacks classification. At the same time, continue to optimize attribute in vectors, extract more representative attributes and then training more accurate model. Because SVM algorithm used in this study is primitive, we will optimize SVM algorithm in the future, improve training speed and detection accuracy of SVM model to put more complete smart IDS into practice.

REFERENCES

[1] Vapnik V. The Nature of Statistical Learning Theory [M]. New York: Springer- Verlag, 1995

[2] LI Hui，GUAN Xiao-Hong，ZAN Xin，and HAN Chong-Zhao, Network Intrusion Detection Based on Support Vector Machine[J], JOURNAL OF COMPUTER RESEARCH AND DEVELOPMENT, no.40, 2003

[3] Xiao Yun, Wang Xuanhong. Support Vector Machine theory and its application in network security. Xi'an Electronic and Science University Publishing House, 2011

[4] LIU Xin-zhe, Application of artificial intelligence techniques to Intrusion Detection System [J]. Railway Computer Application, no.08, 2004

[5] RAO Xian, DONG Chun-xi , YANG Shao-quan, Detecting intrusions by using support vector machines [J] JOURNAL OF XIDIAN UNIVERSITY, no.03, 2003

[6] Guo Chi Chen Zhuo, Research of an Intrusion Detection Model Based on Support Vector Machine [J], Computer&Digital Engineering , no.09, 2010

Published by Atlantis Press, Paris, France.