Deterministic Flow Marking for IPv6 Traceback (DFM6)

Vahid Aghaei-Foroushani and A. Nur Zincir-Heywood
Faculty of Computer Science
Dalhousie University
Halifax, NS, Canada
Email: {vahid, zincir}@cs.dal.ca

***Abstract*—Although some security threats were taken into consideration in the IPv6 design, DDoS attacks still exist in the IPv6 networks. The main difficulty to counter the DDoS attacks is to trace the source of such attacks, as the attackers often use spoofed source IP addresses to hide their identity. This makes the IP traceback schemes very relevant to the security of the IPv6 networks. Given that most of the current IP traceback approaches are based on the IPv4, they are not suitable to be applied directly on the IPv6 networks. In this research, a modified version of the Deterministic Flow Marking (DFM) approach for the IPv6 networks, called DFM6, is presented. DFM6 embeds a fingerprint in only one packet of each flow to identify the origin of the IPv6 traffic traversing through the network. DFM6 requires only a small amount of marked packets to complete the process of traceback with high traceback rate and no false positives.**

Keywords—Flow Based IP Traceback; DDoS Attacks; IPv6; Network Security

I. Introduction

IP traceback is a mechanism which aims to identify the true source of an IP datagram. However, as many current IP traceback schemes are proposed concerning IPv4 network, they cannot be directly used in IPv6 network. Implementing those techniques for IPv6 networks require modifications because of the technological differences; such as the differences in the IP header.

To this end, we propose DFM6, a new traceback approach under IPv6 that helps network administrators actively and effectively traceback to the source of attacks in a short time when suffering from an attack with spoofed source IP addresses. Our proposed architecture falls in Deterministic Flow Marking (DFM) [1]–[3], one of the marking techniques used for IPv4 networks. DFM has two unique features that lead us to propose our IPv6 traceback approach based on this technique. First, unlike other marking methods that mark the traffic at the packet level, DFM marks the traffic at the traffic flow level. This feature makes the DFM to have low marking rate. It requires a small amount of marked packets to find the source of DDoS attack at the victim side. Second, it can traceback up to two levels behind the marking router, which most of the times it can infer not only the attacking network, but also the attacker node behind the network address translation (NAT) or proxy devices.

II. Related Works

There are some IPv6 traceback approaches that mark the packets by taking advantage of the IPv6 extension header and store the marking information in either the Hop-by-Hop Option or the Destination Option extension (DOH) headers. To the best of the authors' knowledge, there are four previous works on this category that aim to adapt the Deterministic Packet Marking IP traceback approach (DPM) [4] on the IPv6 networks. In particular, Obaid et al. [5] uses 24 bytes hop by hop extension header to store the marking data into every outgoing packet. The interface of the router closest to the source of attack marks the packets.

You-ye et al. [6] uses the same concept as Obaid et al. [5] applies, but employs the Destination Option Header to store the 24 octets marking data. They also introduce two thresholds, L_min' and L_max terms. Only when the load is between L_min and L_max, the packets are marked. For reconstruction, the victim finds the marked packets by looking at the DOH field, and extracts the ingress IP address of the marking router.

Animesh et al. [7] divides the 128 bits ingress address to K segments, and marks every packet by the K bits IPv6 fragment, d bits hash of the IPv6 address and 2^a bits for the fragment offset. With the suggested number of $k = 16$, $d = 11$ and $a = 2$, the marking data for each packet is 8 octets which should be stored in the Destination Option Header. Reconstruction procedure consists of mark recording and ingress address recovery. Mark recording process indicates which mark fragment arrived to the destination. Address recovery reconstructs the IP address segments and determines which ones are valid.

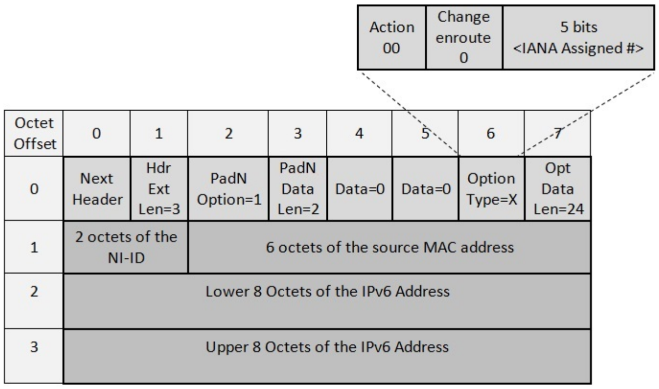
Ashwani et al. [8] also introduce a modified Deterministic Packet Marking approach for IPv6 networks that uses 40 octets hop by hop extension header to store the ingress IP address of the edge router as well as its hash, into every outgoing packet.

In our previous work, we proposed the Deterministic Flow Marking, DFM, to reduce this overhead by marking every flow instead of every packet, given that all packets in the same flow belong to the same source [1]–[3]. In this paper, we aim to extend the DFM approach for IPv6, to have both of the advantages: a high traceback accuracy low processing overhead for the IPv6 networks.

III. DFM for IPv6 Networks (DFM6)

We aim to design the DFM6 general enough so that it can be used to find the source of any kind of IPv6 traffic. To achieve this, we assume the following: 1) Attackers may be aware they are being traced, 2) Attackers may spoof the source

978-3-901882-77-7 © 2015 Crown



Action	Change enroute	5 bits
00	0	<IANA Assigned #>

Octet Offset	0	1	2	3	4	5	6	7
0	Next Header	Hdr Ext Len=3	PadN Option=1	PadN Data Len=2	Data=0	Data=0	Option Type=X	Opt Data Len=24
1	2 octets of the NI-ID		6 octets of the source MAC address					
2	Lower 8 Octets of the IPv6 Address							
3	Upper 8 Octets of the IPv6 Address							

Fig. 1: The Proposed Format of the Destination Options Header for Storing the Marking Data

MAC and IPv6 addresses. We also assume that the medium access control (MAC) filtering is enabled at the edge routers. We believe that such an assumption is realistic since most of the routers have this function enabled.

We have designed the DFM6 by two modules, DFM6 Encoding (DFME) and DFM6 Decoding (DFMD) modules. In the following, we describe these two modules in detail.

A. *Selecting the Marking Field*

To apply the DFM to the IPv6 networks, an appropriate field in the IPv6 header for embedding the marking data should be selected. Several extension headers are defined in IPv6 to support a broad range of applications, and new extension headers may be defined in the future [9]. Among these extension headers, only the Destination Options Header (DOH) is designed to be examined and processed at the destination node [9]. In addition, DOH can provide a large enough space for storing the entire marking data. Therefore, for our DFM6 approach, DOH is selected for storing the marking data in the IPv6 networks. By employing the DOH for storing the marking data, we do not have the marking data fragmentation overheads that are experienced in the previous DFM design for the IPv4 network. DOH is identified by a Next Header value of 60 in the immediately preceding header [9].

B. *Mark Encoding*

The mark encoding task is assigned to the DFM6 Encoding module (DFME), which runs at the marking routers. The architecture of the DFME is based on the fact that all packets in a flow have the same source. Thus, if the origin of even one packet in a flow can be found, then the origin of all of the other packets in the same flow is also discovered. Flow detection is an embedded feature in almost all manageable routers (i.e. Cisco has NetFlow, InMon has sFlow, and Juniper uses JFlow). The DFME module marks the outgoing IPv6 flows by selecting and marking only the first packet of each flow. In DFM6, only the edge routers mark the flows, and the rest, including the core routers among the traffic path from the source to the destination, are not involved in the marking process.

The DFME module uses three identifiers to mark the flows in order to trace up to the attacker node. These three identifiers are:

- The IPv6 address of the egress interface of the edge router (16 octets): An edge router is the closest router to the attacker node with at least one public IPv6 address to its egress interface.
- Node-ID (6 octets): An identifier assigned to each source MAC address observed from the incoming IPv6 packet to the edge router. If the edge router is the closest router to the end network, then the source MAC address would be the MAC address of the source of the packet; otherwise, the source MAC address would be the MAC address of the previous hop (usually the previous router in the path).
- The network interface identifier, NI-ID (2 octets): This is an identifier assigned to each interface of either the MAC address of a network interface on the edge router, or the VLAN ID of a virtual interface if the edge router uses VLAN interfaces. The NI-ID specifies from which subnet a traffic flow comes.

Using the above three identifiers (24 octets marking data) to mark the first packet of each IPv6 flow, DFM6 is able to traceback not only up to the source edge router, but also to the exact source network interface of the edge router and even one step further to the source node located in a LAN behind the edge routers.

The tuple of source and destination IPv6 address and the flow label is intended to uniquely identify a particular flow during its lifetime (plus a subsequent quarantine period) [10].

In this paper, we define the 3-tuple of source and destination IPv6 addresses and the flow label as the flow ID. The DFME module maintains a table to keep track of the marked packets, and their respective flow IDs. This table is called the Marking-Table. Once the DFME module observes an outgoing packet, first of all it extracts its flow ID and then looks for the existence of a table record for this flow ID in the Marking-Table. If this flow ID is not in the Marking-Table, it means that this flow is new to the DFME module, so the DFME module takes the following steps:

1) Creating a table record for the new Flow ID in the Marking-Table;
2) Calculating the 24 octets flow marking data;
3) Marking the selected packet by the marking data. If the selected packet does not have a Destination Option Header, create one and store the marking data; otherwise add the marking data to the available Destination Option Header.

However, if this flow ID is already in the Marking-Table, it means that this flow has been already marked by the DFME module. In this case, the DFME module just forwards the packet without any change. It should be noted that because a flow label of zero indicates that the packet is not part of any flow, therefore the DFME module should mark every packet with a flow label of zero. To eliminate the mark spoofing attack, the DFME module should watch the Destination Option Header of the incoming packets. If it faces a packet with the same DFM6 identifier in the Destination Option Header, this packet should be considered as a mark spoofing attack. In this case, the DFME module overwrites the Destination Option Header with the correct marking data.

C. Marking Data Format

As described before, the IPv6 Destination Option header is selected for storing the marking data. Figure 1 shows the proposed format of the Destination Options header for storing the marking data, which is based on the formatting guideline described in [9], and has the following fields:

- Next Header (One octet): Identifies the type of header immediately following the Destination Options header.
- Hdr Ext Len (One octet): Length of the Destination Options header in 8-octet units, not including the first 8 octets. As the total size of our proposed Destination Option Header is four 8 octets, the value of this field should be 3.
- The PadN option is used to align subsequent options and to pad out the containing header to a multiple of 8 octets in length. For N octets of padding, the PadN Data Len field contains the value N-2, followed by N-2 zero-valued octets. In our proposed Destination Option Header, we need 4 octets of padding, so the PadN Data Len should be 2, followed by 2 zero-valued octets.
- Option Type (One octet): The Option Type is internally encoded into three fields, such that their highest-order two bits specify the action that must be taken if the processing IPv6 node does not recognize the Option Type. The third-highest-order bit specifies whether or not the Option Data of that option can change en-route to the packet's final destination. The remaining five bits along with these three bits produce a unique identifier of the option. By zeroing the 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
- Option Data Len (One octet): Length of the Option Data field, in octets. In our scheme, we have 24 octets marking data, so we set this field to 24.
- Option Data: Variable-length field. We store our 24 octets marking data in this field.

D. Calculating the Path Maximum Transmission Unit

In our method, when the sender sends a packet that is equal to the size of the path MTU, the DFME module cannot add the Destination Option Header to the packets to mark the flows. It is because marking the packet with an extra 32 bytes will increase the size of the packet more that the path MTU, and given that the intermediate routers cannot do fragmentation, the packets will be dropped.

There are some previous researches to avoid this problem [11]. In this work, we modify the previous algorithms to adapt them with our proposed flow based IPv6 traceback approach. In our scheme, the path MTU should be reduced by 32 bytes, so that the marking router is able to add an extra 32 bytes destination option header to the first packet of each flow. To this end, our proposed DFME module on the edge router reduces the MTU value of the "ICMPv6 Too Big" packets by 32, and returns the packet back to the sender. The sender then sends the packet according to the size of this modified MTU. According to [9], an IPv6 device cannot have less than 1280 bytes MTU. Therefore, the minimum MTU size received by the DFM6 must be 1312 bytes (1280 plus 32 bytes for the Destination Option Header).

E. Mark Decoding

The DFM6 Decoding module (DFMD) is located at the destination network and its goal is to infer the origin of the incoming traffic. As the whole marking data is stored in the Destination Option Header, mark decoding consists of a simple process which extracting the marking data from the DOH header. Given that the origin of all packets in a flow is the same, once the marking data of a flow is extracted from one packet, then the origin of all other packets in the same flow are also discovered. Using PFMD, the destination is able to distinguish the traffic of different nodes behind an edge router. As a result, when an abnormal traffic is observed, the victim is able to distinguish between the attack and the legitimate traffic and infer the source of an attack, even if it is behind a NAT or a proxy device.

IV. Experimental Results

To evaluate our proposed DFM6 approach and compare its performance with the other previous deterministic methods, we implemented our approach and four DPM approaches described in the literature [5]–[8], and evaluated them on a testbed network. We also employed the CAIDA IPv6 5 June 2012 Anonymized Internet Traces as the evaluation traces. This dataset is standard tcpdump traffic and is publicly available [12]. The evaluation traffic is sent from a local area network behind a marking router and directed to the destination. For this purpose, we replayed the data sets on the testbed network using tcpreplay and tcprewrite open source applications [13]. In addition, we implemented the mark encoding and the mark decoding programs to mark the packets at the edge router and traceback the source of traffic at the destination.

To evaluate our DFM6 approach, and compare it with the other deterministic marking approaches, we use the following four evaluation metrics:

- Traceback Rate: The ratio of the number of successfully traced back packets to all packets.
- Marking Rate: The ratio of the marked packets to all packets.
- Bandwidth Overhead: The ratio of the volume of the overhead traffic to the volume of the original traffic.
- Number of Required Packets: The number of required marked packets at the destination to complete the traceback process.

Naturally, the desired outcome has higher traceback rate, lower marking rate, lower bandwidth overhead and lower number of required packets to complete the traceback process. Table I presents our evaluation results of our proposed DFM6 approach and the four previous deterministic approaches on IPv6 traceback from the literature [5]–[8]. These results show that all of the deterministic IPv6 traceback methods have 100%

TABLE I: The Evaluation Results of our proposed DFM6 approach, as well as four previous deterministic approaches on IPv6 traceback.

Method	Traceback Rate	Marking Rate	Bandwidth Overhead	Number of Required Packets
DFM6	100%	13.7% (One Packet/flow)	3.92%	1
DPM2006	100%	100% (Every Packet)	18.57%	1
DPM2011	100%	100% (Every Packet)	18.57%	1
DPM2012	100%	100% (Every Packet)	6.18%	16
DPM2014	100%	100% (Every Packet)	30.95%	1

traceback rate. However, they achieve this at the expense of 100% marking rate.

On the other hand, the proposed approach DFM6 achieves a 100% traceback rate by only marking approximately 14% percent of the packets. It should be noted here that the marking rate of DFM6 could be different based on the dataset. However it would always be less than the other DPM approaches because the marking rate of all other DPM approaches is always 100%, regardless of the traffic pattern, as they mark every packet, whereas DFM6 only marks one packet per flow.

As described is section II, the DPM2006 [5], DPM2011 [6], DPM2012 [7] and DPM2014 [8] approaches mark each packet by an additional 24 octets, 24 octets, 8 octets and 40 octets marking data, respectively. DFM6 marks the packets by extra 32 octets, but only one packet per flow need to be marked. Therefore as we expected, the bandwidth overhead for the DFM6 is much lower than the other four DPM approaches (3.92%), and it can be even lower for DDoS attack traffic as the number of packet per flow in DDoS traffic is usually more than the normal traffic.

Finally, the number of required marked packets to complete the traceback process at the destination for all approaches, except the DPM2012, is one. It is because all of these approaches store the complete marking data in each marked packet. Therefore, at the destination, only one marked packet is enough to complete the traceback process. However at the source side, DFM6 marks only one packet per flow, while the other approaches mark every packet.

Our results show that our proposed DFM6 approach outperforms the other deterministic marking approaches, as it has the same traceback rate, but the lowest bandwidth overhead, marking rate and number of required packets.

V. Conclusion

In this paper, we have presented DFM6, a novel Deterministic Flow Marking for IPv6 networks, which is able to traceback up to the attacker node behind a NAT or a proxy server. DFM6 performs by selecting and marking only the first packet of each flow. This leads us to have both advantages of the high traceback accuracy as well as the low processing and marking overhead. DFM6 consists of two modules, the DFM6 encoding module, DFME, and the DFM6 decoding module, DFMD. The DFME module runs at the egress interface of the marking router, and marks the outgoing traffic. The PFMD module runs at the destination network and tries to infer the source of traffic by extracting the marking data from the marked packets. We have compared the DFM6 with four deterministic IPv6 traceback approaches. Our results based on the CAIDA IPv6 datasets show that DFM6 has the same traceback rate with the other four IPv6 traceack approaches (100%), but it outperforms the other approaches as it has the lowest bandwidth overhead (3.92%) and marking rate (13.7%). It also requires only one packet per flow to complete the traceback process. For future work, we will explore the tradeoff between the traceback rate of the scheme with the number and the location of the participating routers.

Acknowledgement

This research is supported by the Natural Science and Engineering Research Council of Canada (NSERC) grant, and is conducted as part of the Dalhousie NIMS Lab at: `https://projects.cs.dal.ca/projectx/`

References

[1] V. Aghaei-Foroushani and A.N. Zincir-Heywood. Ip traceback through (authenticated) deterministic flow marking: an empirical evaluation. *EURASIP Journal on Information Security*, 5:., 2013.

[2] V. Aghaei-Foroushani and A.N. Zincir-Heywood. Deterministic and authenticated flow marking for ip traceback. *The 27th IEEE International Conference on Advanced Information Networking and Applications (AINA), Barcelona*, 5:25–28, March 2013.

[3] V. Aghaei-Foroushani and A.N. Zincir-Heywood. On evaluating ip traceback schemes: a practical perspective. *IEEE International Workshop on Cyber Crime (IWCC 2013), San Francisco*, pages 127–134, May 2013.

[4] A. Belenky and N. Ansari. On deterministic packet marking. *The International Journal of Computer and Telecommunications Networking, July*, 51(10):2677–2700, Jul 2007.

[5] Syed Obaid Amin and Choong Seon Hong. On ipv6 traceback. *The 8th International Conference on Advanced Communication Technology (ICACT)*, pages 2139–2143, Feb. 2006.

[6] You ye Sun, Cui Zhang, Shao qing Meng, and Kai ning Lu. Modified deterministic packet marking for ddos attack traceback in ipv6 network. *11th IEEE International Conference on Computer and Information Technology*, pages 245–248, Aug. 2011.

[7] Animesh Tripathy, Jayanti Dansana, and Debi Prasad Mishra. A secure packet marking scheme for ip traceback in ipv6. *the International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, pages 656–659, Aug. 2012.

[8] Ashwani Parashar and Ramaswamy Radhakrishnan. Improved deterministic packet marking algorithm for ipv6 traceback. *International Conference on Electronics and Communication Systems (ICECS)*, pages 1–4, Feb. 2014.

[9] Internet protocol, version 6 (ipv6) specification, rfc 2460. http://tools.ietf.org/html/rfc2460, Dec. 1998.

[10] Use of the ipv6 flow label as a transport-layer nonce to defend against off-path spoofing attacks, internet-draft, accessed september 14, 2015. http://tools.ietf.org/html/draft-blake-ipv6-flow-label-nonce-02, Oct. 2009.

[11] Syed Obaid Amin, Muhammad Shoaib Siddiqui, and Choong Seon Hong. A novel ipv6 traceback architecture using cops protocol. *annals of telecommunications - annales des tlcommunications*, Apr. 2008.

[12] The caida anonymized internet traces on world ipv6 day and world ipv6 launch day dataset, accessed september 14, 2015. http://www.caida.org/data/passive/passive_ipv6day_and_ipv6launch_dataset.xml.

[13] A. turner, tcpreplay suite, accessed february 13, 2015. http://tcpreplay.synfin.net/.