On IPv6 Traceback using Deterministic Packet Marking

Syed Obaid Amin*, Choong Seon Hong*, Il Joong Kim**
*Dept. Of Computer Engineering Kyung Hee University
**National Computerization Agency
e-mail : obaid@networking.khu.ac.kr, cshong@khu.ac.kr, ijkim@nca.or.kr

abstract

The motivation of IP traceback is to identify the true source of an IP datagram in internet. These techniques now emerging as effective deterrent for current cyber threats, especially (D)DoS. Deterministic Packet Marking (DPM) is one of the algorithm used for IP traceback. This paper elucidates the implementation of deterministic packet marking scheme on IPv6 networks. The proposed scheme is capable of single packet traceback. We also examined the issues regarding IPv6 header and show that this scheme is practical, scalable, efficient and can be implemented on existing IPv6 networks easily.

1. Introduction

IP traceback is a mechanism to identify the true source of an IP datagram. These techniques are now emerging as an effective deterrent against (D)DoS attacks. IP traceback techniques neither stop nor prevent the attacks; they are just used to identify the source(s) of the offending packets actively or passively. The architecture of the IP protocol makes it difficult to identify the true source of an IP datagram, if the source wishes to hide it. The routing infrastructure is based largely on destination addresses; none of the entity ensures the correctness of source address. In such instance conventional methods of determining the location of the system with given IP address on the internet no longer work. Therefore more advanced methods of identifying the source of a packet are required.

IP traceback techniques can be categorized in following three groups

1) *Packet marking:* Routers probabilistically or deterministically mark path information in packets as they travel through the Internet. Victims reconstruct attack paths from path fragments embedded in received packets.
2) *Messaging:* Routers probabilistically send ICMP messages, which contain the forwarding nodes the packet travels through, to the destination node. Victims reconstruct attack paths from received ICMP messages.
3) *Packet Digesting:* Routers probabilistically or deterministically store audit logs of forwarded packets to support tracing attack flows. Victims consult upstream routers to reconstruct attack paths.

Along with above, many routers employ a technique called ingress filtering to limit source addresses of IP datagrams [1], but not all routers have the resources necessary to examine the source address of each incoming packet. Ingress filtering provides no protection on transit networks. Furthermore,

This work was supported by ITRC of MIC and by NCA.

spoofed source addresses are legally used by network address translators (NATs), Mobile IP, and various unidirectional link technologies such as hybrid satellite architectures.

IPv6 networks are still in experimental phase and there are quite probable chances of DDoS attack on these networks. So far all of the research has been targeted towards ipv4 networks, and implementing those techniques for IPv6 networks require modifications because of the technological differences. In this paper we present a traceback approach for IPv6 networks. Our proposed architecture falls in Deterministic Packet Marking (DPM) category which will be discussed in detail in later sections.

The rest of this paper is articulated as follows: In section 2, we describe related work. Section 3 will give an overview of DPM. Section 4 outlines our proposed technique, section 5 compare our proposed schemes with evolution metrics. Finally, we summarize our findings in Section 6.

2. Related Work

2.1. Current IP traceback techniques

Regardless of technological differences between IPv4 and IPv6 networks; its worth to discuss briefly the operation of existing proposed schemes for IPv4 traceback. Current IP traceback methods are either reactive or proactive.

2.2. Reactive Measures

Reactive traceback technique initiates the traceback in response of an attack and must complete their operation while attack is active. Controlled flooding [2] and Input debugging [3] are the example of reactive measures. The major advantages of these techniques are compatibility with existing protocol as well as scalability but on the other hand a lot of human intervention is required along with a

requirement that attack must last long enough for successful trace; which is impractical.

2.3. Proactive Measures

On the contrary, proactive techniques record the trace record as the packet traverse through the internet and a victim then used the recorded data for traceback. Examples of proactive measures are logging, messaging, Hash based and packet marking. Logging is to record the packets at the key routers and then to use data mining techniques to extract information from it [3], [4]. This scheme is also backward compatible but requires plenty of resources i.e. processing and storing capabilities.

In July,2000 IETF proposed an ICMP based traceback technique called iTrace. In this technique ICMP messages probabilistically say 1 in 20,000 is send along with path information to the victim. This scheme presents a very expandable technology if implemented with encryption and key distribution schemes but the additional traffic generated consumes a lot of bandwidth even with very low frequency (1/20,000) and without encryption an attacker can inject false ICMP traceback messages also ICMP traffic is filtered in many organization to avoid several attack scenarios which make iTrace not that much useful.

Hash or message digest base scheme was introduced by Snoeren et al, this scheme officially called Source path isolation engine (SPIE) [5]. In this scheme, every router captures partial packet information of every packet to be able to determine whether the packet passed through it or not. This techniques is very efficient and capable of identify a single packet source but on the other hand very computational and resource intensive as well as a lot of ISP involvement is required.

2.3.1 Packet Marking Techniques: Various forms of packet marking algorithms are discussed by Savage et al like node sampling, edge sampling and compressed edge fragment sampling [6]. The schemes were ranging from simply appending an address in packet to probabilistically marking of encoded addresses. Here we will consider the working of final scheme i.e. Fragment Marking Scheme (FMS). In FMS identification field of IPv4 header is divided into three portion 3 bits for fragment ID, 5 bits for distance i.e. no. of hops from victim and 8 bits for edge fragment At each router Ri , as the packet P travels through, the Ri modifies the field in following way.

- Encoding the router's IP address to accommodate the 32 bit address in identification field.
- Marking IP packets with some probability q
- Reconstruction of attack path

One of the major problem with FSM is a large number of packets are required to have a full attack graph. Even if the marking probability is 1/25, 2000 packets are required for 15 hops. In case DDoS this scheme fails drastically due to high computation overhead to check a large number of combinations of the fragments. Second, large number of false positives, because the redundancy check is insufficient and the false positives at a closer distance to the victim can cause even more false positives further away from the victim. Andrey Belenky and Nirwan Ansari in presented an idea of deterministically marking a packet to cope with above problems [8]. The main idea is that the interface closest to the source will mark a packet as soon as it enters the network this mark will remain static throughout the lifetime of a datagram. Each mark is partial information of the address. This scheme use 16 bits ID field of IPv4 and 1 flag bit. This

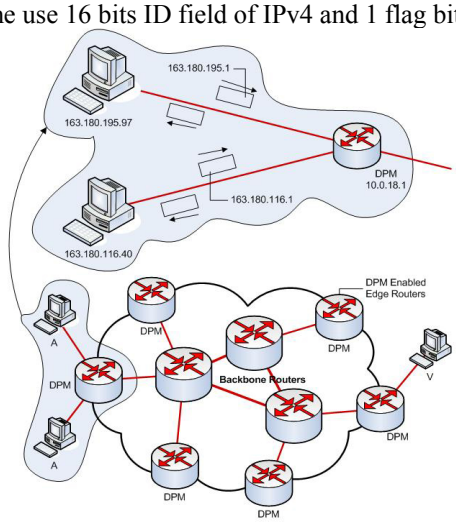

Figure 1. Deterministic Packet Marking (DPM)

scheme will be discussed in more details in sec 3.2.

Mainly all the IP traceback techniques especially; which are discussed above are proposed for IPv4 network and most of these techniques are just theoretical and far away from implementation. So far there are two traceback mechanism available for IPv6 network one is the modification of SPIE for IPv6 [9] and second is to use ICMPv6 but both of these techniques still have the limitation which have been discussed above for hash based and message based IP traceback techniques.

3. Overview of DPM

Scrutinizing all above techniques reveal DPM as more practical and suitable marking algorithm due to its less computation and bandwidth overhead, ease of implementation, security and capability of handling variety of attacks. DPM seems the right scheme to implement for traceback on future IPv6 networks.

3.1. Assumptions

Before starting our discussion it will be useful to discuss DPM for IPv4 in detail. Including DPM, almost all of the techniques discussed above some how follow the following assumption took from [8].

- attackers may be aware they are being traced;
- packets may be lost or reordered;
- an attack may consist of just a few packets;
- packets of an attack may take different routes;
- routers are both CPU and memory limited;
- routers can mark a packet;
- routers are not compromised.

3.2. Introduction of DPM

DPM uses 16 bit ID field and 1 bit reserved flag of IPv4 header each packet is marked with partial address of an

interface of a router closest to the source as it enters the internet. This mark remains unchanged through out the lifetime of a datagram in a network. The interfaces of a router are managed in way that it makes difference between incoming and outgoing packets; as shown in Fig. 1. Incoming packets are marked and outgoing packets are not marked so the mark cannot be overwritten by egress router. As it is apparent, 17 bits are not enough to carry 32 bit address at least 2 packets are required to get the compete IP address. This is where the DPM technique is fairly affected by the size limitation of IPv4 header. The IP address is split into two parts of 16 bits each. The ID field is filled with either of these parts with some probability say 0.5 and the reserved flag bit is set to 0 for first 16 bits and 1 for rest of the bits. In this scheme victim on an average can receive the complete IP address in 7 packets (P=1-0.57 ~ 0.9922) and only 10 packets with the probability of more than 99.9%. It's very simple to find these figures in DPM as unlike other packet marking techniques we don't have multiple attack paths here.
The reconstruction mechanism uses the data structure called Reconstruction Table (RecTbl). The destination would first put the address segments in RecTbl, and then only after correctly identifying the ingress address, out of many possible address segments permutations, would transfer it to IngressTbl.

3.3. Problems with Simple DPM

To understand the problem clearly its better to explain one key metric of DPM i.e. false positive. A false positive is defined as an incorrectly identified ingress address. The rate of false positives refers to the ratio of the incorrectly identified ingress addresses to the total number of identified ingress addresses. It is shown in equation 1

$$\text{rate of false positive} = \frac{\text{incorrectly identified ingress address}}{\text{the total number of identified ingress address}} \quad (1)$$

The problem occurs when two hosts with the same Source Address (SA) attack the victim. The ingress addresses corresponding to these two attackers are A0 and A1, respectively. The victim would receive four address segments: A0[0], A0[1], A1[0], and A1[1] and eventually reconstruct four ingress addresses, since four permutations are ultimately possible: A0[0].A0[1], A0[0].A1[1], A1[0].A0[1], and A1[0].A1[1], where '.' denotes concatenation. Only two of the four would be valid. For 3 attackers the rate of false positive would be 62.5 and starts to get worst as the number of attackers increases. To cope with this problem Andrey Belenky and Nirwan Ansari came up with slightly modified approach i.e. hash-based DPM [11] which in turns produced a significant amount of processing along with large number of packets requirement. Keeping this problems in mind the implementation of DPM in IPv6 network had to be designed in such a way that above problems shouldn't occur and don't add processing overhead of existing networks.

4. Proposal

Implementation of DPM technique to IPv6 requires a thorough analysis of IPv6 header so we can efficiently mark the packet [10]. IPv6 header is simpler but longer than IPv4 header as shown in Figure 2. Two of the fields of IPv6 header i.e. Flow Label and Hop-by- Hop options header can be used

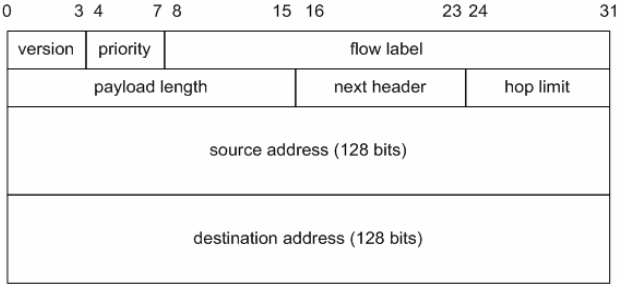

Figure 2. IPv6 basic header format

to store the mark. The generation of mark is dependent on the choice of implementation. Using Flow label has many disadvantages first of all flow label field is general purpose and can't be used for specific purposes. Furthermore to accommodate 128 bit IPv6 address to 20 bits again require encoding which is certainly a computational intensive job. Below we'll discuss the other technique i.e. using separate option in Hop-by-Hop option header.

4.1. Separate option in Hop-by-Hop Options header

Another way of marking a packet deterministically is by adding a new option in Hop-by-Hop options header. The interface of a router closest to the source will mark this new option with the IPv6 address of that interface. As we mark each and every packet with complete IP address so that single packet would be enough to get the complete the traceback. Once the victim aware that he is under attack he can immediately drop the packets coming from the interface(s) near to the attacker and then can request the concern ISP to monitor the traffic to get any rough idea about exact source of attack(s). The later task entirely depends on mutual agreement of ISPs and discussed in detail in section 5.
The new option should be design in such a way that packet overhead is minimal for this purpose RFC 2460 provides formatting guideline to add new options in hop-by-hop or destination options header. These guidelines are on the following assumptions.

- Fields of width n octets should be placed at an integer multiple of n octets from the start of the Hop-by-Hop or Destination Options header.
- The Hop-by-Hop or Destination Options header should take as little space as possible, provided that the header be an integer multiple of 8 octets long.
- When either of the option-bearing headers is present, they carry a very small number of Options.

Based on above guidelines the basic option format is shown in Figure 3, without presence of any other options. The

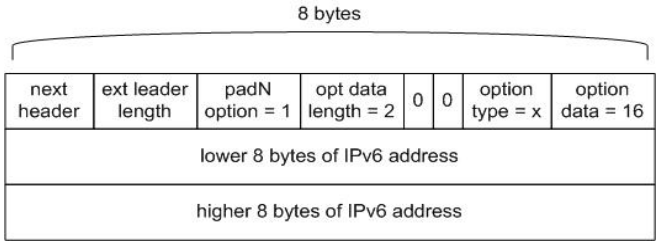

Figure 3. Hop-by-Hop option

alignment requirement is 8n+6. The Option type is internally subdivided into three fields, first two bits determines what action should be taken if the device processing the option the doesn't recognize the option type third bit specify whether option data can be changed en route the remaining five bits along with these three bits produce a unique identifier of the option. By zeroing first two bits, this scheme shows that nodes not recognizing this option type should skip over this option and continue processing the header. Setting third bit to 0 shows that this option must not change en route, the details of other field can be found in [10].

5. Discussion

Marking complete IP address give solid shore up to this technique by eliminating the drawback of DPM when multiple hosts cast the attack [11]. Below we are comparing the efficiency of implemented scheme with key evaluation metrics discussed in [7].

ISP involvement: Regarding ISP involvement it's weird to think all the ISPs in the world are going to implement this technique but the ISP doesn't implementing it can sincerely inform others ISPs or an ISP can also examine by observing the absence of the traceback option in this case the ISP doesn't implementing DPM would be considered as potential hacker and DPM should be implemented on the interface(s) connected to that client ISP. It's worthy to mention that for other IP traceback mechanism if intermediate nodes don't participate than it's nearly impossible to trace back an attack path.

Number of Attacking Packets: Only one packet is enough to complete traceback which also eliminates the path reconstruction problem; one of the major weakness of PPM techniques.

Processing overhead: For an ideal traceback scheme the processing overhead of traceback should be minimum. This additional processing for traceback could occur on the devices of an ISP and/or at probable victim(s) of invasion. It is apparent; the proposed scheme doesn't require any calculation of hash values or message digests, encoding/decoding or any other computational intensive job either on intermediate routers or at victim side.

Bandwidth overhead: On the other hand, we have to slightly compromise on bandwidth consumption due to addition of one option header but this compromise is acceptable as we already have much bigger routing header in IPv6 specification.

Ease of Evasion: Refers how easily an attacker can circumvent the traceback technique. Evasion is not possible in DPM as each incoming packet is marked and any spoofed mark of an attacker will be overwritten by the router interface closest to the source.

Protection: Relates to produce the meaningful traces if some of the devices included in traceback are undermined. DPM is highly protective as intermediate router don't participate in traceback and the single point of consideration is the router interface closest to the attacker if this interface or a router is down then there would be no way for an attacker to invade. Ability to handle major (D)DoS attack Marking of each and every packet with complete address information make this scheme capable to trace not only DDoS but also other sort of incursion.

6. Conclusion and future work

In this paper we first gave an introduction of IP traceback and a brief overview of current IP traceback trends. Subsequently, we gave a detail description of DPM, its working and discussed it merits demerits. Finally we gave a brief introduction of IPv6 header and proposed application of DPM in IPv6 networks. We vindicated our scheme by comparing with several evaluation metrics and found that this scheme is practical, easy to implement, difficult to evade and only one packet is sufficient to identify the source of attack. In future we will observe the effects of marking packet with some probability to reduce the computation time of modifying each and every packet.

7. References

[1] P. Ferguson and D. Senie. Network Ingress Filtering: Defeating Denial of Service Attacks Which Employ IP Source Address Spoofing. RFC 2267, Jan. 1998.

[2] H. Burch and B. Cheswick. Tracing Anonymous Packet to Their Approximate Source. Unpublished paper, Dec. 1999.

[3] R. Stone. CenterTrack: An IP Overlay Network for Tracking DoS Floods. In to appear in Proceedings of the 2000 USENIX Security Symposium, Denver, CO, July 2000.

[4] G. Sager. Security Fun with OCxmon and cflowd. Presentation at the Internet 2 Working Group, Nov. 1998.

[5] A. Snoeren, C. Partridge, L. Sanchez, C. Jones, F. Tchakountio, B. Schwartz, S. Kent, and W. Strayer. Single-packet IP traceback. ACM/IEEE Transactions on Networking, Dec.2002.

[6] S. Savage, D. Wetherall, A. Karlin, and T. Anderson, .Network support for IP traceback,. IEEE/ACM Transactions on Networknig, vol. 9, no. 3, pp. 226.237, June 2001.

[7] A. Belenky and N. Ansari, .On IP traceback,. IEEE Communications Magazine, vol. 41, no. 7, July 2003.

[8] A. Belenky and N. Ansari, .IP traceback with deterministic packet marking, IEEE Communications Letters, vol. 7, no. 4, pp. 162.164, April 2003.

[9] W. Timothy Strayer, Christine E. Jones, Fabrice Tchakountio, and Regina Rosales Hain, SPIE-IPv6: Single IPv6 Packet Traceback, Local Computer Networks, 2004. 29th Annual IEEE International Conference on 16-18 Nov. 2004 Page(s):118 – 125.

[10] S. Deering, R. Hinden, Internet Protocol, Version 6 (IPv6) Specification, RFC 2460, IETF, December 1998.

[11] A. Belenky and N. Ansari, Tracing Multiple Attackers with Deterministic Packet Marking (DPM),Communications, Computers and signal Processing, 2003. PACRIM. 2003 IEEE Pacific Rim Conference on Volume 1, 28-30 Aug. 2003 Page(s):49 - 52 vol.1.