ISSN(Online) : 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 2, February 2017

Preventing DDoS Attack Using Fuzzy C Mean Clustering

M.Durga Devi [1], V.Gokula Priya [2] , S.Sarumathy [3], Ms .R. Sujatha [4]

Department of Computer Science & Engineering, Sri Krishna College of Technology, Kovaipudur, Coimbatore, India[1,2,3]

Associate Professor, Department of Computer Science & Engineering, Sri Krishna College of Technology,Kovaipudur, Coimbatore, India[4]

ABSTRACT: Distributed denial of service attack(DDoS) brings a very serious threat to send to the stability of the Internet.This paper analyzes the characteristic of the DDoS attack and recently DDoS attack detection method. Presents a DDoS attack detection model based on data mining algorithm. FCM cluster algorithm and Apriori association algorithm used to extracts network traffic model and network packet protocol status model. The threshold is set for detection model. Experimental result shows that DDoS attacks can be detected efficiently and swiftly.

KEYWORDS: DDoS, cluster algorithm, association algorithm, FCM, Apriori

I. INTRODUCTION

Distributed denial of service(DDoS)[1] attacks make the resources of host occupied largely via sending many malicious packets, which results in the failure of normal network services. DDoS attack the target host through construting a lot of illegal packets, this kind of attacks changed traditional peer to peer attack mode and used distributed attack mode instead that causes the extent of hosts participating in attack wider, data flow generated by attack present irregular status. All of this make DDoS attacks launched easily, prevented and tracked difficultly and so forth. So far DDoS attacks have become one of the essential threats to network security.

In this paper cluster algorithm and association algorithm are used to build the traffic threshold model and packet protocol status model, so as to automatic, real-time, effective detection of DDoS attacks.

II. ESTABLISHMENT OF DDOS DETECTION MODEL

A. DDOS ATTACK PROCEDURE

A DDoS attack launched by the attacker includes mainly three steps, that is, searching the attack target, attacking and occupying the zombie and actual attacks. The specific process is as follows:

1. Before attacking, the attacker firstly searches the hosts in the network with security vulnerabilities from which the hosts with good link state and performance are picked out, and then intrudes these hosts so that corresponding administration authority is achieved to install control programs.

2. The attacker through network gives the handlers of the attack control instructions that cause the handlers give orders to the agents. Generally the attack agents are controlled by more than one handlers.

ISSN(Online) : 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 2, February 2017

3. The agents send the victims a large quantity of packets. It is difficult to distinguish between malicious requests and normal connection requests because these packets are masqueraded and could not be recognized where they are from as well as the protocols used by attackers are very common.

B. THE CHARACTERISTICS OF DDOS ATTACK

The characteristics of DDoS Attack are as follows after the analysis of it:

1. Abnormal traffic. A lot of useless packets transmitted by the attacker in order to occupy the resources of the victims(bandwidth or host resources). Such a large number of packets would cause the victims system-halted and fail to provide external services.
2. Most DDoS attacks take the three times handshake mechanism and use "SYN" status flag to send the victim connection requests . However, this does not mean to build a real connection, which makes the victim maintain a great deal of half-opened connection and consume the resources of the victims.
3. The attacker makes use of one of the characters of TCP/IP protocol that some non-compliant packets could be used so as to launch DDoS attacks.

Among the characteristics above, the data of the first character could be received from network device via SNMP protocol, the data of the second and third characters would be received after the analysis of the captured network packets. These characters are used as DDoS detection parameter in this paper when the detection model is intended to be built.

C. RELATIVE RESEARCHES

Recently there are three detection methods in DDoS attacks: DDoS attacks detection based on protocol analysis[2], DDoS attacks detection based on cluster[3], DDoS detection based on the model of network traffic statistics[4,5]. However, these methods present some problems. For example, DDoS attacks detection based on protocol analysis is effective relatively only for the attacks with obvious abnormal protocol characters, whereas it does no significant effects to DDoS attacks without obvious protocol characters. DDoS attacks detection based on cluster often make the high error rate, and there is large data needed to be conducted. In addition, it is unable to tell whether the abnormal network traffic is caused by the visit of normal users or by DDoS attacks.

Detection methods of DDoS attacks mentioned above, the only way to improve the detection efficiency of DDoS attacks is to combine various detection methods so as to make up for the deficiency of these detection methods above.

Ref. [2] studied TCP connection status, which is used to identify TCP SYN FLOOD attack. This method is only for TCP SYN FLOOD attack and can not detect or protect UDP FLOOD and ICMP FLOOD. Ref. [3] proposed a DDoS attack detection model based on protocol analysis and cluster. This model uses data mining algorithm to analyze the protocol information in packets, the advantages of it are that the attack detection method need not any manual-construction data, and it keeps a comparatively high detection rate. However, the number of network connections in unit time will range from ten thousand to one million for a large scale network, so the number of network connections can not be reduced to an optimal level with association algorithm. Because of the large calculation data the situation of detection delay will appear which results in the failure to achieve the real-time detection of DDoS attacks. Ref. [4,5] detects and prevents DDoS attacks through building detection model of abnormal traffic statistics, this model could not conduct abnormal traffic well and make sure whether the abnormal traffic is caused by legal high traffic or DDoS attacks.

ISSN(Online) : 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 2, February 2017

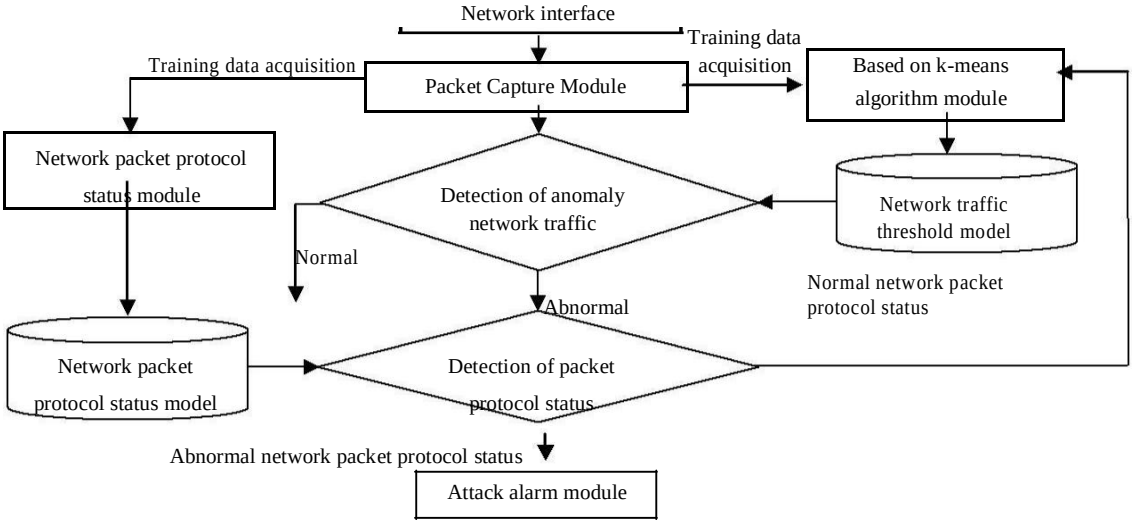

Fig1. DDoS attack detection model

Fig.1. is the working trafficchart of this model that combines the abnormal traffic detection and the packet protocol status detection to detect DDoS attacks. The specific detection procedures are as below:

Step 1 : Network packets Capture module achieve current network traffic value and packets.
Step 2 : Build each detection module through Step 3 and Step 4.
Step 3 : Capture the network traffic value based on k-means data mining module to build initiating the threshold value module of network traffic in which the threshold value would adjust automatically to the result ofpackets protocol status detection.
Step 4 : All of the captured packet are used to build the packet protocol status model based on Apriori and FCM algorithm.
Step 5 : Import the current traffic to abnormal traffic detection module. Once the current network traffic is beyond the threshold value, the network packet protocol status module will be started immediately.
Step 6 : Detect the packet protocol status via the network packet protocol status detection module. If there are abnormal packets, this module will give an alarm, whereas no abnormal packets the current network traffic will be clustered again by way of k-means data mining module to build a new threshold value model.

D. THE ESTABLISHMENT OF ABNORMAL TRAFFIC DETECTION MODULE

In this module used the most obvious characters of network traffic are used to build the detection model. Once the network traffic detection module detect the traffic value is abnormal, the packet protocol status detection module will be started immediately to detect these packets in order to make sure if the current abnormal network traffic value results from DDoS attacks, at the same time the traffic threshold model could dynamically update according to the result of the packet protocol status detection module. It could adapt to the traffic variety caused by the increase of users or the change in application. The achievement of this module will greatly reduce the data which needs to be detected and detect DDoS attack in real-time.

k-means[6] is one of cluster algorithm in data mining that preset a data set which contain n data object and k cluster that needs to create. The main idea of k-means algorithm is to split the data object set into k

ISSN(Online) : 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 2, February 2017

cluster($k \leq n$) that could make a standard measure function optimization and make high similarity of data object in the same cluster.

In this model the network traffic is captured in fixed time window for a period of time. All of the data captured is used to calculate the mean value of every time window via cluster algorithm.

The particular algorithm procedure is as follows:

Step1 Select at random K initial cluster center
$K_1, K_2, ..., K_k$ in m time window

Step 2 Calculate the distance between each network traffic data x_i and initial cluster center through $D_j = \min\{\| x - K_I \|\}$, the sample point that is the nearest to cluster center would be assigned to the cluster whose center is K_v

Step 3 Move every K_w to its cluster center and recalculate the cluster center according to new data added in cluster. Then calculate the deviation including sample value in each cluster domain through formula:

$$d(x_i, K_r)^2].$$

Step 4 The repetitive execution of step3 and step4 until the convergence of D value and all the cluster center will not move. After that the cluster center is the traffic mean value in different time window.

E. THE ESTABLISHMENT OF PACKET PROTOCOL STATUS DETECTION MODULE

Time windows are used as the unit to deal with packet protocol status in this model. For the accurate description of DDoS attacks, Dst-ip, Dst-port, flag, src-byte and dst-byte are used as detection parameters of the packet protocol status: Dst-ip represents destination IP, Dst-port represents destination port, flag represents the status of connection end whose values includes SF(normal connection end) and REJ(connection requests refusal) etc, src-byte represents source byte, dst-byte represents destination byte.

1. The feature extraction of packet protocol status

Apriori association algorithm is used in mining of packet protocol status. The packet protocol status appearing frequently in the network could be combined into one association record, which has massive packet protocol status compressed and reduces the data numbers. The following is an association record extracted from frequent items of the packet protocol status records:

(Dst-ip:192.168.9.8 and Dst-port： 8081--->flag:sf) [support=2.5% confidence=98.5%]

This association record shows the destination IP is 192.168.9.8 and the destination port is 8081,the support of normal service is 2.5% and the confidence is 98.5%.

(Dst-ip:192.168.96.9andDst-port:8000 and dst-byte=0--->flag:s0)[support=8.01% confidence=96.7%]

support of packets in which the number of bytes received is zero is 8.01% , the confidence is 96.7%, and this is the abnormal connection.

ISSN(Online) : 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 2, February 2017

2. Establishing the threshold value

Several protocol status characteristics generated through the use of association algorithm are used to calculate the distance between packet protocol status feature vector and each normal cluster in detection model. If the packet protocol status feature vector are beyond all the normal cluster, this feature vector is viewed as abnormal. If some feature vector in several consecutive time windows is marked abnormal, the attack alarm module will send alarm. Fuzzy C-means algorithm[7] is used to build the network protocol status model.

Fuzzy C-means clustering algorithm is described

As follows: Suppose a data set $X=\{x_1, x_2, ..., x_n\}$, fuzzy matrix $U=[u_{ij}]$ stands for its fuzzy C division, u_{ij} should meet the following condition:

$$\forall j, \sum_{i=1}^{C} u_{ij} = 1; \forall i, j \quad u_{ij} \in [0,1] \quad \forall i, \sum_{j=1}^{n} u_{ij} > 0$$

Present the widely used cluster rule is to use the minimal value of weighted sum squared error in cluster. That is

$$(\min) J_m(U, V) = \sum_{j=1}^{n} \sum_{i=1}^{C} u_{ij}^m d_{ij}^2 (x_j, v_i)$$

U is data and the No.j cluster center. In order to get the minimum value of cluster rule function $J_m(U, V)$, Lagrange multiplier method could be used to reach the necessary condition to get the minimum value of $J_m(U, V)$ that is:

$$ij = 1 / \sum_{l=1}^{K} (d_{ij} / d_{rl})^{2/(m-1)}, \forall i \qquad (1)$$

$$ij = (\sum_{i=1}^{m} u_{ij}^m x_j) / (\sum_{r=1}^{m} u_{ij}^m), \forall j \qquad (2)$$

U2 is sample space, V is clustering Prototype and $_d 2(x, v)$ is the euclidean distance between the No.i i j data and the No.j cluster. In order to get the minimum value of cluster rule function $J_m(U, V)$, Lagrange multiplier method could be used to reach the necessary condition to get the minimum value of $J_m(U, V)$ that is:

$$ij = 1 / \sum_{l=1}^{K} (d_{ij} / d_{rl})^{2/(m-1)}, \forall i \qquad (1)$$

$$ij = (\sum_{i=1}^{m} u_{ij}^m x_j) / (\sum_{r=1}^{m} u_{ij}^m), \forall j \qquad (2)$$

In this model the packet is captured in fixed time window. The data collected continuously for a period is used to calculate the packet protocol status threshold through the FCM cluster algorithm.

ISSN(Online) : 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 2, February 2017

The specific algorithm procedure is as follows:

Step 1 Select all association record of the packet protocol status in very time window

Step 2 Use number at random in initialize the subordinate-matrix U ,and should meet the condition $\sum_{i=1}^{n} u_{ij} = 1, \forall j = 1,..., n$.

Step 3 Calculate the cluster center through the formula m,

$$c_{ij} = (\sum u_{ij}^{m} x_j) / (\sum u_{ij}^{m}), \forall j$$

Step 4 Calculate the new subordinate-matrix U in each time window through formula k,

$$u_{ij} = 1 / \sum_{l=1} (d_{ij} / d_{rl})^{2/(m-1)}$$

Step 5 Loop computations step3 and step4 until the cluster rule function is less than a threshold value. The value is the cluster result.

Ⅲ. EXPERIMENT

Attack detection experiment platform is built in some campus network, the target host is in the network computer center, the operating system is Windows 2000 server, and WWW and FTP service are also used, attack hosts are distributed in this campus network. When the attack detection experiment is under way, the network traffic and packets of the target host would be collected and lasted for two weeks, then the collected network traffic value and packets would used as training samples to generate the threshold model of network traffic and detection model of packet connection status. DDoS attack tools include TFN2K, Stacheldraht, SYN Flood, Trinoo. DDoS attacks in this experiment will be detected in different duration.

From the analysis of DDoS attacks in this experiment, it is found that this system has a high detection efficiency, the detection rate of it could reach more than 97%. moreover, with the increase in the duration of DDoS attacks, higher is the attack detection rate of this system. The function test result of this system shows that it could meet the daily detection needs well.

IV. CONCLUSION

This paper, aiming at the characteristics of difficult detection and prevention as to DDoS attacks, proposes a detection model based on data mining. This model could receive the currently normal network traffic model with data mining algorithm. Once network traffic appears abnormal, this model could detect the packets maintaining in abnormal traffic duration. In this way the system load will be greatly reduced and its real-time can be improved. Through the test in LAN, this system is able to effectively. Through the above calculation step the deviation of membership of every sample point could be determined in its time window. According to the rule of the maximum deviation of membership in fuzzy date set, it should be determined that which cluster center every sample point belongs to so as to make sure whether there are DDoS attacks in current network.

ISSN(Online) : 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 2, February 2017

REFERENCES

[1] Meyer L, and Penzhorn WT. "Denial of service and distributed denial of service-today and tomorrow,"In: Proc. of the IEEE 7th AFRICON Conf. Vol.2,pp.959-964,2004.
[2] Wang HN, Zhang DL, and Shin KG. "Detecting SYN floodingattacks,"In:Proc.of the 21st Annual Joint Conf. of the IEEE Computer and Communications Societies,Vol.3,pp.1530-1539,2002.
[3] Gao Neng, Feng Deng-Guo, and Xiang Ji. "A Data-Mining Based DoS Dectection Technique," Chinese Journal of Computers,Vol.29, pp.944-951,2005.
[4] Li J, and Manikopoulos C. "Early statistical anomaly intrusion detection of DOS attacks using MIB traffic parameters,"In:Proc.of the IEEE Systems,Man and Cybernetics Society,Information Assurance Workshop,pp.53-59,2003.
[5] Kim YW, Lau WC, Chuah MC, and Chao HJ. "Packetscore:Statistical-Based overload control against distributed denial-of-service a ttacks,"In:Proc.of the 23rd Annual Joint Conf.of the IEEE Computer and Communications Societies,Vol.4,pp.2594-2604,2004.
[6] SUN Ji-Gui, and LIU Jie. "Clustering Algorithms Research," Journal of Software,Vol.19,pp.48-61,2008.
[7] YANG De-Gang. "Research of the Network Intrusion Detection Based on Fuzzy Clustering," Computer Science, Vol.32, pp.86-91,2015.
[8] N. Japkowicz S. Stephen "The Class Imbalance Problem: A Systematic Study" Intelligent Data Analysis vol. 6 no. 5 pp. 429-449 2002.
[9] G.M. Weiss F. Provost "Learning When Training Data Are Costly: The Effect of Class Distribution on Tree Induction" J. Artificial Intelligence Researchvol.19pp.315-354-2003.
[10] R.C. Holte L. Acker B.W. Porter "Concept Learning and the Problem of Small Disjuncts" Proc. Int'l J. Conf. Artificial Intelligence pp. 813-818 1989.
[11] J.R.Quinlan "Induction of Decision Trees"Machine Learning vol. 1 no. 1 pp. 81-106 1986.
[12] T. Jo N. Japkowicz "Class Imbalances versus Small Disjuncts" ACM SIGKDD Explorations Newsletter vol. 6 no. 1 pp. 40-49 2004.
[13] N. Japkowicz "Class Imbalances: Are We Focusing on the Right Issue?" Proc. Int'l Conf. Machine Learning Workshop Learning from Imbalanced Data Sets II 2003.
[14] R.C. Prati G.E.A.P.A. Batista M.C. Monard "Class Imbalances versus Class Overlapping: An Analysis of a Learning System Behavior" Proc. Mexican Int'l Conf. Artificial Intelligence pp. 312-321 2004.
[15] S.J. Raudys A.K. Jain "Small Sample Size Effects in Statistical Pattern Recognition: Recommendations for Practitioners"IEEE Trans. Pattern Analysis and Machine Intelligence vol. 13 no. 3 pp. 252-264 Mar. 1991.