Improved Support Vector Machine Wireless Network Security Detection Algorithm Model

Xiujian Lv[1, a], ZhiCheng Sun[2,b] and Jing Wang[3,c]

[1]Shijiazhuang Vocational and Technology Institute , Department of Information Engineering Shijiazhuang, China

[2] Hebei College of Industry and Technology ,Department of Computer Technology Shijiazhuang, China

[3] Shijiazhuang Engineering Vocational College, Department of Art & Design Shijiazhuang,China

[a]lvxji@126.com, [b]hbszc@126.com, [c]45076446@qq.com

Keywords: SVM; WLAN; information gain; Intrusion detection

Abstract. This paper proposed a support vector machine SVM algorithm for WLAN intrusion detection, first calculates the information gain of network intrusion data, and selects the characteristics of properties having a greater impact on the classification from the raw data to optimize the parameters of SVM, Finally, an optimized SVM algorithm is used to detect the wireless network data to obtain the behavior of network detection. Simulation results show that model of SVM-based WLAN intrusion detection has a high correct detecting rate, a low negative rate and wrong alarm rate.

Introduction

Because of the saleularity of the WLAN, the attacker without physical connection can carry on the attack, make WLAN security question seems particularly outstanding. Web 802.11 because wireless channel openness and 802.11 agreement own many loopholes, its security certification authorization mechanism exists great potential safety hazard, vulnerable to the effects of various network threat. For wireless encryption protocol attack platform emerge in endlessly, which is the famous Beini attack platform. Now computer operation and technology rapid development, single operation speed is very considerable, can reach 100 K kys/s, if use cloud computing, operation time will markedly shortened. In order to ensure the safety of the wireless local area network data transmission and communication, establish an effective WIDS to strengthen the security of WLAN becomes very important. Support vector machine is a kind of based on statistical learning theory of the new machine learning method, can well solve the high dimension and small sample problems such as study, generalization ability is very good, the SVM used in intrusion detection system, can obtain the better detection performance.

This paper analyzes the characteristics of WLAN, in research network intrusion detection technology was proposed based on the SVM based on optimization of access point mode wireless local-area network (WLAN) intrusion detection model structure, can have a malicious user attack and intrusion behavior for early detection.

The intrusion detection technology in the application of WLAN

The popularity of the WLAN is mainly it bring great convenience for users, but it is this convenience raises cable network security problems do not exist. The attacker without physical connection can be attached to the network, and uses relative set each steal to carry transmission of broadcast packets. Based on the network intrusion detection system NIDS analysis the original network packets as data source, to judge intrusion events, can not affect the performance of use case detection and response intrusion events.

Published by Atlantis Press, Paris, France.

At present, the W LAN intrusion detection are mostly under test, such as open source Snort intrusion detection system by Snort - wire - less beta, increased the Wife agreement field and option key word, the rule matching method for intrusion detection, the AP by administrator manual configuration, so can well identify unauthorized fake AP, in expansion when AP also need to reconfigure. However, due to its rules file without effective rule definition, make detection function is limited, and can not very well testing MAC address camouflage and flooding denial of service attack.

Based on the SVM wireless local-area network (WLAN) intrusion detection algorithm

Intrusion detection feature attribute selection.

Due to record some characteristic attributes and classification results has nothing to do, in the classification process analysis these attributes will increase time complexity, and reduce the anomaly intrusion detection rate. For more efficient classification and clustering, the use of information gain method, to select the great influence to the result of classification characteristic attributes. Specific steps are shown below:

(1) Calculation with 1 - N value of different mark field X entropy:

$$H(X) = -\sum_{i=1}^{N} P(x_i)\log p(x_i) \tag{1}$$

(2) Calculation mark field X for each feature attribute Y_j condition entropy:

$$H\left\langle X \middle| Y_j \right\rangle = -\sum_{i=1}^{N} P(y_{ik}) \sum_{i=1}^{N} P\left\langle x_i \,|\, y_{ik} \right\rangle \log\left\langle x_i \,|\, y_k \right\rangle \tag{2}$$

(3) Calculation X for each Y_j information gain value:

$$G\left\langle X \middle| Y_j \right\rangle = H\left\langle X \middle| Y_j \right\rangle - H(X) \tag{3}$$

Experiments adopt conservative estimate, set information gain of threshold value is 0.3, the select information gain more than 0.3 the characteristic attributes used in experiment, and in the light of L2 mark, a total of 12 selection feature attributes; And according to L5 mark, pick the 15 characteristic attributes.

Support vector machine intrusion detection algorithm.

SVM is based on structural risk minimization principle of machine learning methods, classifier and support vector only the number of related. That is the essence of its training according to the limited sample solve a quadratic programming problem, get the global optimal solution. The advantages of SVM classifier is generality is good, can improve the generalization performance, solve the nonlinear problem, and the classification accuracy is high, the classification speed and training sample number has nothing to do, will greatly improve the accuracy of WLAN in intrusion detection and fail.

A known a group of independent distribution of training sample:

$$(x_i, y_i), i = 1, ..., n, x \in R^d, y \in \{+1, -1\} \tag{4}$$

Use classification function: $f(x) = \operatorname{sgn}(w \bullet x + b)$ (5)

To normalization, SVM linear classification can be transformed into a quadratic regression problem:

$$\begin{aligned} &\min(\frac{1}{2}\|w\|^2 + C(\sum_{i=1}^{I}\xi_i)^p) \\ &s.t.\, y_i(w \cdot x_i + b) > 1 + \xi_i \\ &\xi_i > 0, i = 1, 2, ..., I \end{aligned} \tag{6}$$

C is the penalty factor, its value is said to experience the error of the greater the punishment.

By using Lagrange multiplier method, equation (6) into:

$$Max(\sum_{i=1}^{n} a_i - \frac{1}{2}\sum_{i,j=1}^{n} a_i a_j y_i y_j (x_i \cdot x_j)) \tag{7}$$

a_i, a_j are Lagrange multiplier.

This will convert to an inequality constrained quadratic function optimization problems. Finally SVM linear decision function as follows forms:

$$f(x) = \mathrm{sgn}\{(w \cdot x) + b\}$$

$$1 = \mathrm{sgn}\left\{\sum_{i=1}^{n} a_i y_i (x_i \cdot x\ ) + b\right\} \tag{8}$$

For nonlinear problem, through the kernel function: $K(x_i, x_j) = \Phi(x_i) \cdot \Phi(x_j)$ (9)

Substituting equation (8), the nonlinear SVM classification function is:

$$f(x) = \mathrm{sgn}(\sum_{i=1}^{I} a_i y_i k(x_i \cdot x_j) + b) \tag{10}$$

SVM classification of the big workload will be in input space and not in high dimensional feature space completed, and input data dimension has nothing to do, to avoid the algorithm can lead to "dimension disaster" problem, fast to solve quadratic programming problem, has the high accuracy of training, so as to identify normal behavior and intrusion behavior as well as the identification of different types of intrusion behavior.

The analysis of experimental results

Simulation experiment.

This paper using MITL laboratory in the simulation environment for network intrusion detection field comparative authoritative KDD99 data set, record 30 d within the original network packets, and translated into about 6 million records stored in the database MySQL5.1 filter and heavy, selected article 891558 valid data record.

KDD99 data set contains normal data and a variety of abnormal data, including four specific types of network intrusion attack behavior: denial of service attack (DoS); Unauthorized get super user permissions attack (U2R)); Remote unauthorized access attack (R2L); Leak detection and scanning attack (PROBING). Each record including: connection time, port, source address, destination address, etc. Each record have been marked as normal or above one of the four types.

In the experiment, joined the two new marker to replace the original mark. (1) binary mark (L2) : if a record is abnormal behavior, mark for 1; If a record is normal behavior, mark to 1. (2) the abnormal behavior type mark (L5), according to abnormal behavior type mark data set. Using LIBSVM function library construction of intrusion detection SVM classifier, and respectively in L2 mark and join L5 mark data sets experiment. Experimental data selection original 40 feature attributes and reduced after the 15 feature attribute, stop the threshold value is 0.02. Will join the L5 mark data set according to abnormal behavior types are divided into 4 sub-niches test data set, each data set contains a kind of abnormal behavior and all normal data. The results are as follows table 1 and table 2 shows, including VN represent the number of support vector.

SVM classification algorithm makes the WLAN intrusion detection strategy flexibility, signature extraction more accurate, fast, can be accurate classification statistics and improve the matching efficiency.

Table 1. Against L2 SVM classification results

	VN	Classification rate%
40 feature attribute	623	83.61
15 feature attribute	574	83.96

Table 2. Against L5 SVM classification results

Abnormal behavior type	40 characteristic attributes.		12 characteristic attributes.	
	VN	Classification rate/%	VN	Classification rate/%
Dos	623	83.61	574	83.96
R2L	180	83.77	150	81.89
U2R	152	98.62	142	98.56
Probing	510	91.56	633	91.97

The results analysis.

This paper introduces a SVM algorithm based on wireless local-area network (WLAN) wireless network intrusion detection algorithm, the four types of attacks type testing. Intrusion detection algorithm for Probing, DoS, U2R, R2L this four data types of detection accuracy is higher, and misinformation rate to maintain a low level, has the high experimental value.

Conclusions

Based on WLAN network intrusion detection system research and implementation has just started, fail and misinformation rate are high, but also the existence of intrusion detection data exist dimension, redundancy higher defect. This paper studies the SVM based on wireless local-area network (WLAN) the wireless local area network intrusion detection problems, the experimental results proved that the proposed WLAN anomaly intrusion detection method is feasible, and has good invasion of recognition accuracy and detection rate, at the same time maintain low error rate, has higher reference significance.

References

[1] Park J H. Synchronization of Genesio Chaotic System via Back- stepping Approach[J]. Chaos Solitons Fractals, 2006, 27(5): 1369- 1375.

[2] Yan Junjun, Yang Yi-Sung, Chiang Tsung-Ying, et al. Robust Synchronization of Unified Chaotic Systems via Sliding Mode Control[J]. Chaos, Solitons and Fractals, 2007, 34(3): 947-954.

[3] Julisch, K. Mining Alarm Clusters to Improve Alarm Handling Efficiency. in: Proceedings 17th Annual Computer Security Applications Conference. New Orleans, LA, USA: IEEE Comput. Soc, 2001. 12.

[4] Goodall, J. R., Lutters, W. G., Komlodi, A. The Work of Intrusion Detection:Rethinking the Role of Security Analyst. in: Proceedings of the Tenth Americas Conference on Information Systems. New York: 2004.

[5] Ammann, P., Wijesekera, D., Kaushik, S. Scalable, Graph-based Network Vulnerability Analysis. in: Proceedings of 9th ACM Conference on Computer and Communications Security (CCS 2002): 2002.

[6]Yanmei Li, Jingmin Wang. Risk Assessment Model of Smart Grid Project Based on Variable Pr ecision RS and LSSVM. Advances in Information Sciences and Service Sciences (AISS). Vol.3, No .10, 2011,11 :375～383.

Published by Atlantis Press, Paris, France.