Detecting DDoS attack based on PSO Clustering algorithm

Xiaohong Hao[1,a]，Boyu Meng[1,b]，Kaicheng Gu[1,c]

[1]School of Computer & Communication, LanZhou University of Technology, Lanzhou 730050

[a]316475958@qq.com；[b]boyu8816@163.com；[c]gkc1314@qq.com

Keyword: application-tier Distributed Denial of Service; browse behavior; particle clustering algorithm; anomaly detection.

Abstract. First, this article analyzes the Application layer Distributed Denial of Service(DDoS)'s attack principle and characteristic. According to the difference between normal users' browsing patterns and abnormal ones, user sessions are extracted from the web logs of normal users and similarities between different sessions are calculated .Because traditional K-mean Clustering algorithm is easy to fail into local optimal, the Particle Swarm Optimization K-mean Clustering algorithm is used to generate a detecting model. This model can been used to detect whether the undetermined sessions are DDoS attacks or not. The experiment show that this method can detect attacks effectively and have a good performance in adaptability.

Introduction

Distributed denial of service attacks is one of the major threats to the security of the Internet, which in the absence of any warning consume resources of the target,it can be made at the network layer or application layer[1].Application layer DDoS have two attack methods[2]: bandwidth depletion mode and the host resource depletion mode. At present, methods to solve these similar problem including: Intrusion detection technology based on data packet[3]Detection method based on flow limitation[4], Detection method based on frequency of access[5], Detection method based on Hidden semi-Markov model[6], Detection method based on the analysis of user behavior data mining[7].The literature[8] proposes a new Dos detection based on data mining, which combined Apriori algorithm and k-mean clustering algorithm. It usingnetwork data to detect DDoS, so it cannot cope with the application layer DDos. The k-mean algorithm have itself flawed, it overly need to select the fit cluster centers and for some initial value, it may converge to sub-optimal solution.

Application layer DDoS detection based on PSO clustering algorithm

Principle and model of detection: This paper establish detection model which is using to identify the application layer DDoS form analysis user behavior. System design as shown in Figure1.

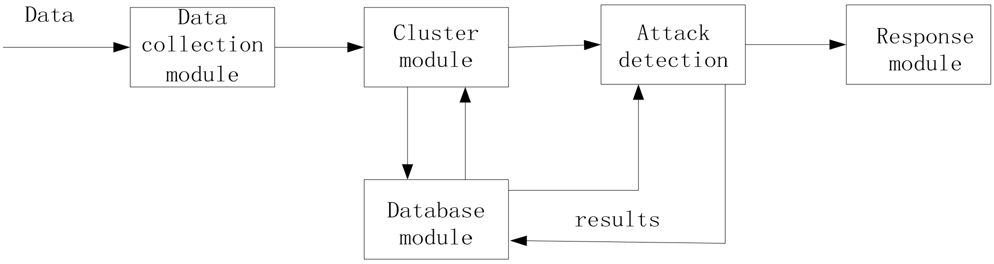

Figure1.system module design

Description of user browsing behavior

The Web log records information about each user access to the server, it including the user's IP address, client, customer identification, time of Web server receives the request, customer requests, request status code, transmitted bytes such as some access data. Extract Web log , preprocess the information and translate the results into Session:

$$S_k=\{ip_k,<t_1,\mathrm{u}_1>,<t_2,\mathrm{u}_2>,\cdots,<t_i,\mathrm{u}_i>\} \qquad (1)$$

Calculate the distance between sessions

In order to more accurately describe the user browsing behavior, better reflects the normal legitimate users and anomaly attacks users browse access to the difference in behavior, so analysis the similarities and differences in content, time, page-views and sequence. This paper refer to the method which use three vectors and a matrix to detailed descript the user's session features.Then calculate the similarity between session δ , the more similarity the distance more small. So the abstract distance can be defined as $\mathrm{d}=\frac{1}{\varphi}$.

Definition 1 (content vector): $W_k=(\mathrm{w}_1,\mathrm{w}_2,\cdots,\mathrm{w}_n)$, length of the vector is n. It indicates the server contains page number. The formula is as follows:

$$\Delta(\mathrm{W}_p,\mathrm{W}_q)=\frac{\sum_{\forall i\in[1,\mathrm{n}]}\partial(\mathrm{W}_p{}^{(i)},\mathrm{W}_q{}^{(i)})}{n} \qquad (2)$$

Definition 2 (time vector): $T_k=(\mathrm{t}_1,\mathrm{t}_2,\cdots,\mathrm{t}_n)$ 1, length of the vector is n. It of user browsing page i.The similarity formula of two hit vectors is as follows:

$$\Delta(\mathrm{T}_p,\mathrm{T}_q)=1-\mathrm{d}(\mathrm{T}_p,\mathrm{T}_q) \qquad (3)$$

Definition 3 (hit vector): $Hit_k=(\mathrm{hit}_1,\mathrm{hit}_2,\cdots,\mathrm{hit}_n)$, length of the vector is n.It indicates times number of a user brows a page, it reflects the user's interest degree each pages.

$$\Delta(\mathrm{Hit}_p,\mathrm{Hit}_q)=1-\mathrm{d}(\mathrm{Hit}_p,\mathrm{Hit}_q) \qquad (4)$$

Definition 4 (sequence matrix): H_k is a $n\times n$ matrix, it records the number of times of jumping between the various pages in the session. The similarity formula of two time vectors is as follows:

$$\Delta(\mathrm{H}_p,\mathrm{H}_q)=\frac{\sum_{\forall i\in(1,\mathrm{n})}\sum_{\forall j\in(1,\mathrm{n})}\partial(\mathrm{H}_p^{(\mathrm{i,j})},\mathrm{H}_q^{(\mathrm{i,j})})}{n^2} \qquad (5)$$

Considering the similarity between three vector and a matrix, the overall similarity $\delta(\mathrm{S}_p,\mathrm{S}_q)$, is as follows:

$$\delta(\mathrm{S}_p,\mathrm{S}_q)=\frac{\Delta(\mathrm{W}_p,\mathrm{W}_q)+\Delta(\mathrm{T}_p,\mathrm{T}_q)+\Delta(\mathrm{Hit}_p,\mathrm{Hit}_q)+\Delta(\mathrm{H}_p,\mathrm{H}_q)}{4} \qquad (6)$$

Numerically greater, the session are more similar, the distance between there sessions is smaller. So the distance is as follows: The formula is as follow

$$\mathrm{d}(\mathrm{S}_p,\mathrm{S}_q)=\frac{1}{\delta(\mathrm{S}_p,\mathrm{S}_q)} \qquad (7)$$

Detection of attacks

The Sessions is defined as $S=\{\mathrm{S}_{i,}\mathrm{i}=1,2\cdots,\mathrm{N}\}$, $\mathrm{S}_{i,}$ is a N-dimensional pattern vector. The solution is to divide $\omega=\{\omega_1,\omega_2,\cdots,\omega_M\}$ 1, let the total dispersion of the all clusters to be minimum.

The total distance of all samples to the corresponding cluster's centers is minimum. The formula is as follow:

$$J=\sum_{j=1}^{M}\sum_{X_i\in\omega_j} d(\mathrm{S}_i,\overline{S^{(\omega_j)}}) \qquad (8)$$

$\overline{\mathrm{S}^{(\omega_j)}}$ is the cluster's center j， $d(\mathrm{S}_i,\overline{\mathrm{S}^{(\omega_j)}})$ is the distance between the sample and the cluster's center j.

PSO Clustering algorithm

This paper consider the cluster's center as a particle's corresponded solution, the particle's location is combined with cluster's center. There are two forms of application layer DDoS attacks and normal user, so the number of clusters is M=3.Algorithm flow chart is as follows：

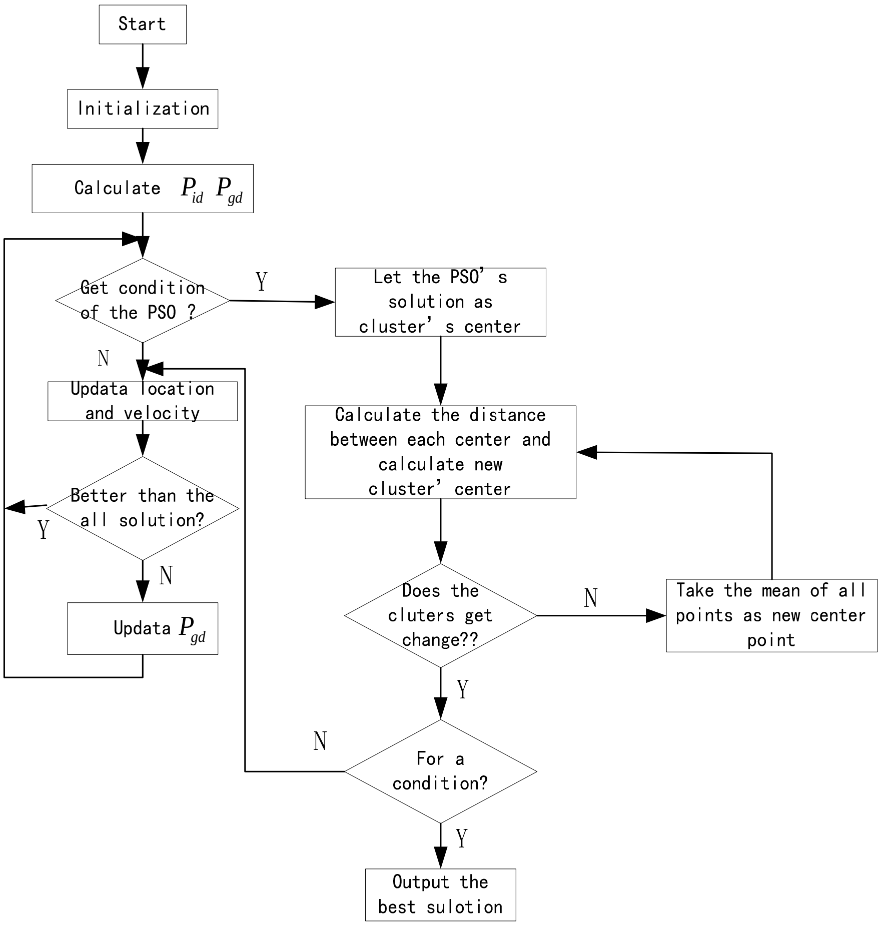
Start
Initialization
Calculate P_{id} P_{gd}
Get condition of the PSO ?
Y
N
Let the PSO' s solution as cluster' s center
Updata location and velocity
Calculate the distance between each center and calculate new cluster' center
Better than the all solution?
Y
N
Updata P_{gd}
Does the cluters get change??
N
Take the mean of all points as new center point
Y
N
For a condition?
Y
Output the best sulotion

Figure 2. Flow chart PSO clustering algorithm

Experimental results and analysis

This paper use the data<VAST Challenge 2013-mini challenge 3, Network security log data>from Central South University's visual research group. TFor the large amounts of the data , the paper randomly collect 100 sample and 20 attack sample data from the Web log of user a access logs. Program development platform is MATLAB2014a.The cluster analysis results in Figure 3.

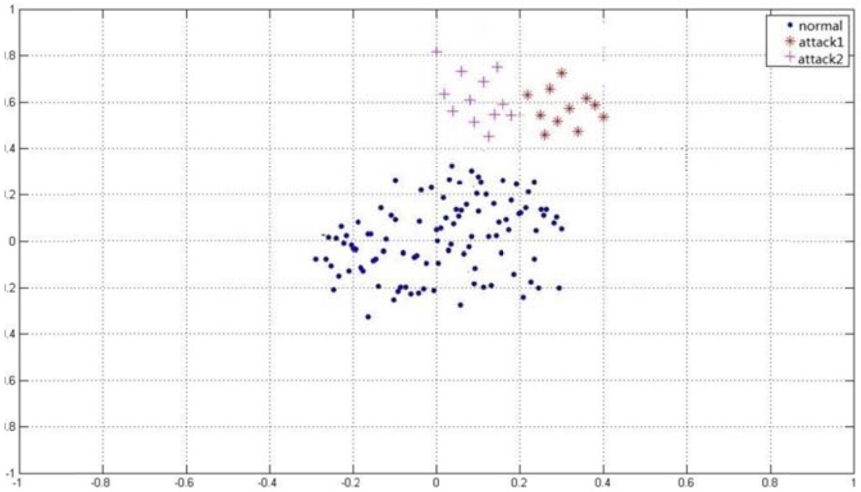

Figure 3. Clustering results of Euclidean space projection

Table 1 Clustering results

Data	Session	Actual attack Session	Detecting attack Session	Accuracy rate
S_k	120	20	23	86%

It can be seen that model detection rate is about 86% from the Table 1. The reason of detects attacks number slightly more than number of actual attacks is the model can not reflect all normal access to legitimate user's behavior. If increase the amount of the data, after corresponding clustering analysis, the accuracy will be increased accordingly.

Conclusion

This paper analysis the principles and characteristics of application layer DDoS attacks , provide a application layer DDoS attacks detection method which is based on Particle Swarm Clustering algorithm and describe user's behavior of browsing Web pages. Consider the attacks as an user's exception access behavior, according to the difference of legitimate and abnormal user's browsing behavior, describe the user's browsing behavior by data mining technique, calculate the similarity between each session, then detect the attacks behavior by using Particle Swarm Clustering algorithm. Simulation experiments show that this method can detect attacks effectively and have a good performance in adaptability.

References

[1] FenYan, JiajiaWang,Jinfeng Zhao.DDoS attack detection summary[J].Study on computer application, 2008,25（4）：966-969.

[2] ChuanXu. Research and implementation of DDoS attack detection algorithms on application layer [D]. Chongqing University，2012.

[3] Dougligeris C, Mitrokotsa A. DDoS attacks and defences machanisms: Classification and st ate-of-art[J], Computer Network,2004, (44):643-666.

[4] Sun Chang-hua, Liu Bin. Survey on New Solution Against Distributed Denial of Service Att acks[J]. ACTE Electronica SINCA. 2009, 7(37):1562-1570.

[5] Muthuprasanna M, Manimaran G. Distributed Based on Web User's Browsing Behaviours[J]. Journal of Software. 2007,4(18):967-977

[6] Yi Xie. Research on key technology of HTTP attack detection on application - layer [D]. Guangdong : Sun Yat - sen University, 2008

[7] Fengyu Wang, Shoufeng Cao, Jun Xiao. A DDoS detection method of community outreach based on Web application layer [J]. Journal of software, 2013,24 (6) : 1263-1273.

[8] NengGao, DengguoFeng,. A DOS attack detection based on data mining technology [J]. Chinese Journal of Computers, 2006,29 (6) : 944-950