

Detecting Distributed DoS/Scanning by Anomaly Distribution of Packet Fields

Chang-Han Jong, Shih-Pyng Shieh

{chjong,ssp}@csie.nctu.edu.tw

Department of Computer Science and Information Engineering,
National Chiao-Tung University, Hsin-Chu, Taiwan

Abstract

To detect distributed denial of service and distributed scanning attacks, we propose Anomaly Dispersion Scheme (ADS). Observing the creation of packet fields in attack programs and normal programs, ADS monitors the distributions of packet fields, which alter when the attack programs using raw socket interface partake in, to detect intrusion. The sets of anomaly packet fields are attack signatures, which can be used to identify the attacks.

Keywords:

Distributed Denial of Service, Scanning, Network Intrusion Detection, Anomaly Detection

1 Introduction

Distributed DoS and scanning (DDoS/DS) are among the most serious problems in network security. DDoS sends numerous malicious packets from multiple hosts to disable the victim hosts [Yu 90][Dittrich]. DS collects network information including live hosts, open ports, and vulnerable services by multiple hosts for future intrusion. The attackers may use various ways to conceal themselves. Table 1 lists the techniques: source IP spoofing hides the origin of the attack packets; destination IP

spoofing conceals the true victims; TCP/IP protocol ambiguity makes the probing stealthy; inter-protocol scanning uses one protocol implementation, such as HTTP proxy, to scan other ports, such as finger port; besides attackers can perform attacks from multiple hosts so that tracing back is hard work.

Attack-Hiding Technique	Description
Source IP Spoofing	Source addresses may not belong to the attack hosts
Destination IP Spoofing	Destination addresses may not belong to the victim hosts by assigning small TTL
TCP/IP Protocol Ambiguity	For packets not well-defined in specification, implementation may response in wrong way or have no response
Inter-Protocol Scanning	Using one protocol implementation to scan other ports
Multiple Host	Using thousands of hosts to attack

Table 1: Attack-Hiding Technique

DDoS/DS remain traces on packet fields so that we can use for detection. For DDoS, fields of address and port field sprawl; for DS, packets with protocol vulnerability of design or implementation may be used to elude detection; Layer 3/4 attacks accompanied with DDoS/DS use special value of packet fields to perform attack. Moreover, tracing the attack programs, we found that attack programs, which use raw socket interface, create network packets whose fields are one of the following three kinds-1) fixed value, 2) created by random function, and 3) created by a specified function. Table 2 is the sample code fragments of the three types. The fields of attack packets are different from ones of the normal traffic, because the attack packets are often created by raw socket instead of by the TCP/IP protocol stacks embedded in the operation system. Therefore, the distribution of network packet fields may be used to detect DDoS/DS if DDoS/DS attack programs actually behave differently in packet fields with other programs.

Generating of packet fields	Sample code fragment
Fixed value	<code>*((u_short*)p_ptr)=htons(242); /*IP ID*/</code>
Random function	<code>ih.ip_id=htons(random()); /*IP ID*/</code>
Certain function	<code>ih.ip_sec.s_addr=k00kip(); /*IP Src Addr*/</code>

Table 2: Packet Fields Generating of Attack Programs

Various related work have been done to detect DoS or scanning. GrIDS detects rapid malicious network activities by modeling the network activities [Staniford 95]. Packet aggregation watches the ICMP messages to detect failure of networks [Kanamaru 00]. SPICE detects scanning by the entropy concept and a correlation engine [Standiford 00]. IP addresses are not trustful if no authentication is applied. GrIDS and packet aggregation are fooled by the spoofed IP addresses. In addition, GrIDS computes the graph search, whose complexity grows exponentially with the size of networks. SPICE has now only been proved functional in detection scanning.

We proposed the Anomaly Dispersion Scheme (ADS), which detects DDoS/DS attacks on the assumption of abnormal distribution of packet fields caused by DDoS/DS attacks. ADS doesn't consider IP addresses as a trustful identifier of packet. Moreover, ADS assumes every packet value can be forged, but nevertheless when forged value distribution is far from the normal distribution, intrusion is detected. It's why ADS detects distribution DoS and scanning.

This paper is organized as follows. In Section 2, we present an Anomaly Dispersion Scheme for detecting DDoS/DS attacks. In Section 3, we show the experiments and discussion. And finally, Section 4 gives the conclusion.

2 Anomaly Dispersion Scheme

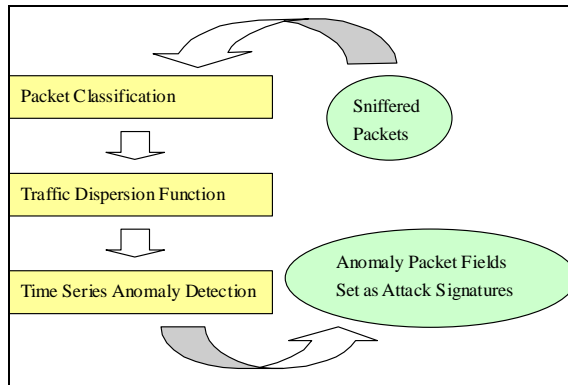


Figure 1: ADS

ADS is composed of three components- 1) packet classification, 2) traffic dispersion function (TDF), and 3) time series anomaly detection (TSA). Packet classification and TDF are like the D box of CIDF while TSA is as the A box. For a set of packet fields, or also called a classification way, ADS gives a digest function to map the fields in to a real value. A classification way is anomalous if that distribution of the corresponding packet fields is anomalous. Concurrently, ADS monitors multiple classification ways to detect anomaly distribution of packet fields caused by DDoS/DS attack packets. In a period, ADS report the set of the anomalous classification ways as the attack signatures.

When ADS receives a packet, packet classification is performed, where each packet is calculated by every packet digest functions of the classification ways. The results show packet counts of different classes, under certain period and classification way. And periodically, ADS performs TDF to transform the packet counts of a classification way into a time series of the network traffic characteristic. TSA then checks if the time series biases too much to be regarded as intrusion. To sum up, ADS detects network attacks on the point of view of packet field distribution and then transforms the statistics to temporal characteristic and then to time series processing.

```

1. //*****ADS*****
2. When packets come, performing Packet Classification
3.     For each c, belongs C, which is a predefined set of classification ways
4.         bitstring digest=PacketDigest(packet p ,c)
5.         // t is the current time period
6.         Dc,t,digest++
7.
8. Periodically, performing Traffic Dispersion Function and Time Series Anomaly Detection
9.     S={ }
10.    For classification way c, belongs to C
11.        //Mc,t refers to a set of Dc,t,i
12.        //tn means the current time period
13.        Time Series TS={TDF(Mc,t0), TDF(Mc,t1), TDF(Mc,t2),..., TDF(Mc,tn)}
14.        If TSA(TS,K)=ANOMOLIOUS then S=SU{c}
15.        Report S as attack signature
16.
17. //*****SUB ROUTINES*****
18. PacketDigest(packet p, classification way c){
19.     bitstring digest, partial_digest
20.     for each rp, belongs to all properties of p
21.         //bit_reduction uses group/aproximated/hash
22.         partial_digest =bit_reduction(rp)
23.         digest =concat(digest, partial_digest)
24.     return digest
25. }
26. TDF(set X){
27.     Run at first time period
28.     C1 is given a large number
29.     While G(X) is less than given threshold T,
30.         C1=C1/2
31.     C3=1;C2=0
32.     Run at the 3 time period
33.     C3= 2/(G(X)2ndc_time_period+G(X)1st_time_period)
34.
35.     G(X)=C3* (C1*Log(x+1)+C2), where x belongs to X
36.     return G(X)
37. }
38. TSA(Time Series TS){
39.     TS={s1,s2,s3,...sn}
40.     DTS={s2-s1,s3-s2,...,sn-sn-1 }={m1,m2,m3,...mn-1}
41.     if mn> C4*stddev(m1,m2,...mn) then
42.         return ANOMOLIOUS
43.     else
44.         return NORMAL
45. }

```

Table 3: Pseudo Code

2.1 Packet Classification

ADS receives packets from routers or via sniffing, and then classifies them in multiple classification ways. Two packets are regarded as the same class, if the results of packet digest function of a classification way are the same. For packet digest

functions, there are four kinds of input, including 1) packet fields, 2) length of the header and packet, 3) integrity and validity of the header and packet, 4) true properties (true packet length, not the recorded one).

If the maximum possible class number of a classification way is too large, ADS may fail because of exhausting search. Reports show that DDoS packets with random spoofed source IP addresses make ATM switches exhausted in routing table searching. To reduce the maximum possible class number, bit reduction should be performed. There are three kinds of bit reduction techniques can be optionally adapted, 1) group mapping, 2) approximated mapping, and 3) hash mapping. Group mapping reduces the output bit according the semantics of the packet fields. For example, the network ID represents the full 32bits IPv4 address and ports are separated into well-known and large-than-1024 ones. Approximated mapping, such as dividing or modulation, is used when the packet fields represent the length. TCP SYN/ACK sequence numbers are also feasible to be used in approximated because a TCP session is supposed to use sequential numbers. Hash grouping, such as MD5 algorithm or simple modulation is used.

2.2 Traffic Dispersion Function

For a give period and classification way, this component produces a real value to present the network characteristic. The values should express the composition of the packets so that DDoS/DS can be detected. The required properties of traffic dispersion function are 1) aggregative, 2) insensitive, and 3) not over-coverage.

Aggregative is that TDF reflects to the whole change but not fractions change. Insensitive is that TDF correlates to the quantities, not only the ratio. Insensitive resists from that the attackers to make the ejected attack packet with the same

composition of the normal traffic. Over-coverage is that extreme quantity of certain class would not affect value of TDF too much. Not over-coverage is to reduce the extreme big or small quantities of few classes

There are extreme examples of TDF that are over-coverage. An IDS uses $f(x)=1/\text{Pr}(x)$ as the malicious function, where $\text{Pr}(x)$ is the probability of the event. In this case, if $\text{Pr}(x)$ is small, then $f(x)$ is significant large. When normal events are comparably with low probability, it may cause the IDS to have high false alarm rate. Another case is the second-order movement, $f(x)=x^2$, which have extreme large value when the packet count of certain classes are high.

If a function $g()$ is positive, increasing and its order is between $(0,1)$, then

$G(X)=g(x)$, where x belongs to X , is a TDF function that satisfies aggregative, insensitive, and not over-coverage. We propose a TDF $G(X)$ based on the above description.

$$G(X)=C3 * (C1 * \text{Log}(x+1) + C2)$$

The order of $\text{Log}(x)$ is between $(0,1)$, increasing and positive when $x > 1$. So $x+1$ adjusted the function to satisfy the require properties. $C1$ controls the order and can be modified to adapt different environments. $C2$ gives weights those classes that have zero packet count. $C3$ normalizes the output of $G(X)$, and is for privacy issue when opening the use of statistics in to public,

2.3 Time Series Anomaly Detection

The differencing technique is used to reduce seasonal effect and trend in time series analysis [Chatfield 89]. First-order differencing transforms a time series $\{x_1, \dots, x_n\}$ to another time series $\{y_1, \dots, y_{n-1}\}$ by performing $y_t = x_{t+1} - x_t$. This technique has also been

used in detection of VoIP traffic anomaly [Mandjes 00]. Therefore this component uses this technique to process the time series produced from the traffic dispersion function component. Denning's Mean and Standard Deviation Model variation with differencing is used in this component for simplicity [Denning 86]. The physical meaning is that, rather than forecasting, the concept of variance tolerance is used to decide anomaly because we may not know how packet fields are used.

3 Experiments and Discussion

We have implemented a prototype based on libpcap library and FreeBSD 4.4. Samples are tcpdump packet captures from live network traffic, without or with injected attack. The traffic is captured on the gateway of a campus class C LAN with over 200 computers inside. Injected attacks are real-time performed by three famous DDoS/DS tools, stacheldraht, nmap and ping. Each sample lasts 60 seconds. The injected attacks start at the 30th second. When processing, ADS skips the first 6 seconds because captured packets bursts in the first few seconds due to libpcap initialization. The detection windows size is 8 and will be discuss later.

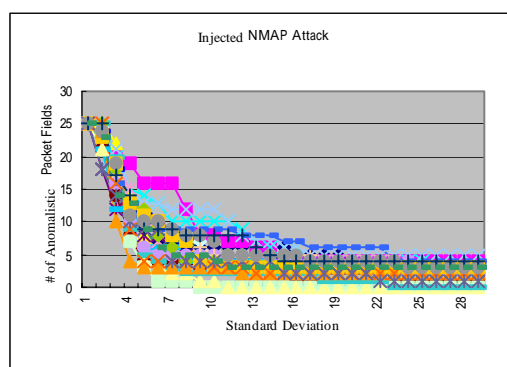


Figure 2: With Injected NMAP attacks

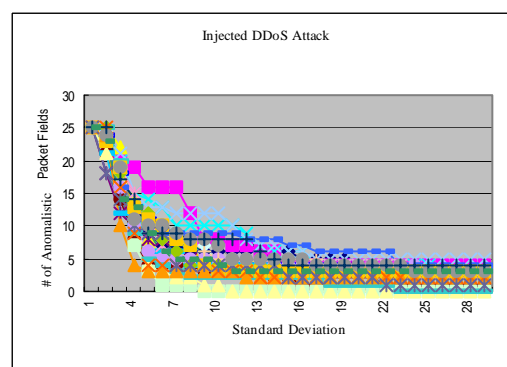


Figure 3: With Injected DDoS attacks

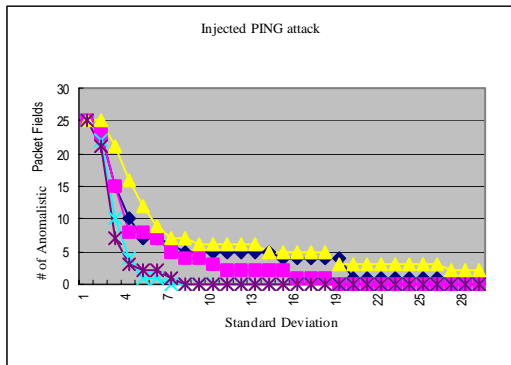


Figure 4: With Injected PING attacks

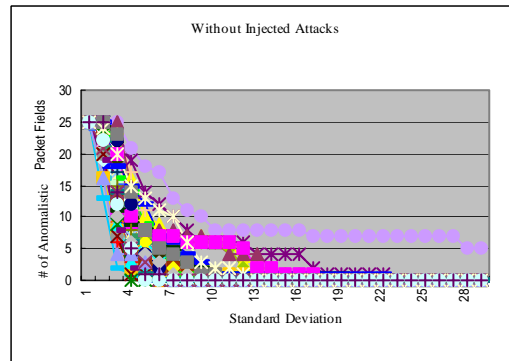


Figure 5: Without Injected Attacks

Figure 2-5 shows the number of anomalous packet fields with and without injected DDoS/DS attacks. X-axis is the tunable C4 threshold, which specifies how statistics biased is treated as anomalous. Samples with injected attacks, Figure 2-4, tend to have more fields anomalous on the same threshold than samples without injected attacks, Figure 5.

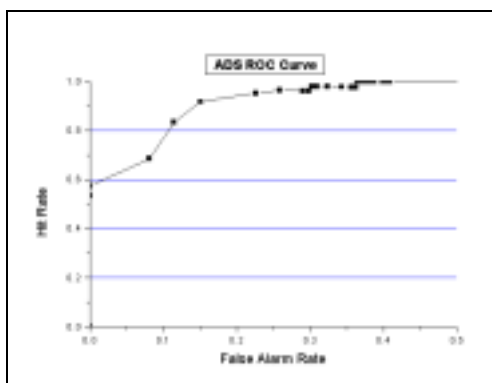


Figure 6: ROC Curve of ADS

The throughput of the ADS prototype with 27 classification ways is 102K Packets/second. Figure 6 shows the R.O.C curve of the ADS. ROC is an IDS performance evaluation graph, where the X-axis is the false alarm rate (false alarms/negative samples) and Y-axis is the hit rate (correct alarm sample/total alarm samples). This testing flow can be used to discover novel attacks. During the experiments, one type of novel network attack based on Microsoft Windows

Universal PnP protocol was discovered.

3.1.1 Anomaly Distribution of TCP/IP Packet Fields

ADS alarms for network attacks as attack signatures. We conclude the reasons why certain fields are anomalistic, which helps security officers to identify the why the packet anomaly. The following table shows why certain fields present anomaly. IP protocol fields set to 255 may be used to elude detection. IP ID generated by various algorithms, such as sequential counter, MD4, and MD5, would have different distribution than fixed valued. IP Offset field is anomalistic when small fragment packets for dissimulating is used or MTU is not suitable. IP and TCP Options field is used for information discovery for a long time. TCP SEQ/ACK field is supposed to be sequential or near sequential. If the attack program uses fixed value on these fields, then anomaly is predictable. OS fingerprinting is performed by examining the response of TCP/IP protocol stacks under undefined or ambiguous parts of the protocols. Therefore rare field values are presented in the OS fingerprinting packets. For attackers to increase the accuracy of OS fingerprinting, they don't use only a few packet fields, OS fingerprinting reveals anomaly in almost all fields.

IP Protocol	Stacheldraht tool sets this field to fixed value 255, which is the reserved value so that the author of the program may want to elude somehow. IP packets, whose protocol not equal to TCP/UDP, may be combined used with IP options to perform scanning.
IP ID	Older Un*x system use a sequential counter to generate IP ID fields; Linux uses MD4 algorithm; FreeBSD uses old fashion sequential counter or MD5 algorithm. So the IP ID field easily presents anomaly in attacking if the distribution of IP ID fields
IP Offset	When a packet is fragmented, IP offset field records the offset of spited packets related to the original packet. If the attack program wants to use tiny fragment packet to hide its activity, the IP offset field will be anomalistic. Fragmentation also happens when the MTU of a network in the routing path is small then the MTU of other network in the routing path.

IP Options	IP Options can be use to get information about routing and time or to use source routing to detour the firewall.
ICMP	Attack programs may use ICMP to detect if the target host alive. ICMP are used to test if the target host alive before scanning. Therefore, ICMP packets reveal another kind of attack signature other than the main attack packets. But ICMP happens too frequent, so the threshold of this fields anomaly should be independent set.
TCP/UDP Port	By Stantiford's observation, Port scanning is to gather enough information from the port. Therefore TCP or UDP port field distribution change may probably be caused by scanning attack.
TCP SEQ/ACK #	If the attack programs send special TCP packets to scan or DoS, the TCP SEQ/ACK # fields change enormously because TCP connection semantics are different from the normal ones. When the TCP connections are anomalistic, TCP SEQ/ACK number may have abnormal distribution. The normal semantics of TCP SEQ/ACK # is sequential, and if the attack programs uses fixed value, anomaly is detected.
TCP Options	Window scale factor may be use to on degrading the performance of TCP sliding window. TCP options can also be used for gathering information just as IP options.
OS Fingerprint	OS fingerprinting sends a series of TCP/IP packets and receives the responses of the target hosts. In the experiments we saw that various fields of the packets anomaly because of the scanning and response packets.

Table 6: Reasons of Packet Fields Anomaly Distribution

3.1.2 Detection Window Size versus Seasonal Effects

Detection window size is the size of time series in the TSA. The detection window size affects the accuracy of the IDS. We observed that if periodically and the detection window size is smaller than the cycle, false alarm rate will be high because the periodical network activities may be impulse. For example, Microsoft windows operation system broadcasts about every ten seconds.

A time series $\{X_n\}$, has a periodical action in cycle L . Without the loss of generality, the time series is denoted $\{m_1, m_2, m_3, \dots, m_L, m_{L+1}, m_{L+2}, \dots, m_{2L}, \dots\}$. Broadcast packets

are on m_{iL} , i is positive integer. When the detection windows $<L$, the samples may be between m_{iL} and $m_{(i+1)L}$ and doesn't contain any m_{iL} . If m_{iL} is much larger than m_{jL+k} , where $0 < k < L$, forecasting according the detection windows lapses. If the detection windows size is large or equal than L , one of the m_{iL} will be included in the forecasting, false alarm rate reduces.

4 Conclusion

ADS derives from tracing and comparing the TCP/IP protocol stacks and DDoS/DS attack programs. The idea comes from that packets created by different programs differ in packet fields especially between native TCP/IP embedded the OS and DDoS/DS attack programs using raw socket interface. The R.O.C curves and the throughput of the prototype show that ADS is practicable. ADS can be easily extend its ability for monitoring more fields by only adding packet digest function for the packet fields that we are interested in.

Because ADS uses primitive Time Series Anomaly Detection and no correlation engine or probability decision mechanism in it, false alarm rate is still high. We would like to replace add these features to ADS in the future.

Reference

[Chatfield 89] C. Chatfield, "The Analysis of Time Series-An Introduction 3rd Edition," page 21, 1989

[Denning 86] Denning, "An Intrusion Detection Model," IEEE Trans. on Software Engineering 1986

[Dittrich] Dave Dittrich, Distributed Denial of Service (DDoS) Attacks/tools homepage, <http://staff.washington.edu/dittrich/misc/ddos/>

[Kanamaru 00] Kanamaru, "A Simple Packet Aggregation Technique for Fault

Detection,” *Int. Journal of Network*, 2000

[Mandjes 00] M. Mandjes, I. Saniee, and A. Stolyar, “Load characterization, overload prediction, and load anomaly detection for voice over IP traffic,” *Proceedings 38th Allerton Conference, Urbana-Champaign, US*, pp. 567-576.

[Staniford 95] S. Staniford-Chen, S. Cheung, R. Crawford, M. Dilger, J. Frank, J. Hoagland, K. Levitt, C. Wee, R. Yip, D. Zerkle, “GrIDS-A Graph Based Intrusion Detection System for Large Networks,” *National Information Systems Security Conference*, 1996

[Staniford 00], Stuart Staniford, James A. Hoagland, Joseph M. McAlerney, “Practical Automated Detection of Stealthy Portscans,” *ACM Workshop on IDS*, 2000

[Yu 90] Che-Fn Yu, Virgil D. Gligor, “A Specification and Verification Method for Preventing Denial of Service,” *IEEE Trans. on Software Engineering*, Vol 16, No 6, June 1990