A real-time DDoS attack detection and prevention system based on per-IP traffic behavioral analysis

Conference Paper · August 2010

DOI: 10.1109/ICCSIT.2010.5563549 · Source: IEEE Xplore

3 authors, including:

Guofeng Zhao
Chongqing University of Posts and Telecommunications
60 PUBLICATIONS **700** CITATIONS

SEE PROFILE

A Real-Time DDoS Attack Detection and Prevention System Based on per-IP Traffic Behavioral Analysis

Yi Zhang
College of Communication and Information Engineering
Chongqing University of Posts and Telecommunications
Chongqing china
cyzhangyi@126.com

Qiang Liu
College of Communication and Information Engineering
Chongqing University of Posts and Telecommunications
Chongqing china
liuqiang_1110@163.com

Guofeng Zhao
College of Communication and Information Engineering
Chongqing University of Posts and Telecommunications
Chongqing china
zhaoguof@gmail.com

***Abstract*—While many offline-based detection approaches have been well studied, the on-line detection of DDoS attack at leaf router near victims still poses quite a challenge to network administrators. Based on per-IP traffic behavioral analysis, this paper presents a real-time DDoS attack detection and prevention system which can be deployed at the leaf router to monitor and detect DDoS attacks. The advantages of this system lie in its statelessness and low computation overhead, which makes the system itself immune to flooding attacks. Based on the synchronization of TCP and UDP protocol behavior, this system periodically samples every single IP user's sending and receiving traffic and judges whether its traffic behavior meets the synchronization or not. A new non-parametric CUSUM algorithm is applied to detect SYN flooding attacks. Moreover, this system can recognize attackers, victims and normal users, and filter or forward IP packets by means of a quick identification technique. Finally, experiment results show that the system can make a real-time detection for flooding attacks at the early attacking stage, and take effective measures to quench it.**

Keywords-DDoS; real-time; traffic behavior; CUSUM algorithm; Early Stage

I. INTRODUCTION

During past decade, DDoS attacks have posed a powerful security threat to many ISPs, and brought enormous economic losses to them. In May 2009, hackers induced DDoS attacks towards some of China Telecom's main DNS servers, resulting in the problem that hundreds of websites stopped services. According to Arbor's survey in 2008, SYN flooding attacks, DNS flooding attacks and Smurf attacks are three main ways of DDoS attacks, and 76% of which are SYN flooding attacks [1].

The major steps of DDoS attacks are shown as follows. Firstly, attackers exploit the technology of client/server, and establish a botnet by combining with multiple vulnerable computers. Secondly, attackers send commands to the botnet and launch the attack of Denial-of-Service (DoS) for one or several targets. The botnet will increasingly amplify the power of the DoS attack and make the target consume many system resources. Finally, the victims can not work normally. Based above analysis, DDoS attacks include the following three main characteristics: (1) the quantity of attack source is gigantic but individual attack traffic is little, (2) attacker's traffics often resemble legitimate traffic and (3) the attack patterns will be mixed up to ignite a real attack.

To deal with above problems, this paper put forward a per-IP behavior analysis approach, which is implemented in an online, real-time DDoS attack detection and prevention system. The main contributions of this paper are shown as follows.

(1) Based on per-IP behavioral analysis, a new DDoS detection system is realized. For each IP user, our system will create records for every single IP user's sending and receiving traffic and judge its behavior whether meets the normal principles. Comparing with recording huge number of flows, our approach can greatly reduce the amount of computation and memory consumption.

(2) A specific packet identification technique is utilized to reach real-time flooding attack detection goal. It has improved the system performances dramatically.

(3) A non-parameter CUSUM algorithm is applied to detect the abnormal behavior of each IP. Based on a decision algorithm, each IP user will be classified as attacker, victim or normal user. After differentiate the attacker, the system will block its traffic and forward the normal user packets.

The remainder of this paper is structured as follows. Section II surveys related work. Section III introduces the system architecture. Section IV describes the proposed flooding detection algorithm. Section V evaluates the system and shows its performance results. Section VI concludes the paper.

II. RELATED WORK

There exists a variety of attack detection and defense mechanisms for DDoS attack detection. With respect to detection time and accuracy, there are two groups of DDoS attack detection techniques. One group is offline-based and the other one is online-based.

978-1-4244-5539-3/10/$26.00 ©2010 IEEE

Generally, for offline detection mechanisms, they are classified into two groups as specific detection and anomaly-based detection. Specific detection uses rule-match methods to justify whether monitored traffic have special attack features [2]. The rule-match approaches maintaining per flow state and matching packets to a pre-defined set of rules [3] has shown a certain good capability. However, rule-match approaches unlikely detect unknown DDoS attacks.

For previous unknown DDoS attacks, anomaly-based detection has higher accuracy than rule-match approach. Anomaly-based detection models the behavior of normal traffic and then reports any anomalies. PCA, entropy and subspace methods have demonstrated accuracy and efficiency in detecting network-wide traffic behavior anomalies. Lakhina et al. [4] made use of maximum and relative entropy and subspace to mine and analyze traffic anomalies. Ringerg et al. [5] used PCA (principal Component Analysis) to analyze the origin-destination flow aggregation and entropy time series of traffic features. However, most of these network-wide anomaly detection and machine-learning approaches are performed offline. Thus, it is difficult for them to take timely preventive measures for DDoS attacks.

In order to real-timely detect and defense DDoS attacks, on-line detection techniques are now paid wide attention. Generally, on-line detection techniques are statistical approaches regarding traffic feature and behaviors. Consequently, computation, memory consumption and detection time are key concerns about on-line detection. Wang et al. [6] proposed a behavioral-distance based anomaly detection mechanism.

III. System Architecture

Basically, our system is deployed at the entrance to the victim subnet, such as campus network, access network. The system architecture is shown in Figure 1 and it can be divided into three layers, i.e. application layer, network layer and driver layer.

A. Application layer

The application layer comprises user-controlled module (M1), system management module (M2) and data unload module (M8). Its functions are as follows: (1) to turn on and off the real-time detection, (2) to set a variety of detection parameters, such as the detection period and so on and (3) to unload data in three buffers. Application layer provides the user with a user-friendly operating platform.

B. Network layer

Network layer includes attack feature training module (M3), attack detection module (M6) and data buffer update module (M7). This layer has three functions. The first one is extracting flow features and storing them into the corresponding IP record; the second one is determining whether the traffic behavior of each IP is abnormal; and the last one is updating the data buffer.

To achieve real-time packet sampling and attack detection, three data buffers are set up, i.e. OIT (Observed IP Table) buffer, SIT (Safe IP Table) buffer and AIT (Alarm IP Table) buffer. During observation period, IP records which have not been determined as attackers are stored in OIT buffer. Besides IP address and relative hash value, for every IP record, there are three kinds of data packet counters which specifically record the number of protocol handshake packets of TCP, UDP or ICMP protocol and other three counters for storing alarm number of negative behavior.

SIT data buffer contains IP records of normal users and victims, while AIT data buffer stores the one of attackers. SIT and AIT data buffer both include IP address, Hash value and a timer (Time_in), which stores the start time while an IP record is established. Finally, according to the timer and updating period, data buffer update module (M7) can delete the overdue records in time.

C. Driver layer

Driver layer consists of two modules of packet capture module (M4) and packet filtering module (M5). Firstly, the network card is set to the promiscuous mode. Secondly, according to the data packet classification algorithm, data is captured and stored in the data buffer. Finally, based on test results, the system automatically filters the attacker's traffic and forwards normal user traffic. In order to quickly classify and store IP packets, Hash function is utilized to calculate source address and destination address matching. Considering that this detection is carried on Linux kernel, the processing time should be short and the collision must be small. So we choose Hash function designed by Xu et al. [7], which has good performance.

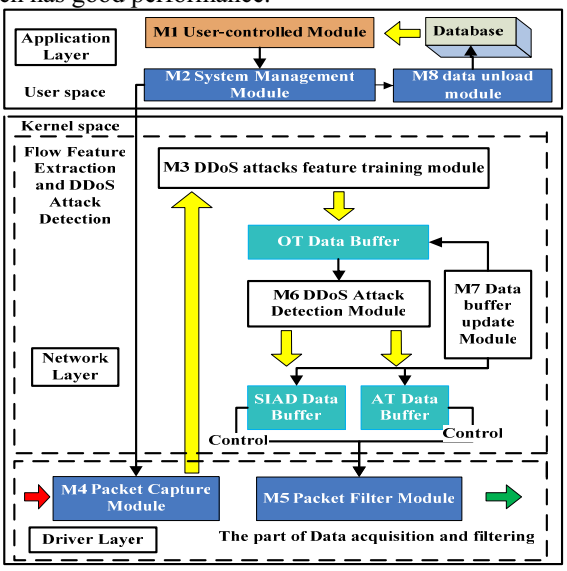

Figure 1. The overall framework of the system

IV. Flooding Detection Algorithm

A. Packet classification algorithm

Packet classification algorithm is implemented mainly in the module 4. Records of source IP address and destination IP address for each packet are respectively maintained in the OIT buffer. Supposing there are N IP addresses within the victim network and K IP addresses outside it, this algorithm

merely needs K+N IP records. However, previous classification based on per-flow would set up $K \times N$ IP records. Obviously, our method reduces memory consumption and computation. Being more important, this classification method can achieve aggregation of attack traffic. In other words, the low-rate traffic of every attack source will be disclosed quickly through IP records related to victim at the early stage of DDoS attacks. Thus, the detection algorithm can rapidly find out the attack.

B. Flood feature extraction

Packet classification determines efficiency of real-time detection and memory need, but extraction and storage of flood feature plays a significant role in flooding attack detection. In fact, TCP, UDP and ICMP protocols are based on Client-Server model, and most of hosts in the network often play the dual role of client and server. Accordingly, after analyzing the relevance between clients and servers, it shows that a strong synchronization exists between client's receiving SYN/ACK packets (SYN/ACK_rev) and its sending ACK packets (ACK_send). There is also a synchronization relationship between following three types of packets: server's receiving SYN packets (SYN_rev), its receiving ACK packets (ACK_rev) and sending RST packets (RST_send).

Generally, most attackers choose many unreachable addresses as spoofed source addresses. Therefore, when a DDoS attack starts, the strong correlation between the protocol handshake packets for TCP protocol will not keep. For DNS flooding attack and Smurf attack, the system can find out the attacks by checking the mismatch between the request packets and response packets.

C. DDoS detection algorithm

The DDoS detection algorithm proposed here consists of two parts: Recognition and Decision. For every IP records in OIT data buffer, Recognition algorithm periodically samples and inspects each IP records, and finds out suspicious IP records with negative behaviors. When it detects the negative behavior of the suspicious IP record, the alarming counter related to corresponding attack type will be automatically accumulated. At the end of an observation period, decision algorithm testes every attack alarming counter of IP records. Considering a large safe margin, the detection period t_d is set to 30 seconds and the observation period t_o is 10 minutes. If one of the alarming counters value exceeds 3, decision algorithm can determine the suspicious IP as an attacker or a victim being attacked by this type of attacking. Moreover, to avoid system's misjudging "flash crowd" to be SYN flooding attacks, when the detection algorithm firstly finds out the attack, t_d will be extended to 1 minute.

The non-parametric CUSUM algorithm is applied to our detection algorithm. The basic formula is proposed in [8]. However, instead of collecting SYN-FIN or SYN-SYN/ACK packet pairs as proposed by them, this DDoS detection algorithm samples the SYN， SYN/ACK, ACK and RST packets of TCP protocol, and the request and reply packets of DNS and ICMP protocol.

For client's traffic behavior detection, let $\{\Delta_n, n=1,2,3,\cdots\}$ be the number of SYN/ACK_revs minus that of the corresponding ACK_sends collected within one detection period. And for server's traffic behavior detection, let $\{\Omega_n, n=1,2,3,\cdots\}$ be the number of SYN_revs minus that of the corresponding SYN/ACK_sends and RST_sends collected within one detection period. To alleviate the dependencies on the observation period, Δ_n is normalized by the average number F_1 of SYN/ACK_revs, and Ω_n is normalized by the average number F_2 of the sum of ACK_revs and RST_sends. F_1 and F_2 can be estimated in real time and updated periodically by the following recursive function.

$$F_1(n)=\eta F_1(n-1)+(1-\eta)\mathrm{SYN/ACK}(n) \quad (1)$$

$$F_2(n)=\mu F_2(n-1)+(1-\mu)\mathrm{ACK\_rev(RST\_send)}(n) \quad (2)$$

Where η and μ are two coefficient lying strictly between 0 and 1. According to the highest weight of our past observations, η and μ are set to 0.7.

Define $\{X_n = \Delta_n / F_1, n = 1,2,3\cdots\}$ and $\{Y_n = \Omega_n / F, n = 1,2,3\cdots\}$. So, $\sum(X_n) = c$ 1 and $\sum(Y_n)$ =d 1. Moreover, $\{X_n\}$ and $\{Y_n\}$ are no longer dependent on the network size and time-of-day. However, the CUSUM algorithm requires a negative drift before a change and positive drift after the change. We define $X_n = X_n - \alpha$ and $Y_n = Y_n - \beta$， where $\alpha > c$ and $\beta > d$. By this way, X_n and Y_n can fulfill the requirement of the CUSUM algorithm. During the DDoS attack, X_n and Y_n dramatically rise and become positive. But X_n and Y_n keep negative for normal traffic.

Let

$$y_n = (y_{n-1} + X_n)^+, \ y_0 = 0 \quad (3)$$

$$z_n = (z_{n-1} + Y_n)^+, \ z_0 = 0 \quad (4)$$

where x^+ is equal to x if $x > 0$ and 0 otherwise. y_n represents a stationary random process of client's traffic behavior, and z_n represents that of server's traffic behavior.

$d_N(.)$ represents the decision at the end of every detection period, '0' for normal operation and '1' for attack (a change occurs). N_a and N_v respectively represent the detection threshold of the attacker and the victim.

$$d_{N_a}(n)=\begin{cases}0 & \text{if } y_n \le N_a \\ 1 & \text{if } y_n > N_a\end{cases}, \quad d_{N_v}(n)=\begin{cases}0 & \text{if } z_n \le N_v \\ 1 & \text{if } z_n > N_v\end{cases} \tag{5}$$

V. PERFORMANCE EVALUATION

This section describes the performance evaluation results of our DDoS attack detection and prevention system which have run at the Network Management Center of Chongqing University of Posts and Telecommunications for 6 days. Figure 2 shows the network topology structure of the tested network. By monitoring the data from the leaf router of the campus network of CQUPT, this system respectively real-timely analyzes and detects the clients' and servers' traffic in the case of SYN flooding attacks and normal operations. The detected objects contain two websites mainly offering the access by HTTP protocol, such as BBS and redrock, and a website offering the service of Downloading Music. They are mainly the representative of servers' normal traffic. 172.22 network segment is chosen as client's normal traffic, which covers three types of areas, such as the office area, the dormitory area and the teaching area.

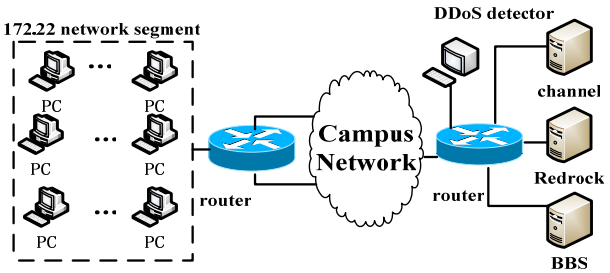

Figure 2. CQUPT network topology structure

A. Normal Traffic Behavior

Based on the principle of SYN flooding attacks, the time dependency and the relevance between TCP SYN, SYN/ACK, ACK, RST packets are tested. Figure 3(a) illustrates the changes of the relevance between server's receiving SYN (SYN_rev) packets, sending SYN/ACK (SYN/ACK_send) packets, and the sum of ACK_rev and RST_send. With time going on, the data of websites of two types shows that the number of SYN_rev minus the total number of ACK_rev and RST_send is very small. Figure 3(b) describes the changing tendency of the total amount of client's sending SYN (SYN_send), ACK (ACK_send) packets and receiving SYN/ACK(SYN/ACK_rev) packets. Obviously, the relevance between SYN/ACK_revs and ACK_sends is very strong. So the relevance is an inherent attribute of traffic behavior, and keeps invariant for different network sizes and detection periods. This is also the theoretical basis of on-line analysis and detection of DDoS attacks.

For detection algorithm, y_n and z_n are the important detection parameters. Figure 4 illustrates the accuracy of CUSUM algorithm in the case of normal traffic. Obviously, y_n and z_n are always 0. Moreover, several bursts don't exceed the threshold of $N_a = 0.4$ and $N_v = 0.5$.

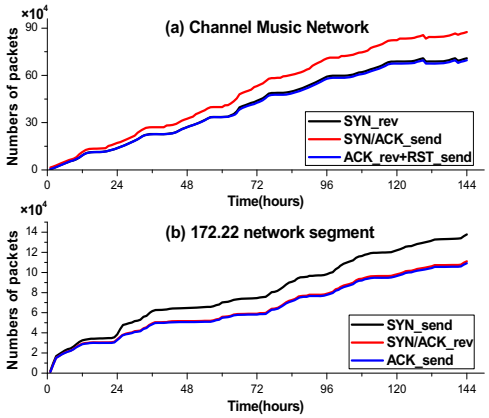

Figure 3. The dynamics of TCP three-way handshake packets of client and server

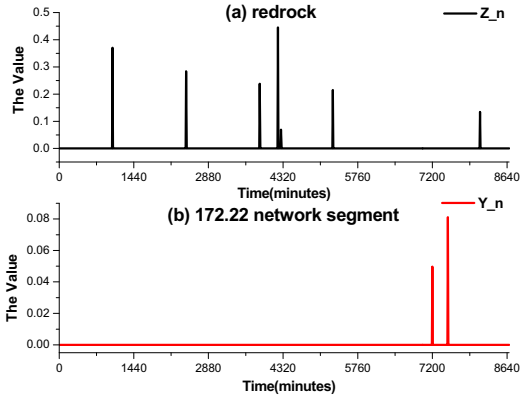

Figure 4. CUSUM test statistics under normal operation

B. SYN Flooding Detection

In order to investigate the accuracy of the DDoS attack detection algorithm and defensive effects, in CQUPT campus network, we launched three SYN flooding attacks to a website redrock. As shown in the Table 1, each attack lasts 10 minutes and has different flow rate. Figure 5 shows the variables of y_n and z_n in the attack case. Theoretically, y_n and z_n should be near to zero. However, when the attack traffics come, the two values may rapidly jump over the thresholds N_a and N_v respectively. The system does not immediately take defensive measures to stopping the attack, but keep observing the suspected IP record. After the alarming of attacks counts more than three, the system starts to filter the traffic from the attackers. Typically, because most attackers spoof its source IP to unreachable addresses, the server cannot receive their ACK packets to complete the TCP connection. Therefore, in the records, the number of transmitted ACK packets from attackers could not be updated. After an IP being determined to be an attacker, y_n always keeps the same value and greater than N_a. However, as for the server, due to the normal communication being recovered z_n can reduce rapidly below N_v when the IP address being judged to a victim.

In order to evaluate the detection sensitivity, there are two parameters being studied which are the period to detect attacks initially and the period to take defensive measures. Based on the testing result of three attack patterns, the above two parameters are summarized in the table 1. Obviously, the detection algorithm always firstly finds out the attackers, and then the victim is detected. However, under three attack patterns, there is a very short interval between the periods to detect the attacker and victim. In the first attack pattern, we use a single attack source to attack the website and the traffic rate is 10 SYNs/s. The attack traffic rate in the third attack pattern is 100 times larger than that of the pattern one. The traffic rate of each individual attack source is also 10 SYNs/s, but the attacker's number is 100.

TABLE I. THE DETECTION RESULTS UNDER THREE ATTACK PATTERNS

No	Attack source	Single attack SYNs/S	Total attack SYNs/S	Detect time (min)	Defense Time (min)
1	1	10	10	1.1	2.5
				4.5	5.5
2	1	80	80	0.8	2
				1.2	2.5
3	100	10	1000	1.1	2.5
				1.05	2.5

However, the system detected attacks both under the first and third attack pattern. The period to detect the victim in third pattern is much smaller than the first pattern's. Such results show that our approach proposed in this paper is effective.

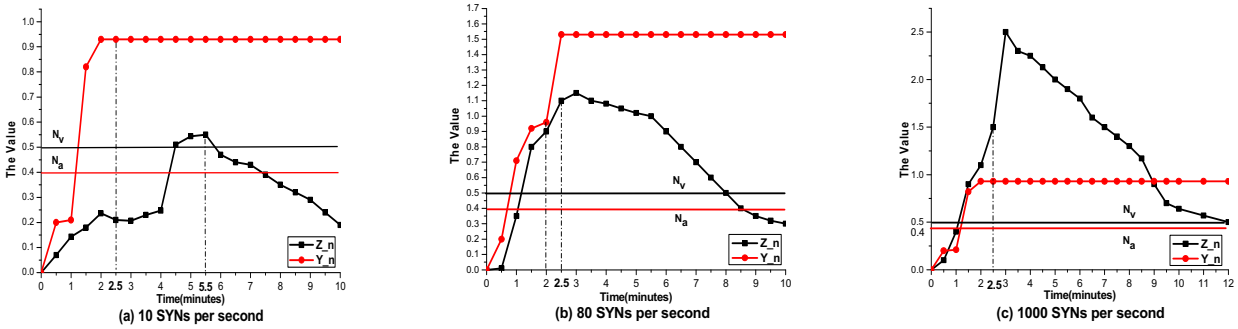

Figure 5. In different attack mode, the sensitivity of SYN flooding detection

VI. CONCLUSION

Based on analyzing per-IP traffic behavior approach, a real-time DDoS attack detection and prevention system is realized. It has three advantages shown as follows. 1) Based on per-IP traffic behavior analyses, it is easier to differentiate the attackers from the normal users. 2) Because our approach needs less computation and memory, the system could be deployed for on-line DDoS detection and prevention. 3) By applying the non-parameter CUSUM algorithm and decision algorithm, this system can detect attacks accurately at the earlier attack stage. Moreover, our system can quickly filter the attack traffics and forward the normal traffics simultaneously by means of the fast identification technology. On a campus network, to investigate our system, many tests have been done. The results show that the system has high DDoS detection accuracy and short detection time.

Besides for SYN flooding attacks detection, the system can be utilized to detect DNS flooding attacks and Smurf attacks. That means our system has a wider applying field.

ACKNOWLEDGEMENT

This work is funded by key project and Chun hui project both supported by Chinese Education Ministry (No.208117) and project supported by Chongqing Education Committee (No. KJ070516).

REFERENCE

[1] Arbor Networks. Worldwide Infrastructure Security Report, http://www.arbornetworks.com/report, Sept 2008.

[2] T. Peng, C. Leckie and R. Kotagiri, "Survey of network-based defense mechanisms countering the DoS and DDoS problems", ACM Comput. Surv. 39, April 2007.

[3] R. Sommer and V. Paxson, "Enhancing byte-level network intrusion detection signatures with context", CCS, 2003.

[4] X. Li, F. Bian, M. Crovella, C. Diot, R. Govindan, G. Iannaccone and A. Lakhina, "Detection and identification of network anomalies using sketch subspaces", IMC, 2006.

[5] H. Ringerg, A. Soule, J. Rexford and C. Diot, "Sensitivity of pca for traffic anomaly detection", SIGMETRICS, 2007.

[6] Hemant Sengar, Xinyuan Wang, Haining Wang, Duminda Wijesekera and Sushil Jajodia, "Online Detection of Network Traffic Anomalies Using Behavioral Distance", IEEE IWQoS 2009 , Charleston, July 2009.

[7] Chuan Xu, Hong Tang and Guofeng Zhao, "Design and Complementation of a Real Time Traffic Measurement System in High-speed Networks", IEEE Proceedings of NPC 2008, Oct 2008, pp.341-344.

[8] Haining Wang, Danlu Zhang and Kang G. Shin, "Detecting SYN Flooding Attacks", IEEE INFOCOM 2002, New York City, June 2002.