IP Trace Back Techniques to Ferret out Denial of Service Attack Source

ADNAN AIJAZ, SYED RAZA MOHSIN and MOFASSIR-UL-HAQUE
Electrical Engineering Department
National University of Sciences and Technology (NUST)
Military College of Signals, Rawalpindi.
PAKISTAN

Abstract: - Today's life has been revolutionized by Internet. Future of Internet is even more promising because of emerging technologies like ubiquitous computing, context sensitive, adaptive and reconfigurable applications. Security is the most important issue concerned with Internet. Internet is exposed to threats like system penetration, financial fraud, theft of proprietary information, Denial of Service (DoS)/Distributed Denial of Service (DDoS) Attacks etc. Currently, DoS/DDoS attacks are the most expensive computer crimes. The attacker launching this type of attacks commonly masks his identity using IP spoofing. It is very difficult to identify the sources of a DoS/DDoS attack. IP trace back methods are used to locate the source of packet causing DoS/DDoS. In this paper we have carried out a survey of different existing trace back schemes highlighting their advantages and disadvantages. We also suggest our own hybrid IP traceback technique.

Key-Words: - Internet, Security, IP trace back, DoS, DDoS, IP spoofing

1. Introduction

The Internet is the basis of number of innovative technologies like the World Wide Web, Email, P2P applications, VOIP etc. It has enabled instant access to vast and diverse resources. But, Internet is also vulnerable to number of attacks from different sources. Major categories of attacks during 2006 were viruses, insider abuse of access, unauthorized access to information, and denial of service (DoS) attack [1]. It is often much easier to disrupt the operation of a network or system than to actually gain access to a network. There are number of freely available tools on Internet, from covertly exchanged exploit programs to publicly released vulnerability assessment software, to degrade performance or even disable vital network services [2].

1.1 DoS/DDoS

The aim of DoS attack is to prevent legitimate users access to system resources by shutting down or seriously slowing down a service provided by a computer system. DoS first received large scale public attention in February 2000 when major Internet sites including CNN, Yahoo, e bay and Amazon were brought down by DoS attacks. CNN and other victims claimed that the attack caused damages totaling $1.7 billion [3].

DoS attacks are classified as either flooding or logic attacks. In flooding attack the victim is overloaded with a large amount of traffic thus consuming resources. Example of flooding attack is the TCP/SYN flooding. Logic attacks are however, based on exploiting the vulnerabilities in the target system and can be carried out even with a single well crafted packet [4]. Example of logic attack is the LAND attack.

In distributed DoS (DDoS) attack, the attacker uses hundreds or thousands of compromised hosts, often residing on different networks, to overload and crash target system [5]. Currently, it is not possible to prevent DoS/DDoS attacks because they are based on exploiting weaknesses in the core internet protocols which are embedded in the underlying network technology.

In DoS/DDoS attack, attacker uses fake source IP addresses to make tracing and stopping of DoS difficult. This technique is called IP spoofing. This technique involves the manipulation of the source IP address in the IP header of a transmitted packet. This gives the attacker a form of anonymity. It is difficult to solve problem of IP Spoofing because of lack of security features in TCP/IP specifications. Ingress filtering, Use of cryptographic authentication , IP trace back are some of the approaches used to handle forged IP source addresses[6]. The purpose of IP traceback is to identify the true IP address of a host originating attack packets. IP trace back is vital for quickly restoring normal network functionality and preventing reoccurrences [7].

2. Classification Of IP Traceback Techniques

Existing trace back schemes can be roughly categorized into three distinct categories: traditional, marking and logging. In traditional scheme, victim develops an attack signature, consisting of some data common and unique to the attack traffic. A query including the attack signature is then sent hop-by-hop to each router along the path. Examples of this type of technique are input debugging and controlled flooding. In packet logging, the IP packet is logged at each router through which it passes. Routers are queried in order to reconstruct the network path. SPIE (Source Path Isolation Engine) is an example of this type of technique. In packet marking, the router marks IP packets with its identification information. The network path can be reconstructed by combining packets containing marks. The marking information may be inscribed in the same attack packets called inbound marking or extra ICMP packets called outbound marking. Current traceback schemes based on marking include variants of PPM (Probabilistic Packet Marking), ATA (Algebraic Based Traceback Approach), DPM (Deterministic Packet Marking), and schemes that use ICMP (Internet Control Message) messages, such as iTrace [8].

3. Traditional Approach

Traditional approach also referred as link-testing methods or hop-by-hop tracing work by testing network links between routers to determine the origin of the attacker’s traffic.

3.1 Input Debugging

Input debugging start from the router closest to the victim and interactively tests its incoming (upstream) links to determine which one carries the attack traffic. This process repeats recursively on the upstream routers until reaching the traffic’s source.

3.2 Controlled Flooding

Controlled flooding works by generating a burst of network traffic from the victim’s network to the upstream network segments and observing how this intentionally generated flood affects the attack traffic’s intensity. From changes in the attack traffic’s frequency and intensity, the victim can deduce the incoming network link on the upstream router and repeat the same process on the router one level above.

Traditional approach is not very efficient as it requires a lot of human effort and other network providers’ support. For a successful trace, attack must last far a long enough duration. Compatibility with existing protocols, routers and network infrastructure is an advantage of this method [9].

4. Logging

The main idea of IP traceback approach based on packet logging is to log packets at each router through which they pass.

4.1 SPIE (Source Path Isolation Engine)

SPIE stores packet digests, instead of packets themselves, in a space-efficient data structure, to decrease the required storage space. For each arriving packet, the router uses the first 24 invariant byte of the packet (20-byte IP header with 4 bytes masked out plus the first 8 bytes of payload) as input to the digesting function. The 32-bit packet digest is stored into the time-stamped digest table. The digest table is paged out before it becomes saturated. Digest tables are archived for one minute for potential traceback operation. During the traceback process, routers are queried in the reverse-path flooding (RPF) manner and the digest tables at queried routers are examined to reconstruct the network path [10].

Logging approach is resource-intensive in terms of processing and storage requirements. This scheme is not scalable. It is difficult to extend this scheme to complete Internet. Sharing of the logging information can lead to logistic and legal issues. Using Hash based IP traceback can reduce the storage overhead significantly [11].

5. Packet Marking

The basic idea of IP traceback approach based on packet marking is that the router marks packets with its identification information as they pass through that router.

5.1 PPM (Probabilistic Packet Marking)

In Probabilistic Packet Marking the mark overloads a rarely used field in IP packet header, i.e., 16-bit IP identification field. The identification of a router could be 32-bit IP address, hash value of IP address, or uniquely assigned number. In the last two cases, the length of identification information is variable and could be less than 16 bits. Since the marking space in packet header is too small to record the entire path, routers mark packets with some probability so that each marked packet carries the information of one node in the path. In addition, based on the length of router identification and the implementation of marking procedure, the router may only write part of its identification information into the marking space. While each marked packet represents only a small portion of the path it has traversed, the whole network path can be reconstructed by combining a modest number of such packets [12].

The PPM approach does not incur any storage overhead at routers and the marking procedure (a write and checksum update) can be easily and efficiently executed at current routers. But due to its probabilistic nature, it can only trace the traffic that consists of a large volume of packets. However, this method increases the packet's length at each router hop and can lead to additional fragmentation [9].

5.2 DPM (Deterministic Packet Marking)

In DPM only ingress edge routers perform the marking. All other routers are exempted from the marking task. Basic DPM uses the 16-bit IP identification field of the IP header and one reserved bit to record the marking information. The IP address of every ingress edge router is split into two segments with 16 bits each. One segment is randomly selected when a packet traverses this router. The idea is that the victim is capable of recovering the whole IP address of an ingress edge router once it obtains both segments from the same router. For the victim to figure out which portion of the IP address the current packet carries, one bit is used as a flag. Therefore, the marking information comprises two parts, the 16-bit partial IP address of the edge router and a 1-bit flag [11].

There are two main differences between DPM and PPM. DPM only marks the first ingress edge router, while PPM marks all routers along an attack path. PPM marks probabilistically, while DPM marks every packet at the ingress edge router. The task of ingress address reconstruction in DPM is much simpler than the task of path reconstruction in PPM.

5.3 ICMP (Internet Control Message Protocol)

In ICMP trace back method, iTrace, each router selects one packet per 20,000 packets and then generates an ICMP message. The ICMP message has the same destination IP address as the traced packet. The ICMP message also contains the IP header of the traced packet, and the IP addresses of the incoming interface and the outgoing interface of the current router. As long as the victim receives sufficient ICMP messages, it may recover the whole attack path.

The marking procedure of iTrace is very similar to PPM. Therefore, it shares similar pros and cons. Unlike PPM, ICMP traceback belongs to outbound marking, because of which ICMP traceback requires additional bandwidth to convey the marking information [11].

5.4 ATA (Algebraic Based Traceback Approach)

ATA is a modified PPM method that uses algebraic techniques from the fields of coding theory and machine learning to encode and decode path information as points on polynomials. The encoded path information is stored in the Fragment ID field. At the victim side, algebraic methods are used to reconstruct the polynomials [13].

6. A Hybrid IP Traceback Technique Based On TTL Identification

The goal of all the traceback techniques is to identify the sources of attacking traffic, but the reconstruction of an attack path actually reveals the identity of the first router on the path. A better approach would be to find an algorithm that reveals the identity of first router without requiring the participation of all the routers on the path [8].

Since the attacker can forge any field in the IP header, he can't falsify the Time to live (TTL) field. The TTL is an 8-bit field that determines the maximum number of hops a datagram can traverse. Each router decrements the TTL value by 1, after forwarding the datagram. The problem of determining the first router on the path can be solved by using this field.

The TTL field is different for different operating systems and is not universally selected, but all the packets sent by a particular operating system will have the same initial TTL value [14]. Default TTL values for different operating systems are shown in Table 1.

The basic idea behind our technique is to create a TTL vs. operating system table and store it on the routers. The matching of a TTL value with any entry of the table is indicative of the fact that this is the first router on the path. The router should then mark the packet with its IP address. For marking purpose, we select the variable length options field in the IP header. The router should in fact overwrite the first 4 bytes of options field with its IP address (since IP address is 32 bits long for IPv4). This overwriting is of no harm since the first 4 bytes would always contain the address of the first router in case any of 'Record Route', 'Strict Source Route' or 'Loose Source Route' options is present. So even if the attacker forges the options field with wrong IP address or unnecessary data, it would still be overwritten with the true IP address of the router.

The flooding DoS attack uses IP spoofing. The problem of this Source address spoofing can be solved by a technique called Ingress Filtering [15], in which the router discards the packets with illegitimate source addresses. The legitimacy of source address can be checked from the network id part of the IP address. A serious limitation of this technique arises when the attacker forges the address to the one that belongs to the same network as the attacker's host.

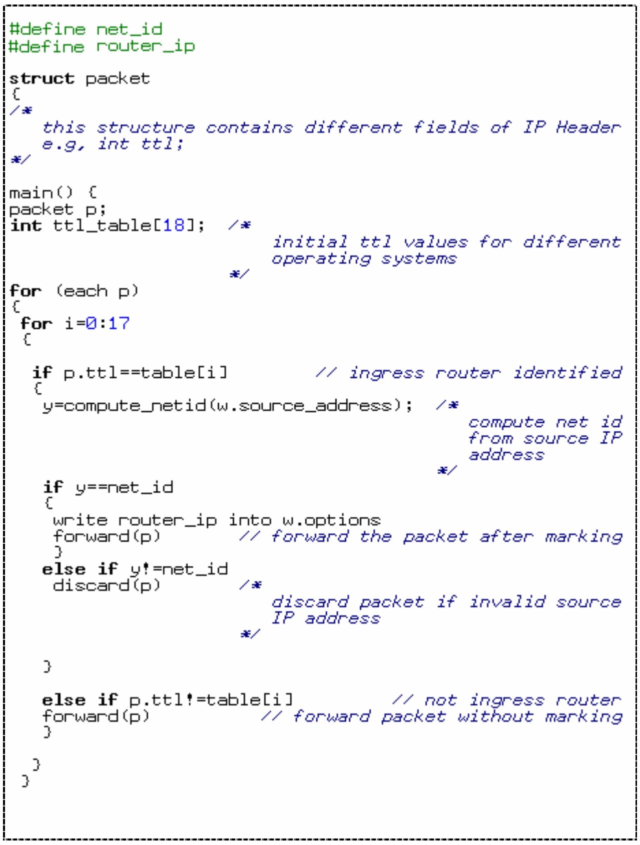
Table 1: Default initial TTL values for different operating systems [14]

OS	Version	Platform	TTL
Windows	9x/NT	Intel	32
Windows	9x/NT	Intel	128
Windows	2000	Intel	128
Digital Unix	4.0	Alpha	60
Unisys	x	Mainframe	64
Linux	2.2.x	Intel	64
FTX (UNIX)	3.3	STRATUS	64
SCO	R5	Compaq	64
Netware	4.11	Intel	128
AIX	4.3.x	IBM / RS6000	60
AIX	4.2.x	IBM / RS6000	60
Cisco	11.2	7507	60
Cisco	12.0	2514	255
IRIX	6.x	SGI	60
Free BSD	3.x	Intel	64
Open BSD	2.x	Intel	64
Solaris	8	Intel / Sparc	64
Solaris	2.x	Intel / Sparc	255

A more effective solution for IP Traceback is to combine Ingress Filtering with this variant of packet marking (based on TTL identification). The packet is first checked for spoofing and is discarded if the source address is forged. If the source address is valid, the packet is marked with router's identity. This would obviously reduce the marking overhead of the router. This technique also requires minimum storage requirements at the routers and the present day routers can efficiently execute this marking procedure. The algorithm for this hybrid technique is shown in figure 1.

This hybrid technique is significantly different from the basic DPM as described in section 5.2 in following respects.

- ➢ Reduced marking overhead due to ingress filtering.
- ➢ No need of address reconstruction.
- ➢ Faster convergence.
- ➢ Usability of identification field is retained for fragmentation purposes.
- ➢ Reliable approach to identify the ingress router.



```
#define net_id
#define router_ip

struct packet
{
/*
   this structure contains different fields of IP Header
   e.g, int ttl;
*/

main() {
packet p;
int ttl_table[18];  /*
                        initial ttl values for different
                        operating systems
                     */
for (each p)
{
 for i=0:17
 {

  if p.ttl==table[i]        // ingress router identified
  {
   y=compute_netid(w.source_address);  /*
                                          compute net id
                                          from source IP
                                          address
                                       */
   if y==net_id
   {
    write router_ip into w.options
    forward(p)        // forward the packet after marking
    }
   else if y!=net_id
    discard(p)        /*
                         discard packet if invalid source
                         IP address
                      */


   }

   else if p.ttl!=table[i]         // not ingress router
   forward(p)           // forward packet without marking
   }

  }
 }
```

Fig. 1: Algorithm for the hybrid IP Traceback technique based on TTL identification

7. Simulations

The above mentioned algorithm has been implemented at the IP (network layer) level in the network simulator-2 (ns-2). The IP protocol in the ns-2 simulation model was modified to incorporate marking in it and the resulting scenario was recompiled to make the changes active. The size of topology, on which this modified protocol was tested, really does not matter as the amount of delay and overhead, which this protocol is causing, is incurred at the very first router on the packet's path. The function of all other routers except the ingress router remains the same.

The network shown in figure 2 was simulated in the normal scenario. The delay and header overhead was recorded. Then the IP protocol was modified and recompiled. Simulating the same network after the incorporation of proposed marking has resulted in the overhead and delay as shown in the simulation results below. The overhead depends on the type of application or service used, for example, FTP application with 1040 bytes, as the packet size of service, is used in the simulation. The graph in figure 3 shows delay and overhead comparison of the normal and marked traffic. This shows the efficiency of this technique.

8. Conclusion

The development of IP traceback techniques is motivated by different DoS attacks in recent years. However, IP traceback is the first step in identifying the attacker behind the attacks. The effectiveness of any traceback technique depends primarily on its overhead, convergence and the ability to trace any type of DoS attack. The hybrid technique presented here is capable of tracing any type of DoS attack because we can trace even a single packet. Today there is a need of practical implementation of an effective technique so that IP traceback could be carried out in real time across the internet.

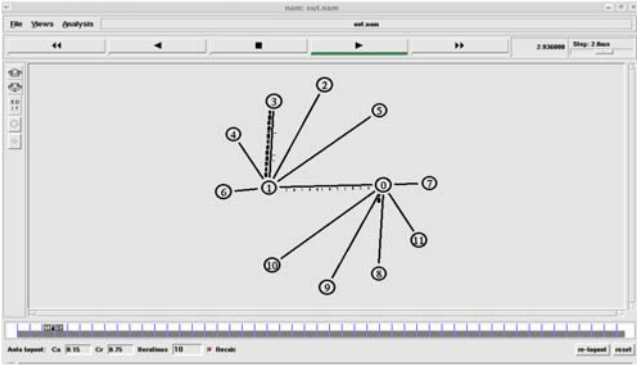


Figure 2: Simulated Topology

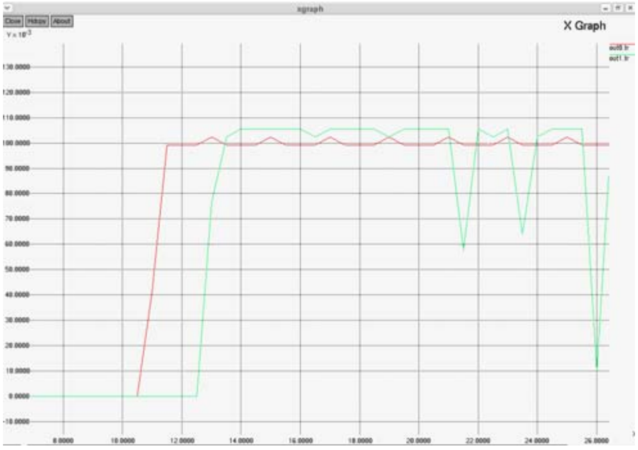


Figure 3: Graph showing delay for the marked and normal traffic.

References

[1] Lawrence A. Gordon, Martin P. Loeb, William Lucyshyn and Robert Richardson, CSI/FBI Computer Crime and Security Survey, 2006.

[2] Computer Emergency Response Team. CERT advisory CA-1999- 17 denial of service tools. http://www.cert.org/advisories/CA-1999-17.html.

[3] Computer Emergency Response Team. CERT advisory CA-2000- 01: denial of service developments. http://www.cert.org/advisories/CA-2000-01.html, 2000.

[4] D.Moore, G. M. Voelker and S. Savage, "*Inferring Internet Denial of Service activity*" ACM Trans. Comput. Syst., vol. 24, no. 2, pp. 115-139, May 2006.

[5] Matthew Hutchinson, "*Study of Denial of Service*", MS Dissertation, Queen's University of Belfast, Aug 2003.

[6] J. Li, J. Mirkovic, M. Wang, P. Reiher and L. Zhang, "*SAVE: Source Address Validity Enforcement Protocol*", in Proc IEEE INFOCOM, 2002, pp 1557 -1566.

[7] S.C. Lee and C. Shields, "*Tracing the Source of Network Attack: A Technical, Legal and Societal Problem*", Proc. 2001, IEEE Workshop on Information Assurance and Security, IEEE Press, 2001, pp. 239-246.

[8] Vadim Kuznetsov, Andrei Simkin and Helena Standstorm, "*An evaluation of different IP Traceback Approaches*", Proc. of 4th International Conference on Information and Communication Security, Singapore, 2002, pp 37-48.

[9] Hassan Aljifri, "*IP Traceback: A New Denial of Service Deterrent*", IEEE Computer Society, 2003.

[10] A.C. Snoeren, C Patriage, L. A. Sanchez and S. Kent, "*Hash-based IP Traceback*", Proc. of ACM SIGCOMM conference, 2001, San Diego, CA, Computer Communication Review vol 31, no 24, Oct 2001, pp. 3-14.

[11] Zhiqiang Gao and Nirwan Ansari, "*Tracing Cyber Attacks from Practical Perspective*", IEEE Communications Magazine, 2005.

[12] Chao Gong and Kamil Sarac, "*IP Traceback based on Packet Marking and logging"*, Proc. of International Conference on Communication, 2005, Seoul, Korea, IEEE Communications Magazine, vol 2, May 2005, pp 1043-1047.

[13] D. Dean, M. Franklin and A. Stubblefield, "*An Algebraic Approach to IP Traceback"*, ACM Trans. Info. and Sys. Sec., vol 5, May 2002, pp. 119-137.

[14] Ankit Fadia, "*Network Security: A Hacker's Perspective*", pp. 125-127, Course Technology Inc, 2006. ISBN 1598631632.

[15] S. Savage, D. Wetherall, A. Karlin and T. Anderson, "*Practical Network Support for IP Traceback*", Proc ACM SIGCOMM Conference, Aug 2000, Stockholm, Sweden, Computer Communication Review, vol 30, no 4, Oct 2000, pp. 295-306.