TRACEBACK OF DOS OVER AUTONOMOUS SYSTEMS

Mohammed Alenezi¹ and Martin J Reed²

¹School of Computer Science and Electronic Engineering, University of Essex, UK mnmale@essex.ac.uk ²School of Computer Science and Electronic Engineering, University of Essex, UK mjreed@essex.ac.uk

ABSTRACT

Denial of service (DoS) is a significant security threat in open networks such as the Internet. The existing limitations of the Internet protocols and the common availability tools make a DoS attack both effective and easy to launch. There are many different forms of DoS attack and the attack size could be amplified from a single attacker to a distributed attack such as a distributed denial of service (DDoS). IP traceback is one important tool proposed as part of DoS mitigation and a number of traceback techniques have been proposed including probabilistic packet marking (PPM). PPM is a promising technique that can be used to trace the complete path back from a victim to the attacker by encoding of each router's 32-bit IP address in at least one packet of a traffic flow. However, in a network with multiple hops through a number of autonomous systems (AS), as is common with most Internet services, it may be undesirable for every router to contribute to packet marking or for an AS to reveal its internal routing structure. This paper proposes two new efficient autonomous system (AS) traceback techniques to identify the AS of the attacker by probabilistically marking the packets. Traceback on the AS level has a number of routers that need to participate in the marking. Our results show a better performance comparing to PPM and other techniques.

Keywords

Autonomous System, DoS, Packet Marking, Network Security, Traceback.

1. INTRODUCTION

Denial of service (DoS) becomes a serious threat in Internet today due to the available variety of DoS forms, size, types, and tools. Attackers can increase the involved machines in the attack from single attacker causing a classic DoS attack to many compromised machines causing a distributed denial of service (DDoS) attack. The attack can be amplified to increase the volume of the attack by using non-compromised machine to reflect spoofed requests containing the victim IP address. Arbor networks [1] reports that DDoS attacks can reach sizes of up to 40 Gigabits/sec.

In general, research into DoS has been divided into three phases: prevention, detection, and mitigation. Identifying the origin of the attack is defined as traceback which is one part of the mitigation phase. Many techniques have been proposed to traceback the origin of the attack. Probabilistic packet marking (PPM)[2] and deterministic packet marking (DPM)[3] are the two main forms of packet marking in general.

Most of the proposed techniques are on a hop-by-hop basis where the topology of the ISP is revealed. However, this paper proposes two new techniques for traceback identifying the AS of the attacker. Our techniques will mark the packet with the AS number (ASN) based on a dynamic

probability to improve the performance of the technique. Knowing the distance, for a packet journey from source to destination, in advance is one of the advantages of using the AS. Unknown distance was one of the main limitations of PPM and other techniques in specifying the marking probability for routers. Additionally, the proposed marking probability will reduce the required number of packets to reconstruct the attack path. In PPM, marks from the early nodes on the attack are overwritten by the downstream nodes. Therefore, the required number of packets to reconstruct the attack path.

The paper is organized as follows. Related work is introduced in section 2. The design and motivations are presented in section 3. The first proposed idea, prevent overwriting AS traceback (POAST), is explained in section 4 while the second the proposed idea, efficient AS traceback (EAST), is presented in section 5. Finally, the conclusion is presented in section 6.

2. RELATED WORK

Many techniques have been proposed to traceback the origin of DoS attack based on different methodologies and on different levels. In the following, different traceback techniques will be presented based on the proposed methodology and level of traceback.

Link testing[2] is based on testing the links in the upstream direction and can be found in two types: *input debugging* and *controlled flooding*. The traffic is filtered and directed between the edges of the network in *input debugging*, while flooding the network by traffic to study the effect on the attack traffic is in *controlled flooding* [4].

Selecting specific routers to log the packets and using different data mining techniques to reconstruct the attack path is termed *logging* [2]. Using special ICMP packets with a probability of 1/20,000 to carry the path information is known as *ICMP Traceback* [5].

Packet marking techniques are divided into main types: probabilistic packet marking (PPM) [2] and deterministic packet marking [3] (DPM). PPM marks the packets with the IP address of the router based on a fixed probability P=1/25. PPM uses edge sampling as a marking algorithm which encodes the edges (two nodes) instead of recording the nodes as in node append or encodes each node as in nodes sampling [2]. However, there are drawbacks with PPM [3, 6, 7]. Marking information inside the packets, which will be used by the victim for tracing the attacker's source, could be explicitly violated by the attacker due to high number of unmarked packets in PPM. The further the router is away from the victim on the attack path, the higher the chances for its marked packets to be overwritten by a router closer to the victim. Therefore, the required number of packets to reconstruct the attack path is high. Furthermore, PPM increases the overhead and processing in routers because it involves all the routers on the attack path in the marking process. DPM is proposed to traceback the attack deterministically. DPM will mark all of the packets at the access router and will not provide the full attack path but rather identify the access router that is closest to the attacker. In other words, DPM marks the packets with a probability of 1. The source IP address of the access router is divided into two parts where each part with a size of 16 bit. Using a probability of 0.5, one of the two parts will be injected inside the packet. Just like PPM, there are drawbacks with DPM [7-9]. First, a scalability problem exists as there are only 25 spare bits of the IP packet that can be used for updating [6]. In addition, routers need considerable storage for packet logging [9]. Moreover, routing software needs to be modified in order to apply DPM and the added overhead would slow down the routers. The IP address (32 bits) of an edge router will be marked in two packets. Therefore, the victim needs to receive two packets to identify the IP address. This raises a problem of accuracy if the one of the packets lost.

Furthermore, as with PPM, modifications by the attacker can be used to mark the message to confuse the victim.

Dynamic probabilistic packet marking for efficient IP traceback (DPPM) [10] is another form of PPM as it marks the packet based on probability. However, it uses a dynamic probability where the change in the probability is based on the deduced distance from the TTL field in the IP packet. Although the number of required packets to reconstruct the path are reduced, it has some drawbacks. The main problem is accurately determining the travelling distance, as DPPM counts on the TTL field to deduce the distance which will be used for calculating the marking probability. Additionally, the TTL field can be spoofed intentionally by the attacker to confuse the technique.

Authenticated AS traceback (AAST) [11] has been proposed to traceback the AS as the origin of the attack instead of specifying the attacker source itself. The technique is divided into two parts: AS traceback and the authentication scheme. The traceback is carried out by probabilistically marking the packets between the AS's using the ASN instead of the IP address. The marking technique is the same as in PPM, however, they have changed the value of the marking probability to P=1/6. The required space is 16 bits for the ASN and 3 bits for the distance. The authenticated scheme was presented to prevent the spoofed information inserted by an attacker. It is based on using symmetric key infrastructure through a hash function and exchange of the secret key between the Autonomous system border routers (ASBR's).

3. DESIGN AND MOTIVATIONS

3.1. Design

Both of the proposed traceback techniques in this paper will be applied on the AS level. The key points of our design for the presented work are based on the following:

- Reduce the required number of packets in the traceback process.
- Reduce the required storage at the victim side by reducing number of required packets and avoid any logging schemes.
- Eliminating the unmarked packets, to avoid spoofing that would be intentionally injected by the attackers.
- Reduce the number of routers involved in packet marking.

3.1. Motivations

Including the AS in the traceback process brings many advantages to overcome limitations with PPM and other techniques. Using the AS in traceback will not reveal the topology, which is advantageous for network operators. Additionally, the number of AS hops is significantly less than actual routers hops. In 99% of cases, a packet passes through less than 6 AS [12, 13] in its journey from source to destination, while it passes over 20 routers in the hop by hop basis. Furthermore, routing on the AS is carried by BGP and the distance can be known in advance between ASs. Knowing the distance in advance, can be used to optimize the marking probability and reduce the required number of packets to reconstruct the path.

4. PREVENT OVERWRITING AS TRACEBACK

In this section our prevent overwriting AS traceback (POAST) technique is presented. It is based on marking the packets with a dynamic marking probability and protects the marked packets from overwriting by downstream routers. The marking is carried out by only one router in each AS by encoding the AS number instead of a router's IP address and thus determines the AS of the attack.

4.1. Available Space and Assumptions

4.1.1. Available Space

Most packet marking techniques use 25 bits or more for overloading the IP packet [2, 11]. Our technique will use the available 25 bits comprising: 8-bit type of service, 16-bit ID and 1-bit reserved flag. These will be used to encode the fields shown in table.1.

4.1.2 Assumptions

Our assumptions can be summarized in the following:

- The attack could be triggered by single or multiple attackers.
- The attacker/attackers may be aware of the tracing.
- There may be packet loss in any number of packets sent.
- AS routes are stable during the attack.
- The AS routers are not compromised.

4.2. Proposed Technique

Border gateway protocol (BGP) is used for routing the packets between the AS. Each ASBR has a full path for each AS reachability. At ASBR, upon receiving the packet, BGP table will be checked and based on the destination, the ASBR selects the path for the packet after matching the prefix (policies are applied). Using the BGP table, the ASBR knows how many required AS hops to the destination [14]. Knowing number of hops in advance is considered to be a significant advantages over most the proposed techniques as the marking probability can be optimized. For example let assume we have 5 AS's (AS src- AS1- AS2- AS3- AS dest) illustrated in Fig.1. A packet traversing from an attacker in AS_src destined to a victim in AS_dest, has to pass through ASBR_src before passing AS1, AS2, and AS3. After the attacker starts to send packets to the victim, the first router will receive the packets is ingress router1 (edge router) of AS1. At the outgoing interfaces of the edge router1 the First Flag is set to 1. First Flag will be set on to represent the first AS and help ASBR1 to filter out spoofed packets in cooperation with AS Dist field. Setting the *First flag* on in the ingress router prevents the attacker from pretending to be from different AS, as the edge router of the attacker's network will set the First_flag on. AS_Dist is incremented by one at each ASBR until the ASBR marks the packet. Thus AS_Dist is the distance between the first AS and the AS that carried out the marking. First_flag and AS_Dist perform an important job in our technique. Once the packet is received by ASBR1, an essential check is performed before the marking process. Since the *First_flag* will be set to 1 from the edge router1 and the AS Dist will be zero for the first AS, packets with First flag unset and AS Dist equal to zero will be filtered out. When a packet is unmarked by the first ASBR, it could be marked by any other ASBR on the path. Any packet received by the victim will indicate how far the packet has traversed from the AS of the attacker before it was marked. ASBR1 will mark packets with its ASN according to its marking probability. If ASBR1 decides to write its ASN,

the hash of ASN of ASBR1 is stored in the ID field. In order to prevent any ASBR in the transit network between the attacker and the victim from overwriting the packet, G_flag , which is a one bit flag, will be set to 1 after marking the packet by any ASBR. Once the packet leaves the ASBR1, the *First_flag* is set to zero and *AS_Dist* incremented by 1. Once the victim receives the packet, the AS can be specified from the ID field of the IP header. Any ASBR will follow the same check and marking process. Within 3 marked packets, our technique, will identify the originating AS with probability of 99% for a distance of 6 hops (maximum number is 6 hops [12]). The marking algorithm is shown in Fig.2.



Figure 1. Network topology

The marking probability that will be used at each ASBR is based on number of hops from the current ASBR to the AS of the destination. Other proposed techniques make the probability based on the packet TTL [10, 15]. However, this is likely to be inaccurate as the TTL varies depending upon end-system. This work determines the AS hops count directly from the BGP table. Using dynamic marking probability over a fixed probability would be advantageous as it will reduce number of packets are required to reconstruct the path and provide the fairness in marking. Using dynamic probability, where P=1/d and d represents the travelling distance, would be an advantage if the distance is not known in advance. However, if the distance is known in advance such as in AS routing, using 1/d will not be the optimal choice to our design. The proposed probability is P=1-(1/d), where d is equal to number of hops the AS will make to reach the last AS which contains the required destination. The advantages of this marking probability are that: it gives more emphasis to the first AS which it is most desirable to mark; it does not mark every packet at each AS.

ID	G_flag	AS_Dist	First_flag
20 bits	1 bit	3 bits	1 bit

Variables: Let C be a counter for inside_hops Let f be the First Flag Let AC be a counter for AS_Distance field Let ASN current be the current ASN Let d be number of hops from the current AS to the last AS on the path Let P be the marking probability for ASBR $P = (1 - \frac{1}{d})$ Let ${\cal G}=0$ the flag to be set in case of marking Begin 1: for each Packet do 2: if (outgoing interface of Ingress router) then f = 13: end if 4: // Note: First ASBR f=1 5: if f = 1 then if AC = 0 then 6: $P = (1 - \frac{1}{d})$ 7: 8: f = 0if (random < P) then 9-Write ASN_current in ID field 10: 11: set G = 112: else 13: AC = AC + 1d = d - 114: 15: end if else 16: 17: drop the packet 18: end if else 19: 20: // Note: Transit ASBR f=0 if f = 0 then 21: 22: if (G = 1) then 23: AC = AC + 1d = d - 124: 25: else if (random < P) then 26 27: Write ASN_current in ID field 28: set G = 129: else AC = AC + 130: d = d - 131: 32: end if 33: end if 34: end if 35: end if 36: end for End

Figure 2. POAST algorithm

4.3. Performance and Analysis

One of the key issues in marking the packets is the marking probability that will be used by the routers. PPM proposed to use fixed probability by all routers along the attack path. Using the fixed probability leads to unwanted results such as high false positives rate which is caused by spoofed packets and unfairness marking due to overwriting problem. Here most of the effort will be put on using a marking probability that will insure a high rate of marking by the far ASBR (first ASBR) from the victim. Once the packet being marked by the first ASBR, the G_{flag} will be set on and none of the transit AS's could mark the packet. Therefore, overwriting problem is not an issue in our approach. Number of marked packets is increased with the new used probability. Increasing the number of marked packets will reduce the uncertainty that would be caused by unmarked packets. However, it will increase the overhead in total for the routers. That is the key advantage of our technique. It is designed to mark many packets based on high

probability and reduce the overhead as much as possible. In order to reduce the overhead, G_{flag} is used, which is a flag to prevent any other ASBR from overwriting the packet. Therefore, the marking is carrying only once, by one of the ASBR's, which is reducing the counted overhead. Note that this technique does not mark packets destined within the same AS. The victim can be aware of this through the lack of packet marking and intra-domain traceback or DoS mitigation would be required. The required number of packets to identify the AS of the attacker is reduced comparing to PPM and AAST. Fig. 3 shows the comparison between our technique and other techniques. It is clear that number of packets is highly reduced.



Figure 3. Comparison for required number of packets of different techniques

5. EFFICIENT AS TRACEBACK

In this section, our second idea is proposed. It will traceback the attack to the originating AS with a *full AS path*. The difference in this technique from the early presented (POAST) is in the marking probability and the design. This technique will provide the full AS path for the victim and use different marking probability. The proposed marking probability will efficiently mark the packets and dynamically changed.

5.1. Available Space and Assumptions

5.1.1. Available Space

Our technique will overload the IP header with the same suggested fields in 4.1.1.The assumptions are similar to most of other traceback techniques [2, 16] with some modifications to fit our design.

5.1.2 Assumptions

Our assumptions can be summarized in the following:

- Autonomous system border routers (ASBRs) are not compromised.
- Single or multiple attackers could launch the attack.
- Attackers may be aware of being traced.
- Stable routes exist between autonomous systems (ASs).

- Packets could be routed differently due to ASs policies.
- Attack could be from any type of packets.

5.2. Proposed Technique

Efficient AS traceback (EAST) is based on marking the packets probabilistically. PPM uses a static marking probability p = 0.04 over all of the routers along the attack path. In PPM, the chances for the packets, marked by the closest router to the attacker to reach the victim without being overwritten by the downstream routers, are very low. Therefore, PPM requires a high number of packets to reconstruct the attack path as the victim has to wait for the marked packets from the earlier router on the attack path. Dynamic PPM uses dynamic marking probability p= 1/d, where d is the travelling distance. The travelling distance was deduced from the TTL value in the IP header by most of the proposed Dynamic PPM techniques [10, 15, 17]. All of the dynamic PPM TTL-based techniques are suffering from two drawbacks. First, the initial TTL value is system dependent which means the value would be changing based on the used system. Second, the attacker can intentionally inject packets with different TTL to confuse the technique. Fig. 4 shows a comparison between using PPM with fixed and changing probability for all routers along the attack path. It is clear that using a changing probability for each router along the attack path would give a significant difference over using a fixed probability. Our technique will mark the packets probabilistically with a dynamic marking probability. However, our marking probability will be p = 1/(a-2), where a is the number of ASs from the AS of attacker to the AS of the victim. Since our technique performs traceback at the AS level, a can be known in advance and this helps optimize the marking probability. Therefore, the attacker will not be able to influence the value of a as it is derived in the ASBR from BGP table, unlike other techniques. Our technique will use node sampling as a marking algorithm unlike PPM where the edge sampling is used as the marking algorithm.



Figure 4. PPM with fixed and dynamic probability.

BGP is used to route packets between ASs via autonomous system border routers (ASBRs). Each ASBR will have a full reachability table for other ASBRs. For any received packets by ASBR, the BGP table will be checked and a full path to the required destination, is provided. For illustration purpose, the example in Fig.1 will be used for explaining our technique. The attacking packets will leave the AS_source from the ASBR_Src. Once the ASBR_Src receives a packet, BGP will be checked and the reachability path is extracted. The marking probability P is calculated from the extracted path. Based on the marking probability, ASBR_Src will decide if the packet is going to be marked or not. If the ASBR_Src decides to mark the packets, the hash

value of the 32 bit AS number is injected inside the AS_ID field which is 22 bits. Distance is a 3 bit field which keeps track of travelling distance and starts from the AS that marks the packet to the destination AS. The distance field will reset to zero by the marking router and increment by 1 for non-marking ASBR. The distance field will help the victim in ordering the samples for path reconstruction. Overwriting is allowed by downstream ASBRs and if a packet is unmarked by ASBR_Src, it has to be marked by downstream ASBRs such as ASBR1 or ASBR2 or ASBR3. The required fields for our technique can be shown in Table 2 and the proposed marking algorithm presented in Fig. 5. By design, number of unmarked packets is significantly reduced comparing to PPM and AS_PPM(see performance and analysis section).

Table 2. Required Fields for EAST.

AS_ID	AS_Dist
22 bits	3 bits

Variables:

```
Let ID_field is 22 bits.
   Let Dist be 3 bits for AS hops distance.
   Let H(ASBR) is the 22 bit hash of the 32 bit ASBR number.
   Let a is the distance to the destination AS.
   Begin
 1: for each Packet do
     if (ASBR) then
2:
3:
        p = \frac{1}{a-2}
        if (random number < p) then
4:
           ID_field = H(ASBR)
5:
6:
           Dist = 0
        else
7:
           Dist + +
8:
9.
        end if
10:
      end if
11: end for
   End
```

Figure 5. EAST algorithm

5.3. Performance and Analysis

In order to design a traceback technique that performs better than PPM in terms of number of required packets to reconstruct the path and uncertainty, the marking probability has to be chosen carefully. We started with the choice of dynamic marking instead of fixed probability. Simulation results in Fig.4 show that using dynamic marking is advantageous over static marking. Then we have studied how to improve and choose the right dynamic marking probability. The problem with most of proposed techniques is the distance. Some techniques assume fixed probability and some deduce the distance from TTL field. Here, we are taking the advantage of the AS full path where the distance is known in advance. According to Muhlbauer [12], in 99% of cases, a packet traversing from source to destination passes 6 ASs or less. We believe that knowing half of the AS path would be enough to determine the AS of the attacker. We propose to exclude the last two AS out of the marking probability considerations since the AS of the victim is part of the AS full path. Indeed, simulation in Fig.6 shows a better performance than PPM and AAST by using a marking probability of p = 1/(a-2). The required number of packets to reconstruct the attack path for EAST is less than AAST and PPM. Unmarked packets are significantly reduced comparing to PPM and AAST. Fig.7 shows number of unmarked packets by each technique. It is clear that there are no unmarked packets with EAST and this is due to the design of marking probability. A comparison between EAST and Dynamic PPM (DPPM) is shown in Fig.8. It is clear that EAST

converges faster than DPPM with the proposed marking probability. The value of p was also tested experimentally to be the best.



Figure 6. EAST and other techniques comparison



Figure 7. Comparison for unmarked packets



Figure 8. EAST and Dynamic PPM comparison

6. CONCLUSION

DoS is considered to be a high security threat in the internet today. To traceback the origin of the attack is a challenging process as the attacker could spoof his identity. Here two AS based packet marking techniques have been proposed. Prevent overwriting AS traceback (POAST) has been shown that it has advantages over PPM in that fewer routers are involved in the marking process, thus reducing router overhead, and additionally intra-domain topology information is not disclosed both of which are desirable features for network operators. The system can uniquely inform the victim of the originating AS and this information is given high emphasis through high marking probability in the first AS hop. The marking probability is investigated and a dynamic marking probability has been suggested which is based on AS hop count that is determined from BGP tables. In order to traceback the AS of the attacker with a full path, efficient AS traceback (EAST) has been proposed. While there have been a number of attempts at overloading the current IP forwarding process to provide traceback, they have significant disadvantages. The EAST technique proposed here circumvents many of the disadvantages of other techniques in that it reduces the number of routers involved in the marking and converges to a valid traceback value in fewer packets. Moreover, EAST eliminates unmarked packets which could be used by the attacker to confuse the traceback technique.

REFERENCES

- [1] ArborNetworks, "Worldwide Infrastructure Security Report," 2008.
- [2] S. Savage, et al., "Network Support for IP Traceback " IEEE/ACM Transactions on Networking (TON), vol. 9, JUNE 2001.
- [3] S. M. A. Belenky and N. Ansari, IEEE, "IP Traceback With Deterministic Packet Marking," IEEE COMMUNICATIONS LETTERS, vol. 7, APRIL 2003.
- [4] H. Burch and B. Cheswick, "Tracing anonymous packets to their approximate source," in LISA Dec 2000, pp. 319-327
- [5] M. L. Steve Bellovin, Tom Taylor, "ICMP Traceback messages," Network Working Group, Internet Engineering Task Force (IETF)February 2003.
- [6] M. T. Goodrich, "Probabilistic packet marking for large-scale IP traceback," Networking, IEEE/ACM Transactions on, vol. 16, pp. 15-24, 2008.
- [7] W. Z. Shui Yu, Robin Doss, Weijia Jia, "Traceback of DDoS Attacks using Entropy Variations " IEEE Transactions on Parallel and Distributed Systems, vol. 22 pp. 412 - 425, 20 May 2010 March 2011.

- [8] H. ALJIFRI, "IP Traceback: A New Denial-of-Service Deterrent?," Security & Privacy, vol. 1, pp. 24-31, 2003.
- [9] X. Wang, et al., "A Fast Deterministic Packet Marking Scheme for IP Traceback," in International Conference on Multimedia Information Networking and Security, Hubei, China 2009, pp. 526-529.
- [10] J. Liu, et al., "Dynamic probabilistic packet marking for efficient IP traceback," Computer Networks, vol. 51, pp. 866-882, 2007.
- [11] V. Paruchuri, et al., "Authenticated autonomous system traceback," in 18th International Conference on Advanced Information Networking and Application (AINA'04), 2004, pp. 406-413 Vol. 1.
- [12] W. Mühlbauer, et al., "Building an AS-topology model that captures route diversity," ACM SIGCOMM Computer Communication Review, vol. 36, pp. 195-206, 2006.
- [13] B. Zhang, et al., "Collecting the internet AS-level topology," ACM SIGCOMM Computer Communication Review, vol. 35, pp. 53-61, 2005.
- [14] G. Huston, "Exploring Autonomous System Numbers," The Internet Protocol Journal vol. Volume 9,Number 1, pp. 2-23, March 2006.
- [15] A. D. a. S. C. Vamsi Paruchuri, "TTL based Packet Marking for IP Traceback," presented at the GLOBECOM '08 New Orleans, LO, 2008.
- [16] Z. Gao and N. Ansari, "A practical and robust inter-domain marking scheme for IP traceback," Computer Networks, vol. 51, pp. 732-750, 2007.
- [17] J. B. Hongcheng Tian, Xiaoke Jiang and Wei Zhang, "A Probabilistic Marking Scheme for Fast Traceback," in 2nd International Conference on Evolving Internet, Valencia. Spain, 2010.

Authors

Mohammed Alenezi received his B.E. in Computer Engineer from Kuwait University, College of Engineering & Petroleum in 2001. He got his Master degree in Information Networks from Essex University, UK, in 2008. Currently, he is a PhD student at Essex University.

Dr Martin Reed is a Senior Lecturer in the School of Computer Science and Electronic Engineering at the University of Essex in the UK. His main research interests are network control, network security and audio/video media transport. He has been involved in, a number of EU or EPSRC research funded projects in these areas. Current projects include the EU FP-7 funded project PURSUIT and a project with BT in the area of audio transport over networks.