Mitigating Distributed Denial-of-Service Attacks Using Network Connection Control Charts

Qingtao Wu
Electronic Information Engineering College, Henan University of Science and Technology
Luoyang, Henan Province, China
86-379-64231192
wqt8921@126.com

Haichao Zhang
Electronic Information Engineering College, Henan University of Science and Technology
Luoyang, Henan Province, China
86-379-64231192
haichaozh@tom.com

Jiexin Pu
Electronic Information Engineering College, Henan University of Science and Technology
Luoyang, Henan Province, China
86-379-64231795
pjx@mail.haust.edu.cn

ABSTRACT
In this paper, we present a simple, automated response model that utilizes the Shewhart's control charts based-on network connection to aid in handling DDoS attacks. This model is designed to prevent incoming traffic from exceeding a given threshold, while allowing as much incoming, legitimate traffic as possible. In addition, this model focuses on requiring less demanding modifications to external routers and networks than other published distributed response models that impact the effect of DDoS attacks. The experimental results show the effectiveness of our scheme in early mitigating DDoS attacks.

Categories and Subject Descriptors
K.6.5 [**Management of computing and information systems**]: Security and Protection –*Insurance, Invasive software, Unauthorized access.*

General Terms
Design, Experimentation, Security.

Keywords
Network security, Distributed Denial-of-Service, Shewhart's control charts, Automated response model

1. INTRODUCTION
Owing to the explosive growth of network-related technology and popularity of Internet, computer network has become a necessary part of our daily life. Hence, it has become critical to protect the availability of Internet services and resources. Traditional security considerations revolve around protecting the network connection's confidentiality and integrity, protecting the system from break-in, and protecting the user's private information from unintended disclosure. However, one area that has long been neglected is that of service availability in the presence of distributed denial of service (DDoS) attacks. As emergency and essential services become more reliant on the network as their infrastructure, the consequences of DDoS attacks could result in serious financial loss on E-commerce, even become life-threaten. So, how to defend against DDoS attacks has become one of the extremely important research issues in the Internet security community [1,2].

Infoscale 2007, June 6-8, 2007, Suzhou, China

A key problem to tackle when solving DDoS attacks is attack detection. A real-time and early detection are critical to defend against the dynamics of DDoS attacks. Previously proposed approaches for detecting DDoS attacks depend on monitoring the volume of traffic that is received from the victim [3,4,5]. Because of the stateless feature of the IP protocol, there is no proper way to authenticate the IP flows. Thus, a potential problem with traffic monitoring is that it is hard to tell the attack traffic from the legitimate traffic. One link is experiencing unusually high traffic is not necessarily the indicator for a DDoS attack. It could be "flash crowd" events that should be warmly welcomed, where a large number of legitimate users access the same website simultaneously. The DDoS attack in the final stage can readily be identified through observing very abrupt changes in the network traffic. However, it is too late to react to the attack at this stage. So, we need an early detection mechanism for the early stage of the attacks so that the victim has more time to take action against the attacker.

In this paper, we develop an efficiently adaptive scheme for early detection of DDoS attacks, which involves modelling detection problem into process control problem. The scheme employs network connection control charts for monitoring suspicious behaviors, and utilizes threshold of test statistics to achieve attack detection. Besides, the scheme is self-learning, which enables it to adapt to diverse attacks. DDoS attacks detection model, as well as results of simulations, are presented.

The rest of the paper is organized as follows. In section 2, we discuss the choice of detection feature. In section 3, we give a detailed definition and description of the attack detection problem. In section 4, we focus on the design and implementation of DDoS attacks detection. In section 5, we present the simulation results of our detection scheme. Finally, conclusions are drawn in section 6.

2. FEATURE SELECTION
There are two problems to solve for mitigating DDoS attacks. The first problem is how to distinguish the attack behaviors from the

normal behaviour. It is necessary to model the normal state of network or system. The second problem is to detect the DDoS attacks as soon as possible without raising a false alarm, so that the victim has enough response time to take action against the attacker before the victim is overwhelmed. A key to solve these two questions is to select the available detection feature, which could accurately tell the abnormal from normal state of system.

In the initial phases of DDoS attack, attacker mainly employs scanning tools, constantly sending request connection packets in order to acquire relevant information of the victim. This phase shows anomalies in the characteristics of network connections in the victim's network. Under the normal state (no attacks existed), we take statistics on the network connections based on source IP address in ten minutes and draw the frequency distribution diagram, as Figure 1 shows.

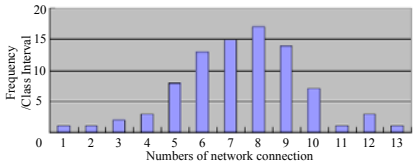


Figure 1. Probability Distribution Based-on Network Connections

The statistics show that in a normal state, the distribution of network connections in a given time period has the characteristics that the middle is high and both sides are low and nearly symmetric. What's more, the symmetry is better and better with more time. That is to say, in a normal state, network connections centered in a specific range. Its maximum and minimum have very low percentage of the whole. We assume that in a normal state, network connections in the victim nearly satisfy normal distribution, and describe it with normal probability distribution function. Therefore, we select network connections as detection feature. The number of network connections and their distribution are used to identify normal network pattern.

3. PROBLEM DESCRIPTION AND MODELLING

3.1 Statistical Description

In order to mitigate a DDoS attack, we need to be able to detect fluctuations in our detection feature over time. However, our detection feature is a random variable due to the stochastic nature of network flows. Consequently, before describing the proposed detection model, we discuss the details of the theoretical background of our detection scheme.

Suppose that random variables (X_1, X_2,$\cdots$, X_k) are the statistics of network connections based on source IP address in the time interval t in normal state, where k is the number of source IP addresses, and their corresponding values of an observation are ($\boldsymbol{x}_1$, $\boldsymbol{x}_2$, $\cdots$, $\boldsymbol{x}_k$) . The expectation $E(\mathbf{x})=\frac{1}{k}\sum_{i=1}^{k}\mathbf{x}_i=\boldsymbol{\mu}$, the standard deviation $\sigma=\sqrt{\frac{1}{k}\sum_{i=1}^{k}[\mathbf{x}_i-E(\mathbf{x})]^2}$, (1< i< k), X approximately obeys normal distribution $N(\boldsymbol{\mu},\boldsymbol{\sigma}^2)$, that is, the probability density function for network connections x_i has:

$$\Phi(\mathbf{u})=\frac{1}{\sqrt{2\pi}}e^{-\frac{\mathbf{u}^2}{2}} \tag{1}$$

where $\boldsymbol{u}$ is standard normal random variable given by

$$u=\frac{u_i-\mu}{\sigma} \tag{2}$$

According to above statistical description, we have definition:

Definition 1 Detection probability Suppose that (x_1, x_2,$\cdots$, x_k) is any group of measurement for network connections of k IP addresses, where $x_{\max}$ is the abnormal maximum, we have:

$$\Phi(\mathbf{u}_{\max})=\frac{1}{\sqrt{2\pi}}e^{-\frac{\mathbf{u}_{\max}^2}{2}} \tag{3}$$

So, detection probability is given by

$$P_d=P\{x_{\max}<x_i<\infty\}=\int_{(x_{\max}-\mu)/\sigma}^{\infty}\frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}dt=1-\Phi[(x_{\max}-\mu)/\sigma] \tag{4}$$

Definition 2 detection threshold Given detection probability is P_d, according to formula (3) and (4), detection threshold is defined by

$$x_{\max}=\mu+u_{max}\cdot\sigma \tag{5}$$

Where u_{max} is the constant looked up on the standard normal distribution table.

We make convention: if $\boldsymbol{x}$ > $\boldsymbol{x}_{\max}$, the source IP address corresponding to $\boldsymbol{x}$ is on invasive attack.

3.2 The Response model

Response model is implemented based on Shewhart's mean-range ($\bar{x}$ -R) control charts [6], as Fig. 2 shows. The $\bar{x}$ -R Control Charts is founded on the basis of statistic inference theory, partitioning the data in time series. In $\bar{x}$ -R control charts, $\bar{x}$ and R respectively reflect the variations of samples mean and range, where the operational characteristic of the data is arranged by the change of sequence data, and judged by the upper and lower control limitation of control charts whether the data variation lies in the normal state.

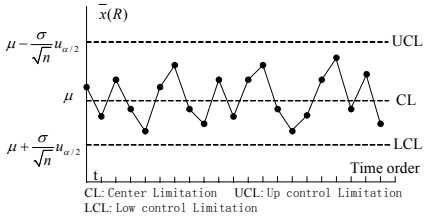


Figure 2. $\bar{x}$ -R control charts

Suppose that the population X obeys normal distribution N (μ, σ^2), n samples of X are observed in the time interval t. For our detection feature, we obtain n values. In terms of the property of normal distribution, samples mean $\bar{x}$ also obeys normal distribution $\bar{x}\sim N$ (μ, σ^2/n). In the light of Shewhart's control principle, if samples mean $\bar{x}$ satisfies

$$P\left\{\left|\frac{\bar{\mathbf{x}}-\boldsymbol{\mu}}{\boldsymbol{\sigma}/\sqrt{n}}<\mathbf{u}_{\alpha/2}\right|\right\}=1-\boldsymbol{\alpha} \tag{6}$$

Then, it is thought that the sample population mean is acceptable. Namely, in the normal state, samples mean lies between

$\mu - \frac{\sigma}{\sqrt{n}}u_{\alpha/2}$ and $\mu + \frac{\sigma}{\sqrt{n}}u_{\alpha/2}$. The central line, upper and lower control limitation of $\overline{x}$ control chart are given respectively by

$$\begin{cases} CL = E(\bar{x}) = \mu \\ UCL = E(\bar{x}) + u_{\alpha/2}\sqrt{D(\bar{x})} = \mu + u_{\alpha/2}\dfrac{\sigma}{\sqrt{n}} \\ LCL = E(\bar{x}) - u_{\alpha/2}\sqrt{D(\bar{x})} = \mu - u_{\alpha/2}\dfrac{\sigma}{\sqrt{n}} \end{cases} \quad (7)$$

$\bar{R}/d_2$ and sample mean $\bar{x}$ are separately used to evaluate σ and μ. Then, formula (7) are transformed to:

$$\begin{cases} CL = \bar{\bar{x}} \\ UCL = \bar{\bar{x}} + u_{\alpha/2}\dfrac{\bar{R}/d_2}{\sqrt{n}} = \bar{\bar{x}} + A_2\bar{R} \\ LCL = \bar{\bar{x}} - u_{\alpha/2}\dfrac{\bar{R}/d_2}{\sqrt{n}} = \bar{\bar{x}} - A_2\bar{R} \end{cases} \quad (8)$$

where A_2 is relevant to sample numbers n, and $A_2 = u_{\alpha/2}/(d_2/\sqrt{n})$. A_2 is constant, which could be known through table look-up.

The central line, upper, and lower control limitation of R control chart are derived similarly.

$$\begin{cases} CL = E(R) = \bar{R} \\ UCL = E(R) + u_{\alpha/2}\sqrt{D(R)} = (1 + u_{\alpha/2}\dfrac{d_3}{d_2})\bar{R} = D_4\bar{R} \\ LCL = E(R) - u_{\alpha/2}\sqrt{D(R)} = (1 - u_{\alpha/2}\dfrac{d_3}{d_2})\bar{R} = D_3\bar{R} \end{cases} \quad (9)$$

Where

$$D_3 = 1 - u_{\alpha/2}\frac{d_3}{d_2}, \quad D_4 = 1 + u_{\alpha/2}\frac{d_3}{d_2}$$

Both D_3 and D_4 are relevant to the number n of samples, and can be known from table look-up.

In the $\bar{x}$-R control charts, when network connections satisfy statistical normal distribution, detection data lies in the control field between the upper and lower control limitations. When data point locates out of the control range, suspicious behavior will be presumably to have been happened in the network. Then, the abnormal data must be further analyzed to investigate whether the suspicious activities indicate a real attack or not.

4. SIMULATION AND IMPLEMENTION

There are two key measures that are used to evaluate DDoS attack detection performance. The first is the detection accuracy, which is one of the biggest concerns among the intrusion detection community. If the attack reaction is taken according to the false detection results, innocent activity will be unfairly punished and normal network services are disturbed. The second is the detection time. One of the advantages for a DDoS detection scheme is to detect the attack as soon as possible so that proper reaction steps can be done earlier to minimize or eliminate the attack damage.

Unfortunately, these two parameters are a confliction pair. The rapid detection requires that we adopt simple method to analyse data so that the system can shorten detection time. The detection accuracy requires detection system could be numerical measurement so that attack action could be definitely judged, but this requires complex computing overhead. Thus, it is crucial for us to detect attacks accurately before the victim becomes overwhelmed. The two design goals, high detection accuracy and short detection time are achieved through modelling detection problem into process control problem and dividing the detection process into two phases. The first phase is to analyze the statistical network connections in terms of the $\overline{x}$-R control charts, finding out the suspicious data of network flows. The second phase is to determine exactly whether the suspicious data is an attack action or not according to detection threshold.

Detection model structure is as Figure 3 shows. It is composed of statistic analysis engine, anomaly analysis engine and attack detection engine. Statistic analysis engine analyzes the historical data, draws the $\overline{x}$-R control charts and determines the detection threshold. Anomaly analysis engine makes dynamic analysis about network connections in terms of the upper and lower control limitation of $\overline{x}$-R control charts, so that tells the suspicious data from network flows. Then, the normal data will be stored in the historical database, and the suspicious data will be sent into attack detection engine for further analysis. Attack detection engine analyses the abnormal data and determines whether the suspicious data is an attack action or not according to detection threshold.

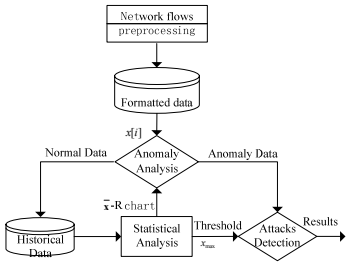


Figure 3. DDoS Attacks Detection Structure

The process of attack detection is as follows:

(1) Compute the sliding mean and range of historical network connections, and set up the $\overline{x}$-R control charts.

(2) Compute the attack detection threshold $x_{\max}$ according to given attack detection probability P_d.

(3) Monitor network flows, and keep counting of network connections during each time slots t. Preprocess these data and form the formatted data timely.

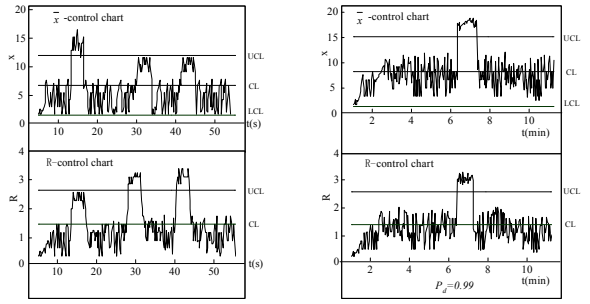
(4) Make anomaly analysis about formatted data according to the $\overline{x}$-R control charts. Store the normal data in the historical data, and send the abnormal data into attack detection engine to analyze further.

Analyse the suspicious data. If the suspicious data surpasses the detection threshold $x_{\max}$, it is indicated that IP address corresponding to the abnormal data makes an attack attempt. Then, attack decision engine will report attack attempt so that system could react to it.

5. EXPERIMENT AND RESULTS

We use DDoS attack tool attacking on a Web server in the Intranet. The historical network connections used in the experiment are collected in 30 minutes from the Intranet when no attacks have been happened. We pick out three hundreds of sequential data among these historical data. By the numeric computation based the formulas given above, we obtain the parameters of the $\bar{x}$ -R control charts and initial detection threshold.

Attack mode in the experiment of anomaly analysis is divided into two cases: (1) adaptive DDoS attack mode. In this case, the attacker tries to control the number of network connections to avoid detection by our scheme. In the experiment, the attack is launched for a short time, then stopped and launched again later, repeating this procedure for many times. At last, attacker launched the true attack. The analysis results of control charts are shown in Figure 4(a). In the $\overline{x}$ control chart only the first attack activity has been monitored. However, all attack activities have been monitored except first attack in the R control chart. (2) Normal DDoS attack mode. In this case, attacker suddenly launches a DDoS attack on Web server. Figure 4(b) illustrates the corresponding result of anomaly analysis. The results show that the $\overline{x}$ -R control charts could reflect the abnormal change of the network connections when the attack occurs.



(a) **(b)**

Figure 4. Results using $\bar{x}$ -R Control Charts in different attack modes

Then, we conducted the attack mitigation under a variety of different detection probability P_d and running time T_r. The average detection accuracy and detection time showed in Table 1.

Table 1 Mitigation results with different P_d and Tr

P_d	T_r=15min	
	Accuracy	Time
0.95	72%	7s
0.96	78%	9s
0.97	82%	10s
0.98	85%	11s
0.99	86%	13s

T_r	P_d =0.99	
	Accuracy	Time
1min	47%	8s
4min	67%	9s
7min	71%	11s
10min	80%	14s
15min	85%	17s

Detection accuracy is measured by Bayesian formula, as formula (10) shown.

$$P(I \mid A) = \frac{P(I)p(A \mid I)}{P(I)p(A \mid I) + P(-I)p(A \mid -I)} \quad (10)$$

where I and $-I$ denotes attack number and all normally corresponding number respectively, A denotes attack alarm number.

The efficiency of our detection scheme is validated by attack tests. The evaluation results show that the scheme has both a short detection time and high detection accuracy. Moreover, due to the dynamic adjustment of $\overline{x}$ -R control charts and detection threshold, our detection mechanism is adaptive to different kind of attacks. We also find that different detection probability will reflect on the detection accuracy and the detection time. Therefore, the selection of detection probability should synthetically consider the balance relationship between detection accuracy and detection time.

6. CONCLUSION

In this paper, we proposed a scheme to detect DDoS attacks by monitoring the network connections in the victim's network. We have also presented a response model based on control charts that can identify suspicious behaviors when an attack has occurred. The feasibility of the scheme is validated through the simulated tests. The experimental results show the effectiveness of our scheme in early detecting DDoS attacks.

There still existed some puzzles we can not involve in our paper. For example, we need a systematic and automatic procedure for setting the parameters of this mechanism, such as threshold value adjustment of the Detection Engine. We hope to build more a perfect solutions to defend against DDoS attacks in the near future.

7. REFERENCES

[1] J. Mirkovic, S. Dietrich, D. Dittrich, and P. Reiher, *Internet Denial of Service:Attack and Defense Mechanisms*. Upper Saddle River, USA: PearsonEducation, Inc., 2005.

[2] L. A. Gordon and M. P. Loeb, *Managing cyber-security resources: A costbenefit analysis*. McGraw-Hill, 2006.

[3] C. Gong and K. Sarac. IP traceback based on packet marking and logging. *in Proc. of IEEE International Conference on Communications*, Seoul, Korea, May 2005.

[4] Lim, B.P., Uddin, M.S. Statistical-based SYN-flooding detection using programmable network processor. *Information Technology and Applications, 2005. ICITA 2005. Third International Conference on Volume 2, Issue , 4-7* (July 2005), 465 – 470

[5] Sherif Khattab, Rami Melhem, Daniel Mossé, *et al*. Honeypot back-propagation for mitigating spoofing distributed Denial-of-Service attacks. *Journal of Parallel and Distributed Computing, Volume 66, Issue 9* (September 2006), 1152 - 1164

[6] M. Basseville and I. V. Nikiforov. *Detection of Abrupt Changes: Theory and Application.* Englewood Cliffs: Prentice Hall (1993).