社団法人　情報処理学会　研究報告
IPSJ SIG Technical Reports

2006－QAI－18（2）
2006／1／20

A DoS/DDoS Attacks Detection Scheme Based on In/Out Traffic Proportion

Zhang FengXiang† and Shunji ABE‡

† Department of Informatics, Graduate University for Advanced Studies, 2-1-2 Hitotsubashi, Chiyoda-ku, Tokyo 101-8430, Japan

‡ National Institute of Informatics, 2-1-2 Hitotsubashi, Chiyoda-ku, Tokyo 101-8430, Japan

E-mail: † fzhang@grad.nii.ac.jp, ‡ abe@nii.ac.jp

Abstract Denial of Service(DoS)/DDoS(Distribute DoS) attacks become the most prevalent threats against the widely used Internet. The goal of DoS/DDoS attacks is to prevent victim machines or networks from offering service to their legitimate users. Many detecting mechanisms based on traffic statistics properties have been proposed. However most of them are essentially based on unidirectional traffic changes. Thus they might result in serious false alarms when legitimately abrupt changes appear. We have proposed a heuristic detection scheme, which mainly checks the In/Out traffic proportion at the protected node's gateway or the router nearby. In normal cases, this kind of proportion is close to a constant value. By checking the likelihood ratio of the proportion distribution between two adjacent periods, we are able to find anomaly changes. After comprehensively considering the feasibility and practicability, we have constructed an anomaly detecting scheme based on in/out traffic proportion, directly towards the significant targets on Internet.

Keyword Denial of service attacks, IP networks, legitimately abrupt change, In/Out traffic proportion, Generalized Likelihood Ratio

1. Introduction

Nowadays, the Internet is increasingly being used to conduct business and even to provide some critical services. On the other hand the DoS/DDoS attacks become the most prevalent threats against the widely used Internet. They can cause serious damage effects on those important Internet applications.

In February of 2000, a series of massive DoS attacks incapacitated several high-visibility Internet e-commerce sites, including Yahoo, Ebay, and E*trade. Then in January of 2001, Microsoft's name server infrastructure was disabled by a similar DoS assault. Many other domestic and foreign sites have also been victims, ranging from smaller commercial sites, to educational institutions, public chat servers and government organizations.

Simply to say, the DoS/DDoS attacks often attempt to disrupt an online service by generating a traffic overload to cause the victim to break down. In this case, the attacks would cause the abrupt changes in traffic. So many detecting mechanisms based on traffic statistics properties have been proposed.

However, most of these existent detection schemes are essentially based on abrupt changes of the unidirectional traffic. This detection base might result in serious false alarms when legitimately abrupt changes appear. To resolve this problem, we have proposed a heuristic detection scheme, and taken the bidirectional traffic into account.

In this heuristic scheme, we would mainly check the In/Out traffic proportion at the protected node's gateway or the router nearby. In normal cases, the proportion of In/Out traffic is close to a constant value[5]. Then we could adopt the likelihood ratio mechanism to test the likelihood of the proportion distribution between two adjacent periods. By inspecting the abnormally low likelihood, we are able to find anomaly changes.

2. Overview of DoS/DDoS attack and detection schemes

A DoS/DDoS attack is an explicit attempt by an attacker to prevent legitimate users of a service from using the desired resources. Attackers carry out DoS attacks by making a resource inoperative. They occupy large amount of shared resources, so that other users have no or little resources left. Though DoS/DDoS attackers do not damage data itself, they intentionally compromise the availability of the resource.

There are many kinds of DDoS attacks, such as Smurf attacks, UDP floods, reflection attacks, and SYN flood attacks.

In Smurf and UDP attacks, the attacker generates many ICMP or UDP packets to exhaust the capacity of the victim's network link. In reflection attacks, the attacker chooses a publicly available service or protocol, such as the Domain Name System, Web, or ping, and sends service requests to many such servers, forging the source address of the victim. Servers then reply back to the victim, and this flood of replies creates denial of service. In SYN Flood attacks, attackers send so many connection requests to one server that users cannot connect to that server. About 90% of all DoS attacks are SYN Flood attacks[1].

There are also many schemes proposed to detect and defense this kind of attack. In the ingress filtering[2], the internal router is configured to block packets that have source addresses from outside the internal network. However, this method cannot remove all attack packets because attack packets with addresses of internal network cannot be blocked.

SYN cache is the mechanism in server nodes. This mechanism has developed the original TCP protocol, and made the server node have more number of half-open states. By this way, the impact of SYN Flood attack could be mitigated. However, this mechanism can not entirely resolve the problem.

In the reference [3] the authors have proposed a statistics-based overload control scheme against DoS/DDoS. It estimates the legitimacy of a packet based on the packet score of its own attribute values. In the reference [4] they have also given an detection scheme against IP networks anomaly, whose key point is to apply some signal processing techniques to the unidirectional network traffic.

However, like many other defense mechanisms based on traffic changes, these two detecting proposals can not solve the problem when legitimately-abrupt-change cases occur.

On Internet, there are some special kinds of public services, whose servers may face the legitimately abrupt change in traffic parameter. This kind of services may be the famous information gateway website, e.g. Yahoo. When bombastic news announced, the web sever would receive much more connection requests than the normal time. In this case, it caused a legitimately abrupt change.

Or, for some special information announce center, e.g. the website of national meteorological agency. If a nature disaster like typhoon, earthquake, or tsunami is said to be coming, the website also might face a legitimately abrupt change in visiting traffics.

In these cases we just mentioned, most detecting scheme which are based on unidirectional traffic changes, would cause the serious false alarms when a legitimately abrupt change aroused.

Hence it is necessary to develop out a new scheme which may touch this problem well.

3. Heuristic DoS/DDoS attack detection scheme

3.1. Network Model of the Detection

In anomaly detection, it's always desired for us to find anomalies close to the attacker, rather than close to the victim, so that malicious packets can be stopped before they can cause any harm. Though this point of security consideration is reasonable and intuitive, but in most scenes of networks attacks, especially when a widely distributed DoS attack happens, it is very difficult for a safe mechanism to assure all networks elements cooperate very well, specially in a world wide range. So in our security scheme, we focus the attention on the protection of the significant objects.

The significant object may have two meanings:

(1) It's attractive enough to attack;

(2) It's valuable enough to protect.

After comprehensively considering the feasibility, practicability, and specialness, we construct an anomaly detecting scheme based on In/Out traffic proportion, directly towards the significant targets on Internet. The detection topology is shown as Fig. 1.

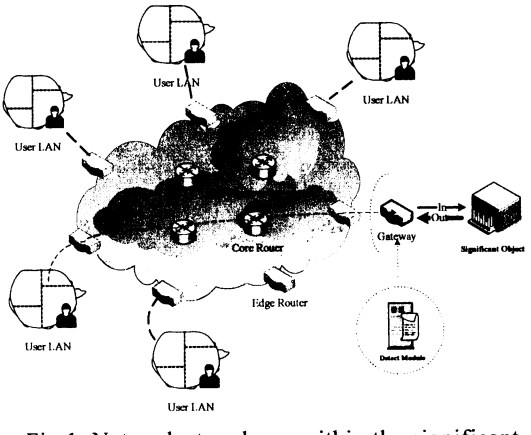

Fig 1. Networks topology, within the significant object, and anomaly detect module.

3.2. Detection scheme using In/Out proportion

Since there would be legitimately abrupt changes in the unidirectional traffics as mentioned in the section 2, we could consider the bidirectional case. Consequently we find if look at the proportion of coming **In** and going **Out** traffic of a network node or service, though the legitimate abrupt changes happen, the In/Out proportion would not change abruptly as well. In the legitimate use of networks, e.g. visiting the website of National Meteorological Agency, more the request packets come in, more the response and feed back packets go out.

So under normal circumstances, the ratio between the packet rate coming in and going out of a server, is close to some constant *k*[5].

But in DoS/DDoS attacks, two kinds of traffic increasing would not be synchronous. In DDoS attacks, the simplest form is merely to send a very large quantity of messages. Flooding attacks work by sending a vast number of messages whose processing consumes some key resource at the target[6]. Obviously for these two cases, the proportional properties of In/Out traffic rate have been destroyed.

So we find another way to detect out the malicious attacks, by checking the In/Out traffic rate proportion of the protected network node. As former analysis, by this kind of mechanism, the false alarms would be decreased.

4. GLR analysis using measured traffic

For an ISP networks showed in Fig. 1, all traffic towards the protected object would go through a gateway. The **In** traffic means that the destination address is the object node, and the **Out** traffic means the source address is this node. Let PR_{in} or PR_{out} be the traffic rate towards the object node or gateway, then the proportion *R* between them should be:

$$R = \mathrm{PR_{in}} / \mathrm{PR_{out}} \qquad (1)$$

As mentioned formerly, the *R* would be close to some constant *k*[5].

Calculate this kind of proportion between In/Out traffic data, and we will get a ratio time series. Since the abrupt changes in those attack processes are short-range dependent, so the proportion time series are also short-range dependent, and could assume it is a AR process with order 1[4].

Hence we could adopt the generalized likelihood ratio (GLR)[7] scheme to test two adjacent windows' likelihood, and then judge whether an attack happened or not. Two adjacent windows are Learning Window and Test Window, showed in Fig. 2.

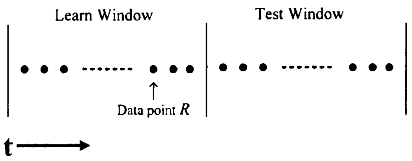

Fig. 2 Adjacent piecewise stationary windows: Learn window and Test window.

From Fig. 2, we can calculate GLR as follows:

$$S_L^2 = \frac{1}{W-1}\sum_{i=1}^{W}(R_{Li} - \bar{R}_L)^2 \qquad (2)$$

$$S_S^2 = \frac{1}{W-1}\sum_{i=1}^{W}(R_{Si} - \bar{R}_S)^2 \qquad (3)$$

$$S_P^2 = \frac{1}{2W-1}\sum_{i=1}^{2W}(R_{Pi} - \bar{R}_P)^2 \qquad (4)$$

$$\eta = \frac{(S_L \times S_S)^{-(W-1)}}{(S_L \times S_S)^{-(W-1)} + S_P^{-2(W-1)}} \qquad (5)$$

R_i : In/Out traffic rate proportion in the relative window;

$\bar{R}$: the mean value of data series in the relative window;

S_L, S_S : the sample variance of the residual in the learn window and test window;

S_P: the pooled sample variance of two adjacent windows

It make the case more conveniently for us to fix the learn window and test window as the same, without hurts to the generality. In equation (2) – (5), we set it as *W*, and then calculate out the generalized likelihood ratio η.

η is a value between 0 and 1. When η is closer to 1, the data distribution in test window is more likely to happen after the learn window. In another word, it is more likely to be anomaly when η is smaller then a preset threshold.

Furthermore, by fixing the length of the observing window, the security scheme could avoid the source exhausts itself.

5. Numerical example and discussion

To affirm the view of the heuristic detection scheme mention in former sections, we have checked the traffic between the Science Information Network (SINET)[8] and other two commercial Internet exchange service networks, JaPan Internet eXchange (JPIX) and JPNAP.

SINET is an information communication network dedicated to academic research. It connects nationwide connection points through high speed communication lines. SINET mutually connects with the Inter-Ministry Research Information Network (IMnet) and commercial Internet service providers to promote the international exchange of information as well as exchange of research data between the industrial, governmental, and academic sectors.

The data sets we've adopted here are showed as follows:

(1) Bitrate in 24 hours on 10 Gigabit Ethernet line of JPIX from 17:44 on May 03, 2005;

(2) Bitrate in 24 hours on Gigabit Ethernet line of JPIX from 13:06 on March 25, 2004;

(3) Bitrate in 4 hours on Gigabit Ethernet line of JPIX from 14:01 to 18:01on March 24, 2004;

(4) Bitrate in 24 hours on Gigabit Ethernet line of JPNAP from 17:44 on May 03, 2005.

The Fig. 3 shows the data set (1), which displays the bitrates distribution of the bidirectional traffic between SINET and JPIX.

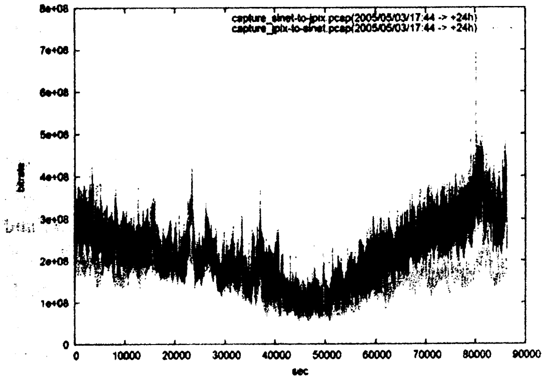

Fig. 3 The 24 hours' bitrates of the bidirectional traffic, between SINET and JPIX from 17:44 on May 03, 2005

In Fig. 3 we may see that for the unidirectional traffic, the bitrate changes frequently and somewhat abruptly. It's difficult for us to get the true statistic distribution of the real network traffic, and also difficult to detect out the anomaly only by checking the unidirectional traffic changes. So we will consider the bidirectional cases.

In this paper we look the traffic coming from JPIX or JPNAP to SINET as **In** traffic, and the traffic going from **SINET** as Out traffic. Then we could get the time series of In/Out bitrate proportion. This result shows that though the proportion would not change so much in short duration, but it would vary in a relatively big range from the whole view. This case is not so suitable for the anomaly detection. So we would check the likelihood between two adjacent windows.

Based on these proportion time series, we may get the generalized likelihood ratio sequence by applying equation (5). In the calculating process, the test window for this time would become the learn window in the next step. Fig. 4 shows the GLR sequence of the 1) bitrate set.

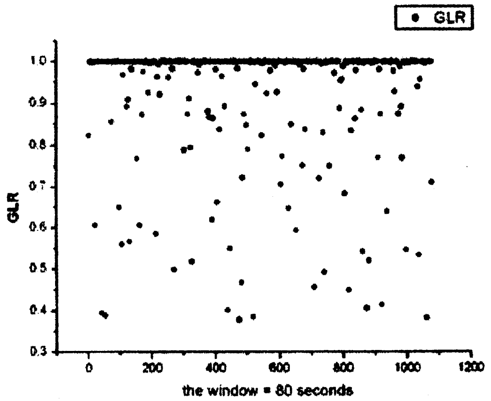

Fig. 4 The GLR sequence of the birtate proportion time series between JPIX and SINET.

In Fig.4 there is 1 day's bitrate data which has been obtained on May 03, 2005. The total sample points are 86398, and the learn and test window is both 80 points/seconds.

We've the similar result as showed in Fig. 4 for other three data sets. In fig. 4, we can find most GLR values are close to 1, and mostly above 0.8. This result has verified our assumption that the In/Out traffic proportion sequence of the normal traffic would not change abruptly. The percentage distribution of GLR shows in table 1.

Table 1. The percentage distribution of the GLR value

GLR Data set	(0. 8, 1]	(0. 7, 0. 8]	(0. 6, 0. 7]	(0. 5, 0. 6]	(0. 4, 0. 5]	(0. 3, 0. 4]	(0, 0. 3]
(1)	84. 2%	3. 8%	3. 2%	3. 9%	3. 8%	1. 1%	0%
(2)	82. 8%	3. 9%	4. 7%	3. 7%	4. 0%	0. 9%	0%
(3)	85. 5%	3. 8%	3. 5%	2. 8%	3. 9%	0. 5%	0%
(4)	84. 9%	3. 9%	3. 8%	3. 3%	3. 4%	0. 7%	0%

From Table 1 we may see that most GLR values are concentrating between 0.8 and 1, which is relatively close to 1. This means that the distribution of In/Out traffic proportion is most likely to its former one.

By picking out the abnormal low value of GLR, we then could check the time point where anomaly happened. By study the details of the packets' header information where anomaly appeared, we may then adopt other proposed defense mechanisms to defend or response to the attacks.

6. Conclusion

We've analyzed the properties of the coming in traffic and going out traffic proportion at the protected node's gateway or the router nearby. The result showed that the proportion of the In/Out traffic would not change abruptly in normal networks. Based on this result we've proposed a heuristic detection scheme directly towards the significant targets on Internet, by checking the generalized likelihood ratio of two adjacent proportion value windows. This would contribute the easiness and accuracy of the anomaly detection.

Reference

[1] D. Moore, G.M. Voelker, and S. Savage, "Inferring internet Denial-of-Service activity," Proceedings of the 2001 USENIX Security Symposium, pp.9–22, August 2001.

[2] P. Ferguson and D. Senie, "Network ingress filtering: Defeating denial of service attacks which employ IP source address spoofing." RFC 2267, January 1998.

[3] Y. Kim, W.C. Lau, M. Chuah, J. Chao, "PacketScore: Statistics-based Overload Control against Distributed Denial-of-Service attack", *IEEE Infocom,* March 2004.

[4] Chuanyi Ji Marina Thottan, "Anomaly Detection in IP Networks", IEEE Transactions on Signal Processing, 51(8):2191-2204, 2003.

[5] Gill, T. M., and Poletto, "M. MULTOPS: a data-structure for bandwidth attack detection". In USENIX Security Symposium (2001).

[6] Jelena Mirkovic, Sven Dietrich, David Dittrich, and Peter Reihter, "Internet Denial of Service Attack and Defense mechanisms", Prentice Hall, December 2004.

[7] H. L. V. Trees, Detection, Estimation, and Modulation Theory. New York: Wiley, 1971, vol. 1.

[8] Science Information Network: http://www.sinet.ad.jp