A Detection Method of Network Intrusion Based on SVM and Ant Colony Algorithm

Jianfeng Pu
Aviation University of Air Force
Changchun Jinlin China
pujianfeng2002@163.com

Lizhi Xiao
Aviation University of Air Force
Changchun Jinlin China

Yanzhi Li
Aviation University of Air Force
Changchun Jinlin China
yzhi_li@163.com

Xingwen Dong
Aviation University of Air Force
Changchun Jinlin China

***Abstract*—This paper researches the intrusion detection problem of the network defense, pointing at the problem of low fitting defects in the traditional detection algorithm of high precision and low forecasting accuracy under the situation of small sample training, and puts forward the algorithm of Support Vector Machine. Aimed at the important influence of SVM kernel function on classification performance, this paper adopts the improved Ant Colony Algorithm as the method of selection SVM characteristics parameters. Experiments show this algorithm is significantly higher than the other algorithm in training and the detection speed, and have a high enhance of the detection rates of attacking sample.**

Keywords-Network Defense; Intrusion Detection; Support Vector Machine (SVM); Ant Colony Algorithm

I. INTRODUCTION

Along with the development of computer and network technology, the network countermeasure is getting more and more attention as a new information countermeasure model. The network countermeasure includes network attack, network defense and computer virus weapons. As an important content of network defense, Intrusion Detection always attracts wide attention by scholars at home and abroad.

Since Dennings put forward the intrusion detection model in 1980s, people have put forward many detection methods, such as: hidden markov model, genetic algorithm, the artificial immune system and neural networks. These intrusion detection methods, have a high detection rate, but their distorting rate is high as well, Hence the performance of these methods has close relations with the size of training data quantity, If the training samples were in low, the system performance would have significantly been reduced. However, it's difficult to gain a large amount of data in the abnormal network intrusion detection, the actual original data source has the properties of variability, different quality, high dimension, small sample and so on. If the traditional detection method is applied here, there will be some defects, such as high precision and low forecasting accuracy, and unsatisfactory actual detection effect.

Support Vector Machine is a new kind of machine learning method based on the principle of structural risk minimization, dedicated to solve the practical problem of small sample, nonlinear, high dimension and local minimum points, by means of the method of generalization ability, improving the defects of slow convergence speed and easy to fall into the local extremum shortcomings exited in the traditional algorithm. In view of the current existed fitting of low detection probability, and high rate of false alarm, the article puts forward a kind of abnormal network intrusion detection method based on Support Vector Machine, and puts forward the thoughts of using ant colony algorithm to select the kernel function, pointing at the problem that the kernel function is difficult to determine. Finally, the paper use the KDD99 invasion data set verified the method by experiment, the experimental results show that the method proposed in this paper has higher rate of intrusion detection accuracy to detect the abnormal network intrusion.

II. NETWORK INTRUSION DETECTION

A. *The principle of network intrusion detection*

Network intrusion is the behavior attempting to break the computer resource integrity, usability and confidentiality. It can carry out intrusion activities not only from outside, but also from the unauthorized activity of internal user. Through collecting some key points of the information in computer network or computer system and making analysis, Intrusion Detection tries to find whether there exist the signs of violating the security strategy and the behavior of being attacked in the computer network or computer system. Intrusion Detection System (IDS) includes the network Intrusion Detection's hardware and software.

B. *The mathematical model of network intrusion detection*

The network data signal can be indicated by the following mathematical expression:

$$\begin{cases} H_0 : R(t) = C(t) + n(t) \\ H_1 : R(t) = S(t) + C(t) + n(t) \end{cases} \quad (1)$$

Among which there is no invasion signal under the hypothesis of case H_0, there is invasion signal under the hypothesis of case H_1. $R(t)$ means the signals received from network flow, $C(t)$ means all the total bytes' length of network traffic data packets during the interval time of sampling , $S(t)$ means the invasion signal, $n(t)$ means the noise of the network data.

The goal of network intrusion detection is to test the invasion signal in the background of the network data with noise, and hope to have the smallest false alarm rate. Figure 1, it is known that in the network intrusion detection system, network intrusion detection model is the core part of network detection system. Because the network data in the network intrusion is a small sample comparing to the total, it usually builds the network intrusion detection model based on SVM.

III. SVM

Based on the structural risk minimizes, the SVM algorithm is the new classification method proposed by the Vapnik, it develop on the theory of statistics, have solid theoretical foundations. The basic thought of SVM is first input the sample and through the kernel function map to the higher dimensional eigenspace, then looking for the optimum boundary in the eigenspace through the maximizing classification interval, the classification interval is maximized and can be transformed into quadratic programming problem, take advantage of Lagrange multiplier and solving antithesis questions of the quadratic programming, thus get the solving of the question.

Does not need any priori knowledge, the SVM can find the optimum hyper plane, and only need a small amount of support vector quantities to show this optimum hyper plane. SVM has overcome the deficiency of the neural network. In solving the small sample, non-linear higher dimensional pattern-recognition, it has a lot of characterized advantages such as simple structure, global optimum and good generalization ability. And it can be popularized and applied in machine learning problems such as function-fitting.

A. Nomal SVM

The optimum classification hyper plane used for target identification of the SVM is generated through maximization the 2 kind targets perigee, consider the target classification issue, the training sample is

$$(y_1, X_1), \cdots, (y_2, X_1), X \in R^n, \quad (2)$$

Where X is the n dimensional characteristic vector; y is category identifier; y∈±1; l is the number of the training sample.The 2 kind targets identifier can be showed as $\mathrm{sgn}(f(x))$, where sgn( ) is sign function, that is:

$$\mathrm{sgn}(x) = \begin{cases} 1, x > 0 \\ -1, x < 0 \end{cases} \quad (3)$$

The final goal of the identifier is to find a $f(x)$, which make all of the training sample satisfy $y_i \,\mathrm{sgn}(f(x_i)) > 0$. When the training sample is nonlinearity, the SVM input the training sample and through the kernel function map the input vector (ϕ) to the higher dimensional eigenspace, the optimization problem for the identification hyper plane is

$$\begin{aligned} &\min_{\alpha} \frac{1}{2}\sum_{i,j=1}^{1} \alpha_i y_i \alpha_j y_j K(X_i, X_j) - \sum_{i=1}^{1} \alpha_i \\ &s.t. \sum_{i=1}^{1} \alpha_i y_i = 0, \\ &0 \le \alpha_i \le C, i = 1, 2, \cdots, 1 \end{aligned} \quad (4)$$

The $K(X_i, X_j)$ is the kernel function stratified the Merce condition. Usually the kernel function include the polynomial function and radial direction base function

$$K(X_i, X_j) = \exp - \frac{\|X_i - X_j\|^2}{2\sigma^2}$$

Where d is the degree of a polynomial, σ is constant. The decision function and b are:

$$f(x) = \mathrm{sgn} \sum_{i=1}^{1} \alpha_i y_i \, K(X_i, X) + b \quad (5)$$

$$b = \frac{1}{N^*} \sum_{X_i \in \Omega_N} y_i - \sum_{X_j \in \Omega_J} \alpha_j y_j (X_j, X) \quad (6)$$

Where Ω_N is the standard SVM sets, Ω_J is SVM sets.

The dual problem is

$$\begin{aligned} &\min_{\alpha} \frac{1}{2}\sum_{i,j=1}^{1} \alpha_i y_i \alpha_j y_j K(X_i, X_j) - \sum_{i=1}^{1} \alpha_i \\ &s.t. \sum_{i=1}^{1} \alpha_i y_i = 0, \\ &0 \le \alpha_i \le \omega_i C, \qquad i = 1, 2, \cdots, 1, \end{aligned} \quad (7)$$

B. Parameters of the SVM

The parameters value of the SVM has great influence to the learning and introduction ability, so it is one important research content. For the SVM taking the RBF as the kernel function, the parameters include adjusting parameter C, the width of the kernel σ and the insensitive coefficient ε. Generally, the value of the C, σ and ε is correlated to the learning sample and issues, C need a trade-off decision between the structure risk and sample error, the greater C permits smaller error, and the smaller C permits greater error. The value of the width of the kernel σ is correlated to the input space range and width of the learning sample, the greater input space range, the great σ value. The insensitive coefficient ε is correlated to the noise, its value has proportional relation with the noise level. In practice, the parameters value should be set according to noise level.

This paper adopt the Ant Colony Optimization to get parameters values of SVM, which can convergence the SVM parameters and get the optimal value, so has better effect.

IV. ANT COLONY OPTIMIZATION

Scientific research shows, the ant colony in the natural world has an intellective character—ants can release a chemical substance called pheromone, they can carry food back to their nest in the shortest route without any visual aid. M. Dorigo, an Italian scientist proposed the "Ant system" method based on such character of ants, and used this method solved problems like TSP (Traveling Salesman Problem), which received great lab results. Further on, M. Dorigo named all ant colony algorithms as Ant Colony Optimization (ACO) in general, which proposed an unique framework model. This algorithm has not only great robustness, positive feedback characteristic and also with parallel and distributed computing feature.

A. *Algorithm structure*

Suppose ant colony scale as *N*, randomly distribute the ant colonies in solution space, then according to the initialized position of the ants' distribution, follow the difference of optimization problem to confirm the initialized pheromone size of ant f:

$$\Delta\tau(i) = \exp(-f^{new}(X_i)) \tag{9}$$

For the initialized position of ants

$$X_i(x_{i1}, x_{i2}, \cdots x_{id}), \qquad i = 1, 2\cdots, N$$

When $f(X_i) \geq 0$, from (9) we know $\Delta\tau(i) \in (0,1]$, when $f(X_i)$ is infinity pheromone thickness will infinitely close to zero, hence, should adjust the adaptability of $f(X_i)$;

$$f^{new}(X_i) = \begin{cases} f(X_i)/a_{vg}, if \quad a_{vg} > a_{vg_0} \\ f(X_i), \qquad otherwise \end{cases} \tag{10}$$

In which, a_{vg} is the average value of $f(X_i)$, $a_{vg_0} \in [1.0\times10^3, +\infty)$; $f(X_i)$ is the adaptability value before adjusted; $f^{new}(X_i)$ is the value after.

After one searching round, ants will make the next searching round by the movement experience accordingly. This article algorithm's movement regulation contains two parts: one is to select individual target through dynamic random, move the other ants to individual target except the optimal ants from the last iteration, we call it as overall long step search; the other refers from the "detective" theory in pattern search, let the optimal ants have short step partial elaborate search in the neighborhood, in order to find the optimal solution.

Firstly, randomly select individual *P* form the sample colonies. In which: $p = [\mathrm{r}\cdot\mathrm{N}]$, *N* is the sample colonies scale, r is the dynamic movement extraction rate, $r = (i_{ter\max} + i_{ter})/2i_{ter\max}$, $i_{ter\max}$ is the maximum iteration times, i_{ter} is the current iteration times. Then calculate that the maximum pheromone thickness from we have chosen colonies as target individual X_{obj}.

$$X_{obj} = \begin{cases} X_j, \qquad if \quad \tau(X_i) < \max(\tau(X_j)) \\ \qquad j = 1, 2\cdots, p \\ X_{best,} \qquad\qquad otherwise \end{cases} \tag{11}$$

In which: X_{best} is the best solution in last iteration. It is due to the bigger pheromone thickness depended on individual, the more attractive this individual to the other ants. It is possible to find better solution in the process when ants make movement to individual ants. Ant *i* according to (12) move to target ant position.

$$X_i = (1-\lambda)X_i + \lambda X_{obj}, \lambda \in (0,1) \tag{12}$$

When ants follow this movement rule, it causes to increase randomness at the beginning stage, accelerates convergence race in the later stage.

The ants which get the optimal solution in above iteration X_{best} take partial elaborate search in their neighborhood, search algorithm as:

Equation (13) and (14)

$$X_{best} = \begin{cases} X_i', \ if \quad f(X_i') < f() \\ X_{best}, \qquad otherwise \end{cases} \tag{13}$$

$$X_i' = X_{best} \pm h\cdot\delta \tag{14}$$

h is the dynamic search step length, the purpose making the dynamic searching step length is to make the process more and more precision during the searching, will help to improve the accuracy.

After finishing overall search and partial search, update pheromone $\tau(i)$ in position ant I, the updated rule is:

$\tau(i) = (1-\rho)\tau(i) + \Delta\tau(i)$ .In which: ρ is the volatility coefficient $\rho \in (0,1)$.

B. *Base on the ant colony SVM optimize parameter*

Define target function as:

$$\begin{aligned} &\min \quad f(C,\sigma,\varepsilon) = \mathrm{sgn}(a_1C, a_2\sigma, a_3\varepsilon) \\ &s.t. \quad C \in [C_{\min}, C_{\max}] \\ &\qquad \sigma \in [\sigma_{\min}, \sigma_{\max}] \\ &\qquad \varepsilon \in [\varepsilon_{\min}, \varepsilon_{\max}] \end{aligned} \tag{15}$$

The concept of optimize parameter is to get a set of parameter (C,σ,ε) through iteration algorithm, to minimize the target function (15). This article makes the optimization to described ant colony algorithm. Make a set of parameter (C,σ,ε) in the domain as the ant's position vector, every ant's adaptness function apply equation (15). The detailed steps as:

Stepl. According to gathered data, constructs training sample set and test sample set.

Step2. Set up parameters, initializes the ant s' position, every position corresponding a set of parameter (C,σ,ε) in SVM model, builds up SVM prediction model by parameters and samples.

Step3. From (10) calculate every ant's adaptness value, and then by (11) calculate every ant's pheromone thickness.

Step4. Randomly select P ants from ant colony, find out the optimal ant's position X_{best} based on the pheromone thickness of every ant's position. Set it up as individual target X_{obj}.

Step5. The non-optimal ants in the ant colony moving to target ant's position by (12) make the overall search.

Step6. The optimal ant make overall search according to its neighborhood.

Step7. Update every ant's pheromone thickness.

Step8. Confirm if satisfies its iteration termination condition, if so, then iteration terminates and outputs the optimal parameter (C,σ,ε); otherwise, return step4.

Step9. Apply the optimal parameter (C,σ,ε) and training sample to build up SVM prediction model.

V. THE ANALYSIS OF THE EXPERIMENTAL RESULTS

A. *Experimental platform and data set introduce.*

KDDCUP99 is formed from the IDS data set of MIT LL in 1998, which only contains network traffic data

- All data set: the training data set (kddcup.data.gz) ,the test data set. (kddcup.testdata.unlabeled.gz)
- 10 percect data set: the training data set (kddcup.data. 10 percent. Gz) , the test data set. (kddcup.newtestdata.unlabeled 10 .percent .gz)
- Corrected.Gz is a test data set contain attacking mark, researchers can compare and analyze the results of their own algorithm test by using this data set.

This paper takes the 10 percent data set to analysis object. In the 10 percent data set offered by KDDCUP99 (including training data set and test data set), it contains 4 kinds of network attack types, which has obvious features in the MIT LL intrusion detection data: the training data set contains 23 kinds of aggressive behavior; test data set contains 38 kinds of aggressive behavior, the type of data sample and distribution is shown in TABLE Ⅰ.

TABLE I. KDDCUP99 DATA DISTRIBUTION

Type	10% Rrain Set		10% Test Set	
	Data	*Rate(%)*	*Data*	*Rate(%)*
Nomal	97 278	19.69	60 593	19.48
Probe	4 107	0.83	4166	1.34
DOS	391 458	79.24	229 853	73.90
U2R	52	0.01	228	0.07
R2L	1 126	0.23	16 189	5.20

B. *The comparative analysis of the algorithm*

Relate to neural network and traditional SVM algorithm, the ACO-SVM algorithm which improves the data sample of network intrusion detection shows a higher accuracy. the Contrasted results that several algorithms apply in training and test in the previous section selected shown in the TABLE Ⅱ.

TABLE II. THE EXPERIMENT'S RESULT

Type	*BP*	*N-SVM*	*ACO-SVM*
Categorised precision（%）	87.3	90.5	99.2
Training time（S）	13.4	12.4	5.6
Testing time（S）	0.93	0.86	0.79
Result tested	ordinary	good	better

The improved ACO-SVM makes separating hyperplane ant colony quickly find the right radial basis kernel function, makes the algorithm substantially higher than the pervious 2 kinds of methods for anomaly detection rate. while the training sample distribution law tends to test sample, the anomaly detection rate reached 99.2%

VI. CONCLUSION

This paper proposes a intrusion detection method based on improved ant colony algorithm which the SVM, the simulation results show that the method is quick convergent, less iteration, and can improve the accuracy of intrusion detection system to a certain extent. The above analysis shows that the ant colony algorithm is rapid, simple and precious in test, relative to the neural network algorithm and the traditional support vector machine algorithm in BP network study.

REFERENCES

[1] WANG Ze-long, YAN, Feng—xia and HE Feng, "A new SVM multi-class c1assification method based on error-correcting code[C] ", Suzhou : 2008 InternationalConference On Computational Intelligence and Security, 2008.

[2] Gjorgji Madzarov, Dejan Gjorgjevikj and Ivan Chorbev, "A multi-class SVM classifier utilizing binary decision tree[J] . Informatica , 2009(33) : 233—241.

[3] Song Guangjun, Zhang Jialin and Sun Zhenlong, "The research of dynamic change learning rate strategy in bp neural network and application in network intrusion detection[C]", 3rd International Conference on Innovative Computing Information and Control, Dalian,Liaoning:IEEE Press,2008:513-513.

[4] Wang Huiran and Ma Ruifang, "Optimization of neural networks for network intrusion detection[C]" , First International Workshop on Education Technology and Computer Science. Wuhan,Hubei, China:IEEE Press, 2009:418-420.

[5] Dorigo M.Optimization, "learning and natural algorithms[D]" , Italy:Ph.D.Thesis,Department of Electronics,Politecnico di Milano,1992.

[6] Dorigo M, Maniezzo V and Colomi A, "The ant system:optimization by a colony of cooperating agents[J]", IEEE Transactions on Systems, Man, and Cybernetics:Part B,1996,26(1):29-41.

[7] LIU Yi-guang, YOU Zhi-sheng and CAO Li-ping, "A novel and quick SVM—based multi-class classifier[J]", Pattern Recognition, 2006, 39(11) : 2258—2264.

[8] S C Lee and D V Heinbuch, "Training a neural－network based intru-siondetector to recognize novel attacks［J］", IEEE trans, on sys-tems, man, and cybernetics, 2001.294－299.

[9] K M C Tan, K S Killourhy and R A Maxion, "Undermining an anomaly based intrusion detection system using common exploits［J］", Raid, 2002. 16－18.