Netw Sci (2013) 2: 42–51
DOI 10.1007/s13119-012-0007-x

RESEARCH ARTICLE

An adaptive probabilistic marking scheme for fast and secure traceback

Hongcheng Tian, Jun Bi (✉), Xiaoke Jiang

Network Research Center, Department of Computer Science, Tsinghua University, Tsinghua National Laboratory for Information Science and Technology (TNList), Beijing 100084, China

Received: 18 September 2011/Revised: 29 December 2011/Accepted: 13 February 2012

Abstract IP traceback can be used to find direct generator(s) and path(s) of attacking traffic. Probabilistic marking schemes, as one type of IP traceback technologies, have been most studied, but they are difficult to fast reconstruct attacking path(s) and defend against spoofed marks generated by attacking source(s). In this paper, we present Adaptive Probabilistic Marking scheme (APM). In APM, when each packet enters the first-hop router, its TTL value is set to a uniform value, and when it is forwarded by routers in the network, each intermediate router decreases the TTL value by one. Consequently, each intermediate router may infer the router-level hop number that each packet has already traveled, and then correspondingly marks the packet with the probability inversely proportional to the router-level hop number. APM is focused on the probability with which a router marks a packet, and APM can cooperate with other probabilistic marking schemes. NS2 simulation experiments prove that, in APM, the time for the victim to receive necessary marks for the path reconstruction is reduced by more than 20% compared with existing probabilistic marking schemes, and spoofed marks cannot reach the victim and influence the traceback process.

Keywords IP traceback, adaptive, probability, marking

1 Introduction

One of the deficiencies of TCP/IP is that the validity of the source address in the IP header is not checked in the Internet. A packet routed is dependent entirely on its destination address. Thus, attacking sources often forge source addresses to escape detection, such as SYN flooding [1], DNS amplification [2], Smurf [3], etc. But it is difficult for victims to block the attack in real time, precisely locate attacking source(s) and pursue legal actions.

The objective of IP traceback [4] is to find the source(s) and attacking path(s) of malicious traffic. Traceback is executed with the assistance of a series of routers. Traceback can also collect statistics for packets' forwarding path(s) in the Internet in order to optimize router configuration, which benefits the research in the area of traffic engineering.

The technology of probabilistically marking packets, as one kind of traceback technology, is much studied in academic circles. Savage et al. [4] have originally proposed Probabilistic Packet Marking (PPM), where each router marks each packet with a fixed probability. In PPM, (1) the victim requires many packets for the path reconstruction, slowing down IP traceback; (2) it is difficult for PPM to defend against spoofed marks, resulting in uncertainty of the path reconstruction. Paruchuri et al. [5] have proposed TTL-based Packet Marking (TPM) and Liu et al. [6] have proposed Dynamic Probabilistic Packet Marking (DPPM), attempting to solve the above two problems introduced by a fixed marking probability in PPM. However, the two approaches only partially improve the performance of PPM in the presence of attacks which spoof TTL values and marks inside IP packets.

In this paper, we present Adaptive Probabilistic Marking (APM) based on the TTL field, where the TTL value of each packet is modified to a uniform number at the first-hop router, and each router can deduce the traveling distance (in router-level hops) of each arriving packet from its source, and then adaptively marks it with the probability

E-mail: junbi@tsinghua.edu.cn

inversely proportional to the traveling distance. Theoretical analysis shows that, in APM, the packet number, that the victim is required to receive for a successfully traceback, is the fewest, and spoofed marks cannot affect the traceback result. NS2 simulation experiments show that, in APM, the time for a victim to collect obligatory marks for the path reconstruction is reduced by more than 20% compared with other schemes, and spoofed marks cannot be received by the victim and cannot disturb the path reconstruction process. APM can be incorporated in other probabilistic marking techniques, such as PPM [4] and the randomize-and-link approach [14]. APM has the following two advantages: (1) the victim requires the fewest packets for the path reconstruction, speeding up IP traceback; (2) APM can eliminate the effect of spoofed marks on the victim.

If the attacking source(s) and attacking path(s) are reconstructed fast, it is helpful to alleviate or eliminate the attack on the victim in time for the following reasons:

(1) The information collected by traceback system can help some ISPs, which attack traffic transits, to react against the attack. Concretely, if there exist attack countermeasure systems (for example, using traffic cleaning solution) in some of these ISPs, traceback system may supply the features of attacking traffic (or attacking packets) for these countermeasure systems (the upstream countermeasure systems are supplied preferentially rather than downstream ones). This is helpful to alleviate or eliminate attacks on the victim in time. Alternatively, traceback system may supply the features of attacking traffic (or attacking packets) for these ISPs (the upstream ISPs are supplied preferentially, too), and then actions (such as packet filtering or traffic constraint) are taken in time on the relative routers of the ISPs against the attack. This may alleviate or eliminate the attack on the victim in time.

(2) Network administrators, who are responsible for the network which attacking source(s) reside(s) in, may be informed of dealing with the attacking source(s) in time, for example, terminating the network connection(s) of attacking source(s).

The rest of the paper is organized as follows: In Section 2, we review traceback literature. Section 3 presents APM. Section 4 and Section 5 compare the performance of PPM, TPM, DPPM and APM from theory and experiment, respectively. In Section 6, we discuss the incremental deployment of APM. And in Section 7, conclusions and our future work are given.

2 Related work

Researchers have proposed various approaches to trace attacking packets back to source(s) of attacking traffic.

2.1 General background

IP traceback technologies can be categorized as six types: link testing, packet marking, ICMP traceback, packet logging, hybrid IP traceback and overlay network for IP traceback. The first type belongs to real-time approaches for IP traceback. The remains can traceback not only in real time but also post mortem. As far as the authors know, there is no Internet-level IP traceback system that is currently deployed.

Link testing can be subcategorized as two types: input debugging and controlled flooding [16]. Input debugging takes advantage of a function of routers, which can identify the input link of attacking packets with a certain signature. The attack signatures are extracted by the victim from attacking packets and are sent to the victim's upstream router, where the input port of attacking packets can be identified. This process is repeated recursively over hop by hop upstream routers. But the operation overhead of input debugging is high. Controlled flooding floods each link of a router with large bursts of traffic and observes changes in the rate that invading packets reach the victim. When the rate of invading packets is reduced, the link the attacking packets come from can be inferred. Controlled flooding has the following shortcoming: (1) the bandwidth overhead is high; (2) under a DDoS attack, false positives or false negatives are high.

Packet marking schemes [4 – 15] are that routers along attacking path(s) mark packets with partial path information, and the victim extracts path information from marks of receiving packets to reconstruct attacking path(s). The shortcomings of packet marking schemes are: (1) false positives or false negatives are high; (2) computing overhead is high at the victim; (3) under a DDoS attack, the packet number required for IP traceback is large.

ICMP traceback [17 – 19] is that when forwarding packets, routers can (with a low probability) send some ICMP traceback messages with some path information to the destination. The destination receives ICMP traceback messages and reconstructs the attacking path(s). The weakness of this approach is that the ICMP traceback message may be spoofed or filtered in the Internet, resulting in high false positives or false negatives.

Packet logging schemes [20 – 22] are that intermediate routers log packet digests, and the source of a single IP packet can be traced back by recursive queries. Packet logging schemes suffer from the high computing and storage overhead at routers.

Hybrid IP traceback [23 – 26] is designed to make use of advantages of packet marking and packet logging schemes and alleviate their weaknesses, but relevant management mechanism is very complicated.

In overlay network for IP traceback [27], tunnels are made between border routers and a central traceback router. And the central traceback router connects to an IDS, which checks whether receiving packets are attacking ones or not. If so, the central traceback router knows which border router the attacking packets come from. The shortcoming of overlay network for IP traceback is that the computing overhead of the central traceback router is high.

2.2 Probabilistic marking schemes

Probabilistic marking schemes have been much studied [4 – 15]. The probability used by intermediate routers to mark packets plays an important role in packet marking schemes. Due to the nature of probability, attacking packets may arrive at the victim without being marked by intermediate routers. And crafty attacking source(s) may send packets with spoofed marks to compromise the traceback process. Paruchuri et al. [5] and Liu et al. [6] attempted to solve the problems introduced by the nature of probability.

TPM has been proposed by [5], in order to reduce the effectiveness of spoofed packets. But TPM has a serious shortcoming when TPM is deployed in the Internet, in which case the first-hop router can't be reconstructed and spoofed packets may reach the victim unmarked.

DPPM has been presented by [6], attempting to precisely pinpoint the attacking source(s) even under spoofed marking attacks. But DPPM doesn't work well enough when attacking sources cunningly set TTL values and spoof marks in packet headers, in which case the path reconstruction process may be significantly confused.

3 Adaptive Probabilistic Marking scheme

In this section, we present APM, an adaptive marking scheme based on the TTL field, in order to minimize the packet number required for the attacking path reconstruction and eliminate the effect of spoofed marks on the victim. We first define several concepts and a Lemma, then propose the design goals, and finally present the APM scheme. APM is focused on the selection of marking probabilities of routers, not concerned about concrete marking contents.

3.1 Assumption, definitions and lemma

We identify the following assumptions that motivate and constrain the design of APM:

- Attacking sources may generate any packet.
- Attacking sources may know they are being traced back.
- Some attacking sources may conspire.
- Routers are not compromised.

Figure 1 describes an attacking scenario aimed at a victim V. V may be a host, a NAT or a firewall. Attacking source(s) may be a host, or a group of hosts distributed in various locations (such as A_1, A_2, A_3 or A_4). *Attacking path* is defined as an ordered list of routers from an attacking source to a victim. For example, in Fig. 1, from an attacking source A_2 to V, the attacking path is (R_2, R_5, R_7), as shown by a red line. When the victim faces a DDoS attack, different attacking paths from different attacking sources form an attack tree rooted at the victim, in which each attacking source is a leaf node. Assume that an attacking packet traverses d routers from A to V, the attacking path η is (R_1, R_2, …, R_i, …, R_d)($1 \leqslant i \leqslant d$). In the following, we will carry out our research on the attacking path η.

Let p_i represent the *marking probability* of router R_i for an attacking packet. In the attacking path η, downstream routers may overwrite the marks of packets left by upstream routers. Define the *reaching probability* for router R_i, denoted by r_i, to be the one that an attacking packet has been lastly marked by router R_i but has not been re-marked by other routers downstream on path η towards victim V. In other words, reaching probability for router R_i is the one that the marking information for R_i can reach the victim. It can be shown that

$$r_i = \begin{cases} p_i \prod_{k=i+1}^{d} (1 - p_k) & \text{for} \quad 1 \leqslant i \leqslant d-1 \\ p_d & \text{for} \quad i = d \end{cases} \tag{1}$$

Define the *unmarked probability*, denoted by U, to be the one that a packet arrives at the victim without having been marked by any router in the path η, which is expressed as

$$U = \prod_{k=1}^{d} (1 - p_k) \tag{2}$$

Lemma 1: if the reaching probability for each router in the attacking path η is equal to $1/d$, the number of packets required by the victim for a successful traceback is the least.

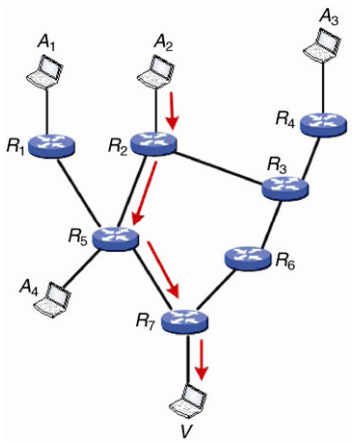

Fig. 1 An attacking scenario aimed at a victim V

Proof: note that the attacking path η consists of d routers. This proposition is equivalent to one of the coupon-collector's problems: If each type of coupon is randomly selected from d distinct types of coupons with an equal probability, the number of selections is the least for collecting all d distinct types of coupons, which has been proved in [28]. In other words, that the mark information from each router is equiprobably received by the victim is similar to that each type of coupon is selected with an equal probability.

3.2 Design goals

If a number of unmarked packets reach the victim, the victim will take additional time to receive more packets so as to obtain enough marking information for the path reconstruction, and countermeasures based on reconstructed attacking path(s) and source(s) cannot be taken in time.

On the other hand, Park and Lee [29] have shown that, in PPM, a proportion of spoofed marks will arrive at the victim, resulting in the uncertainty of the path reconstruction. In other words, the attacking source cannot be identified precisely. Furthermore, the uncertainty is significantly amplified under a DDoS attack. In TPM and DPPM, when attacking sources intelligently set TTL values of packets, a proportion of spoofed marks will also reach the victim, disturbing the reconstruction process.

Therefore, we hope that a newly designed marking scheme should achieve the following design goals:

- Celerity. Packets as few as possible are required by the victim for the path reconstruction, speeding up IP traceback. Therefore, actions can be taken in time along the attacking path η against the attack.
- Security. Spoofed marks as few as possible can reach the victim, lessening the effect on the path reconstruction.

3.3 Proposed scheme

APM can achieve the two design goals mentioned above. APM works in the following way: the TTL value of each packet is set to a uniform number (such as 255) at the first-hop router, and each router deduces the traveling distance (in router-level hops) of each arriving packet from its source, and then adaptively marks it with the probability inversely proportional to the distance. For the given attacking path η, let h $(1 \leqslant h \leqslant d)$ be the traveling distance of an attacking packet from the attacking source A to router R_i $(1 \leqslant i \leqslant d)$. Obviously, h=i, router R_i adaptively chooses the probability

$$p_i = \frac{1}{h} = \frac{1}{i} \qquad \text{for} \quad 1 \leqslant i \leqslant d \tag{3}$$

to mark the packet. The marking procedure is described in Fig. 2.

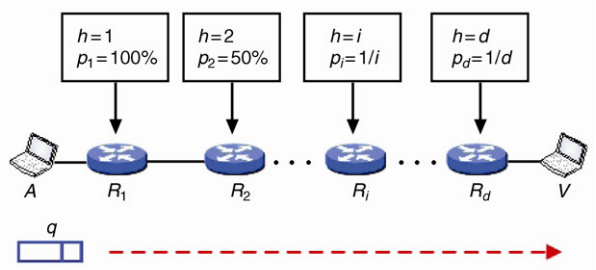

Fig. 2 When an attacking packet q travels from the attacking source A to the victim V (shown as a red dashed line), its TTL value is set to a uniform number (such as 255) at the first-hop router R_1, and each router deduces its traveling distance (in router-level hops) from its source, and then adaptively marks it with the probability inversely proportional to its traveling distance. Let h and p_i represent the traveling distance of q and the marking probability of R_i for q, respectively

The reaching probability for each router along the attacking path η is computed through Eq. (1) and Eq. (3):

$$r_i = \frac{1}{d} \qquad \text{for} \quad 1 \leqslant i \leqslant d \tag{4}$$

Therefore, each router along the attacking path η has a same reaching probability of $1/d$. According to the Lemma 1, the victim requires the fewest packets to reconstruct the attacking path η, and actions can be taken fast along the attacking path η against the attack. Thus, APM meets the first goal optimally.

The unmarked probability (U_{APM}) is calculated in terms of Eq. (2) and Eq. (3):

$$U_{\text{APM}} = \prod_{k=1}^{d}(1-p_k) = (1-1)\left(1-\frac{1}{2}\right)\cdots\left(1-\frac{1}{d}\right) = 0 \tag{5}$$

Since the unmarked probability is zero in APM, spoofed marks will be all overwritten by intermediate routers in the attacking path η. Consequently, APM meets the second goal optimally, too.

Thus, APM meets the two design goals mentioned in Section 3.2. From Eq. (3), it can be seen that the marking probability of a router only depends on the traveling distance of a packet from its source. A key question must be answered: How can a router deduce the traveling distance (in router-level hops) of an arriving packet from its source? We will answer this in the following subsection.

Determination of the traveling distance

It is known to all of us that the TTL field in the IP header concerns the traveling distance of a packet from its source. As a packet is forwarded by routers in the network, each router decreases the TTL value by one. Routers drop any packet with a TTL value of zero. Therefore, if a router knows

the initial TTL value of a packet, the traveling distance of the packet from its source could be computed accordingly. But different operating systems and protocols set different initial TTL values for newly generated packets [3]. Therefore, when a router receives a packet, it is difficult to identify its initial TTL value. However, if each packet uses a same initial TTL value, this will take on a new look.

We consider: (1) it is meaningless that different operating systems and protocols use different initial TTL values for the same Internet; (2) the initial TTL value can be manually modified in operating systems. For examples, for Windows systems the initial TTL value can be modified in the Registry, and for Linux and UNIX in the configuration file; (3) a malicious attacking source can forge the initial TTL value randomly.

Thus, we propose to modify the TTL value of each packet to a uniform value at the first-hop router, such as 64, 128 or 255. Subsequently, each intermediate router can identify the traveling distance of the packet from its source by computing the difference between 255 (64 or128) and the TTL value of the packet, and then mark the packet with the probability inversely proportional to the traveling distance.

We can use the following mechanism to enable a router know it is the first hop in itself. If a port of the router is connected to an access network, we can configure the router in order to let the router know it is the first hop for the access network. When packets are sent into the port from the access network, the TTL value of each packet is set to a uniform number. Subsequently, routers along the attacking path η can deduce the traveling distance of each packet from its source, and then adaptively choose marking probability.

In addition, APM has the effect on some applications, such as tracert tool, IP multicast and so on. IP multicast may use TTL values of multicast packets to control multicast scope. In order to solve these issues, there is a processing rule: firstly, relative packets of these applications are identified based on their respective features; secondly, the special process is made on these packets. In the following portion, we will take tracert and IP multicast for example. As we all know, the tracert tool is concerned with three types of messages: ICMP Echo Message, ICMP Echo Reply Message and ICMP Time Exceeded Message [30]. And multicast packets use dedicated multicast address within the range from 224.0.0.0 to 239.255.255.255 [31]. We may deploy the detection software at each first-hop router's port to the access network. When the detection software detects the ICMP Echo Messages or multicast packets generated from the access network, the first-hop router doesn't set their TTL values to a uniform value, but marks these packets with the probability 100%. This measure eliminates the effect of APM on the tracert tool, IP multicast and so on.

APM marking algorithm

The APM marking algorithm is shown in Fig. 3, where t is the TTL value of a packet, t_u is the unified initial TTL value, and h is the deduced traveling distance of the packet from its source. We may choose $t_u = 255$. To elaborate, the router marks the packet with the probability of $1/h$. Thus, a packet that has traveled a short distance is marked with a higher probability, while a packet which has traversed a long distance is marked with a low probability. For the given attacking path η, the sequence of the marking probabilities for intermediate routers is 1, 1/2, …, $1/d$. In Fig. 3, line 3 and line 4 eliminate the effect of APM on the tracert tool and IP multicast.

```
1  for each packet
2   if it goes into a first-hop router then
3    if it is an ICMP Echo message or multicast packet then
4      mark the packet;
5    else
6      t ← t_u;
7   h ← t_u − t + 1;
8   t ← t − 1;
9   let r be a random number in [0,1);
10  if r <= 1/h
11    mark the packet
```

Fig. 3 APM marking algorithm

4 Theoretical analyses

In the following portion, we will analyze PPM [4], TPM [5] and DPPM [6], and then compare three methods with APM in terms of the reaching probability and the unmarked probability.

4.1 Analysis of Probabilistic Packet Marking

In PPM, the marking probability of each router along the attacking path η is constant. The following analysis is based on the Compressed Edge Fragment Sampling Algorithm [4], which implements PPM well.

In PPM, the marking probability for each router along the attacking path η is

$$p_i = p \qquad 1 \leqslant i \leqslant d \tag{6}$$

Since the edge-id of each edge is just used to reconstruct the address of the start router of the edge, the reaching probability for the start router is the one for the edge-id. Thus, according to Eq. (1) and Eq. (6), the reaching probability for router R_i is

$$r_i = p(1-p)^{d-i} \qquad 1 \leqslant i \leqslant d \tag{7}$$

In terms of Eq. (2) and Eq. (6), the unmarked probability (U_{PPM}) is computed:

$$U_{\mathrm{PPM}} = \prod_{k=1}^{d}(1-p_k) = (1-p)^d \tag{8}$$

The average path length is around 16 in the Internet [32], we may choose $d = 17$. And $p = 1/25$ is recommended as the optimal choice for PPM [4]. Thus,

$$U_{\mathrm{PPM}} = (1-p)^d = \left(1-\frac{1}{25}\right)^{17} \approx 50\% \tag{9}$$

4.2 Analysis of TTL-based Packet Marking

TPM is focused on the marking probability of routers, which is the same as DPPM and APM. In TPM, the marking probability for each packet is determined based on the TTL value of a packet. TPM marking algorithm is depicted by Paruchuri et al. [5].

According to the initial TTL value of a packet, denoted by t_{ini}, TPM marking algorithm is analyzed in the following two cases:

(1) $t_{\mathrm{ini}} < 24$. The marking probability for a packet at router R_i in the attacking path η is

$$p_i = \frac{1}{24+i-t_{\mathrm{ini}}} \qquad 1 \leqslant i \leqslant d \tag{10}$$

According to Eq. (1) and Eq. (10), the reaching probability for router R_i is

$$r_i = \frac{1}{24+d-t_{\mathrm{ini}}} \qquad 1 \leqslant i \leqslant d \tag{11}$$

In terms of Eq. (2) and Eq. (10), the unmarked probability (U_{TPM}) is computed:

$$U_{\mathrm{TPM}} = \prod_{k=1}^{d}(1-p_k) = \frac{24-t_{\mathrm{ini}}}{24+d-t_{\mathrm{ini}}} \tag{12}$$

The average path length is around 16 in the Internet [32], and an attacking source can spoof TTL values and marking information to confuse the path reconstruction at the victim. For example, we may choose $d = 17$ and $t_{\mathrm{ini}} = 18$. Thus,

$$U_{\mathrm{TPM}} = \frac{24-18}{24+17-18} \approx 26\% \tag{13}$$

(2) $t_{\mathrm{ini}} \geqslant 24$. The marking probability for a packet at router R_i in the attacking path η is

$$p_i = \begin{cases} 1 & i = 1 \\ \dfrac{1}{i\text{-}1} & 2 \leqslant i \leqslant d \end{cases} \tag{14}$$

According to Eq. (1) and Eq. (14), the reaching probability for router R_i is

$$r_i = \begin{cases} 0 & i = 1 \\ \dfrac{1}{d-1} & 2 \leqslant i \leqslant d \end{cases} \tag{15}$$

In terms of Eq. (2) and Eq. (14), the unmarked probability (U_{TPM}) is calculated:

$$U_{\mathrm{TPM}} = (1-1)(1-1)\cdots\left(1-\frac{1}{d-1}\right) = 0 \tag{16}$$

4.3 Analysis of Dynamic Probabilistic Packet Marking

DPPM is focused on the marking probability of routers, which is the same as TPM and APM. In DPPM, the marking probability for each packet is computed based on the TTL value of a packet. DPPM scheme is described by Liu et al. [6].

According to the initial TTL value of a packet, denoted by t_{ini}, DPPM scheme is analyzed in the following two cases:

(1) $t_{\mathrm{ini}} < 32$. The marking probability for a packet at router R_i in the attacking path η is

$$p_i = \frac{1}{32+i-t_{\mathrm{ini}}} \qquad 1 \leqslant i \leqslant d \tag{17}$$

According to Eq. (1) and Eq. (17), the reaching probability for router R_i is

$$r_i = \frac{1}{32+d-t_{\mathrm{ini}}} \qquad 1 \leqslant i \leqslant d \tag{18}$$

In terms of Eq. (2) and Eq. (17), the unmarked probability (U_{DPPM}) is computed:

$$U_{\mathrm{DPPM}} = \prod_{k=1}^{d}(1-p_k) = \frac{32-t_{\mathrm{ini}}}{32+d-t_{\mathrm{ini}}} \tag{19}$$

The average path length is around 16 in the Internet [32], and an attacking source can spoof TTL values and marking information to disturb the path reconstruction at the victim. For example, we may choose $d = 17$ and $t_{\mathrm{ini}} = 18$. Thus,

$$U_{\mathrm{DPPM}} = \frac{32-18}{32+17-18} \approx 45.1\% \tag{20}$$

(2) $t_{\mathrm{ini}} \geqslant 32$. The marking probability for a packet at router R_i in the attacking path η is

$$p_i = \frac{1}{i} \qquad 1 \leqslant i \leqslant d \tag{21}$$

According to Eq. (1) and Eq. (21), the reaching probability for router R_i is

$$r_i = \frac{1}{d} \qquad 1 \leqslant i \leqslant d \tag{22}$$

In terms of Eq. (2) and Eq. (21), the unmarked probability (U_{DPPM}) is calculated:

$$U_{\mathrm{DPPM}} = \prod_{k=1}^{d}(1-p_k) = (1-1)\cdots\left(1-\frac{1}{d}\right) = 0 \tag{23}$$

4.4 Performance comparisons

PPM, TPM, DPPM and APM are compared in terms of the reaching probability and the unmarked probability in Table 1.

The reaching probability is associated with the number of packets which the victim requires for a successful traceback. According to Lemma 1, if the reaching probability for each router is equal to $1/d$, the number of packets for a successful traceback is the least. In PPM, the reaching probability for each router is unequal according to Eq. (7). When TPM is deployed in the Internet, the reaching probability for each router is unequal according to Eq. (15). In DPPM, when the spoofed initial TTL values of packets are smaller than 32, the reaching probability for each router is equal according to Eq. (18), but is not $1/d$. For APM, no matter how an attacking source spoofs the initial TTL values of packets, the reaching probability for each router is equal to $1/d$ in terms of Eq. (4).

The unmarked probability concerns what percentage of spoofed marks at most can reach the victim. In PPM, the unmarked probability (U_{PPM}) is greater than zero according to Eq. (8). In TPM, when the spoofed initial TTL values of packets are smaller than 24, the unmarked probability (U_{TPM}) is greater than zero according to Eq. (12). In DPPM, when the spoofed initial TTL values of packets are smaller than 32, the unmarked probability (U_{DPPM}) is greater than zero according to Eq. (19). In APM, no matter how an attacking source spoofs initial TTL values of packets, U_{APM} is zero in terms of Eq. (5). For example, the average path length is around 16 in the Internet [32], and we choose 17 as the spoofed initial TTL value. Thus, according to Eq. (9), Eq. (13), Eq. (20) and Eq. (5), U_{PPM}, U_{TPM}, U_{DPPM} and U_{APM} are around 50%, 26%, 45.1% and 0, respectively.

Table 1 Performance comparisons

	Reaching probability	Unmarked probability
PPM	Unequal	> 0 (e.g., 50%)
TPM	Unequal, when TPM is deployed in the Internet	> 0 (e.g., 26%), when an attacking source spoofs initial TTL values of packets
DPPM	Equal but is not $1/d$, when an attacking source spoofs initial TTL values of packets	> 0 (e.g., 45.1%), when an attacking source spoofs initial TTL values of packets
APM	$1/d$	0

From Table 1, we can see that APM is superior to other schemes. In APM, the victim requires the fewest packets for a successful traceback, and the attacking path reconstruction is not affected by spoofed marks.

5 Simulation experiments

We carried out the simulation experiments using NS2 and BRITE [33]. We made the experiments in the following way: Firstly, we made use of BRITE to obtain 5 random network topologies in different scales: 50 routers, 100 routers, 200 routers, 500 routers and 1000 routers. Secondly, we made the experiments on PPM, TPM, DPPM and APM in each network topology to take statistics to the time for a victim to collect all the obligatory marks (i.e., the collecting time) and the unmarked percentages. Especially, for PPM, we chose 2 marking probabilities of 0.04 and 0.1 to make the experiments, respectively. We considered: (1) attacking sources could spoof initial TTL values of packets randomly; (2) spoofed initial TTL values of packets should be great enough in order that they can reach the victim. Thus, we arranged the initial TTL values of packets randomly distributed between 11 and 255. For each approach, 500 experiments were made in each network topology to take statistics to the collecting time and the unmarked percentages.

Figure 4 describes the collecting time for each approach in each network topology with 99 percent confidence intervals of sizes of ±3.3 percent. Especially, for TPM, it is very difficult for the victim to collect all the obligatory marks in each network topology, because packets are twice marked with the probability 100% by the first-hop and second-hop routers when these initial TTL values are not smaller than 24. The collecting time for TPM is longer than other approaches. Thus, the collecting time for TPM is not listed in Fig. 4. The experimental results show that, the collecting time for APM is the shortest, which is reduced by more than 20% compared with other approaches.

Figure 5 shows the unmarked percentage for each approach in each network topology with 99 percent confidence

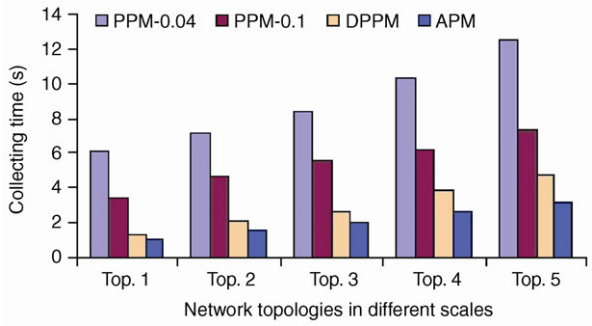

Fig. 4 Collecting time in different marking methods and topologies

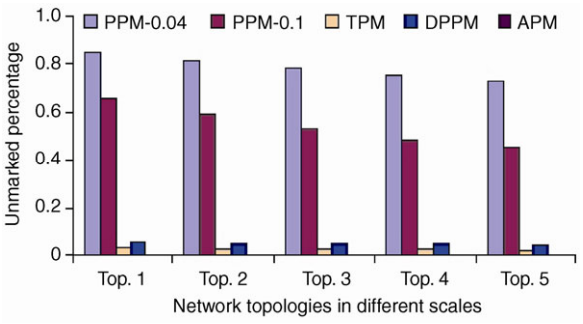

Fig. 5 Unmarked percentage in different marking methods and topologies

intervals of sizes of ±2.8 percent. The results show that, for PPM-0.04 and PPM-0.1, the unmarked percentages are great, for TPM and DPPM small, and for APM zero. So APM is the best to prevent against spoofed marks.

6 Discussions on incremental deployment

For incremental deployment of APM, the key issue is how a deployed router knows whether it itself is the first-hop deployed one for a packet. We discuss two mechanisms of incremental deployment: one is by manual configuration; the other is by a new protocol module.

6.1 A mechanism of incremental deployment

When an AS begins to deploy APM in a small scope of the AS, where all routers are deployed, All border routers are configured to be the first-hop ones. When APM is incrementally deployed in the AS, two rules need to be followed: (1) a legacy router, which is connected to a deployed router through physical link(s), is preferentially deployed; (2) new border routers need to be configured to be the first-hop ones instead of old border ones. If the AS is entirely deployed and a neighboring legacy AS intends to join the deployment, legacy routers of the neighboring AS need to follow the above rules.

For the deployed scope, the analysis of the traceback performance (reaching probabilities for deployed routers and unmarked probability) is the same as Section 3.3.

6.2 Further discussion on incremental deployment

By checking the newly defined IP option, named ciphertext ID, in the IP header, every deployed router can identify whether it itself is the first-hop deployed one for the packet, and can correspondingly decide if the TTL value of the packet is set to the uniform value.

Definition

Along an attacking path, if there exists an ordered list of k_1 deployed routers ($k_1 \geqslant 1$), whose up and down stream neighboring routers (if exist) are non-deployed, we refer to this sequence as *the deployed segment*. If there exists an ordered list of k_2 non-deployed routers ($k_2 \geqslant 1$), whose up and down stream neighboring routers (if exist) are deployed, we refer to this sequence as *the non-deployed segment*.

Thus, an attacking path has three typical cases: (1) it is just a non-deployed segment; (2) it is just a deployed segment; (3) it consists of deployed segment(s) and non-deployed segment(s), which are interleaved with each other. For the first case, the attacking path cannot be reconstructed. For the second, the related traceback performance has been analyzed in Section 3.3. For the third, we will discuss the traceback performance below.

Assumptions

We identify two assumptions that motivate our design: (1) each deployed router supports multiple symmetric-key algorithms, and adopts the same symmetric-key algorithm; (2) each deployed router has the same shared secret, which can be periodically safely distributed and updated.

The mechanism of incremental deployment

Deployed routers form a confederation. Each deployed router encrypts the confederation name with the shared secret by the symmetric-key algorithm in advance, and gets the pre-generated ciphertext ID value, which is locally stored. And each deployed router has the same pre-generated ciphertext ID value.

When an attacking packet reaches a deployed router, two different cases will be described: (1) If the ciphertext ID value carried by the packet does not exist or equal to the pre-generated one, the router knows it itself is the first-hop deployed router for the packet, and writes a uniform initial TTL value and the pre-generated ciphertext ID value into the TTL field and the IP option field, respectively, and then marks it with the probability 100%. (2) If exists and equals, the router marks the packet with the probability inversely proportional to the hop number from the first deployed router to this router along the attacking path.

Traceback performance in the incremental deployment

We will analyze traceback performance in the typical scenario (shown in Fig. 6).

(1) Reaching probability. Let Q represent the probability that the attacking packet q is not marked by all downstream deployed routers of R_w. Assume that t_u is the uniform TTL

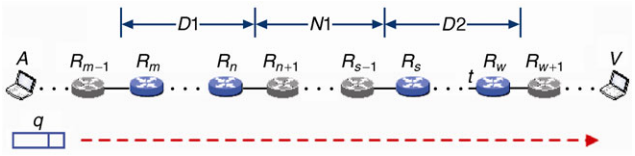

Fig. 6 A typical scenario of incremental deployment. Thereinto, blue and gray routers are deployed and non-deployed, respectively; $D1$ and $D2$ are two randomly chosen adjacent deployed segments, between which there is only one non-deployed segment, $N1$; R_i ($i \geqslant 1$) is the ith router from the attacking source to the victim; $m \leqslant n$, $n+1<s$, $s \leqslant w$; t is the TTL value of the attacking packet q when it arrives at the router R_w ($w \geqslant 1$)

value in APM. The reaching probability for each deployed router in $D2$ can be calculated as:

$$r_w = p_w \times Q = \frac{Q}{t_u - t + 1} \tag{24}$$

$$\begin{aligned} r_{w-1} &= p_{w-1} \times (1 - p_w) \times Q \\ &= \frac{1}{t_u - (t+1) + 1} \times (1 - \frac{1}{t_u - t + 1}) \times Q \\ &= \frac{Q}{t_u - t + 1} \\ &\cdots \end{aligned} \tag{25}$$

Therefore, we have

$$r_w = r_{w-1} = \cdots = r_s = \frac{Q}{t_u - t + 1} \tag{26}$$

Similarly, the reaching probability for each deployed router in $D1$ is derived as

$$\begin{aligned} r_n &= r_{n-1} = \cdots = r_m \\ &= \frac{(t_u - t - w) + s}{(t_u - t - w) + n + 1} \times r_w \end{aligned} \tag{27}$$

From Eq. (26) and Eq. (27), we can conclude that the reaching probabilities for deployed routers in one deployed segment are the same, and larger than those in the downstream deployed segments, since $s > n + 1$. That is to say, for an attacking path, the closer to the attacking source (the stepping stone or the reflector) the deployed segment is, more quickly it is reconstructed at the victim.

(2) Unmarked probability. The unmarked probability is zero, since the first-hop deployed router marks all traversing packets. Thus, we can conclude that marks spoofed by attacking sources cannot reach the victim.

7 Conclusions and future work

Adaptive Probabilistic Marking (APM) minimizes the number of packets required for the path reconstruction, and eliminates the effect of spoofed marks on the victim. NS2 experimental results show that, in APM, the time for a victim to collect the obligatory marks is reduced by more than 20% compared with PPM, TPM and DPPM, and spoofed marks cannot reach the victim and confuse its reconstruction process. Our future work includes: (1) to optimize the APM algorithm; (2) to further study the issues on incremental deployment of APM; (3) to apply APM to hybrid IP traceback [23 – 26] and make experiments in the real network environment constructed according to the topology of CNGI-CERNET2 [34].

Acknowledgements

We would like to thank Associate Prof. Jianwei Zhuge for the helpful discussion and suggestion on botnets. This work was supported by National Science Foundation of China under Grant 61073172, Program for New Century Excellent Talents in University, Specialized Research Fund for the Doctoral Program of Higher Education of China under Grant 20090002110026, and National Basic Research Program ("973" Program) of China under Grant 2009CB320501.

References

[1] W. Eddy and Verizon, "TCP SYN Flooding Attacks and Common Mitigations," RFC 4987, Aug. 2007.

[2] R. Vaughn and G. Evron (2006, Mar). DNS amplification attacks. Available: http://www.isotf.org/news/DNS-Amplification-Attacks.pdf

[3] CERT (1998). CERT Advisory CA-1998-01 Smurf IP Denial-of-Service Attacks. Available: http://www.cert.org/advisories/CA-1998-01.html.

[4] S. Savage, D. Wetherall, A. Karlin, and T. Anderson, "Network support for IP traceback," *IEEE/ACM Trans. Netw.*, vol. 9, no. 3, pp. 226 – 237, Jun. 2001.

[5] V. Paruchuri, A. Durresi, and S. Chellappan, "TTL based packet marking for IP traceback," in *Proc. IEEE GLOBECOM 2008*, New Orleans, USA, pp. 2552 – 2556.

[6] J. Liu, Z.-J. Lee, and Y.-C. Chung, "Dynamic probabilistic packet marking for efficient IP traceback," *Comput. Netw.*, vol. 51, no. 3, pp. 866 – 832, Feb. 2007.

[7] S. Savage, D.Wetherall, A. R. Karlin, and T. Anderson, "Practical network support for IP traceback," in *Proc. ACM SIGCOMM*, Stockholm, Sweden, 2000, pp. 259 – 306.

[8] A. Belenky and N. Ansari, "Accommodating fragmentation in deterministic packet marking for IP traceback," in *Proc. IEEE GLOBECOM 2003*, San Francisco, USA, pp. 1374 – 1378.

[9] D. X. Song and A. Perrig, "Advanced and authenticated marking schemes for IP traceback," in *Proc. INFOCOM 2001*, Anchorage, USA, vol. 2, pp. 878 – 886.

[10] M. Ma., "Tabu marking scheme to speedup IP traceback," *Comput. Netw.*, vol. 50, no. 18, pp. 3536 – 3549, Dec. 2006.

[11] A. Castelucio, A. Ziviani, and R. Salles, "An AS-level overlay network for IP traceback," *IEEE Network*, vol. 23, no. 1, pp. 36 – 41, Jan. – Feb. 2009.

[12] M. Adler, "Trade-offs in probabilistic packet marking for IP traceback," *J. the ACM*, vol. 52, no. 2, pp. 217 – 244, Mar. 2005.

[13] A. Yaar, A. Perrig, and D. Song, "FIT: fast Internet traceback," in *Proc. INFOCOM 2005*, Miami, USA, vol. 2, pp. 1395 – 1406.

[14] M.T. Goodrich, "Probabilistic packet marking for large-scale IP traceback," *IEEE/ACM Trans. Netw.*, vol. 16, no. 1, pp. 15 – 24, Feb. 2008.

[15] Y. Xiang, W. Zhou, and M. Guo, "Flexible deterministic packet marking: An IP traceback system to find the real source of attacks," *IEEE Trans. Trans. Parallel Distrib. Syst.*, vol. 20, no. 4, pp. 567 – 580, Apr. 2009.

[16] H. Burch and B. Cheswick, "Tracing anonymous packets to their approximate source," in *Proc. 14th USENIX Conf. Systems Administration*, Berkeley, USA, 2000, pp. 319 – 328.

[17] S. M. Bellovin, "ICMP traceback messages," Internet Draft, draft-ietf-itrace-04.txt, 2003.

[18] H. C. J. Lee, V. L. L. Thing, Y. Xu, and M. Ma, "ICMP traceback with cumulative path, an efficient solution for IP traceback," in *Lecture Notes in Computer Science*, vol. 2836, S. Qing, D. Gollmann, and J. Zhou, Eds. Berlin: Springer, 2003, pp. 124 – 135.

[19] A. Mankin, D. Massey, C.-L. Wu, S. F. Wu, and L. Zhang, "On Design and Evaluation of 'Intention-Driven' ICMP Traceback," in *Proc. 10th Int. Conf. Computer Communications and Networks*, Scottsdale, USA, 2001, pp. 159 – 165.

[20] A. C. Snoeren, C. Partiridge, L. A. Sanchez, C. E. Jones, F. Tchhakountio, S. T. Kent, and W. T. Strayer, "Hash-based IP traceBack," in *Proc. ACM SIGCOMM*, San Diego, USA, 2001, pp. 3 – 14.

[21] A. C. Snoeren, C. Alex, C. Partridge, L. A. Sanchez, C. E. Jones, F. Tchakountio, B. Schwartz, S. T. Kent, and W. T. Strayer, "Single-packet IP traceback," *IEEE/ACM Trans. Netw.*, vol. 10, no. 6, pp. 721 – 734, Dec. 2008.

[22] T. Korkmaz, G. Chao, S. Kamil, and S. G. Dykes, "Single packet IP traceback in AS-level partial deployment scenario," *Int. J. Security Netw.*, vol. 2, nos. 1 – 2, pp. 95 – 108, 2007.

[23] B. Duwairi and G. Manimaran, "Novel hybrid schemes employing packet marking and logging for IP traceback," *IEEE Trans. Parallel Distrib. Syst.*, vol. 17, no. 5, pp. 403 – 418, May 2006.

[24] C. Gong and K. Sarac, "IP traceback based on packet marking and logging source," in *Proc. IEEE Int. Conf. Communications (ICC)*, Seoul, Korea, 2005, pp. 1043 – 1047.

[25] C. Gong and K. Sarac, "A more practical approach for single-packet IP traceback using packet logging and marking," *IEEE Trans. Parallel Distrib. Syst.*, vol. 19, no. 10, pp. 1310 – 1324, Oct. 2008.

[26] M. Sung, J. Xu, J. Li, and L. Li, "Large-scale IP traceback in high-speed Internet: practical techniques and information-theoretic foundation," *IEEE/ACM Trans. Netw.* vol. 16, no. 6, pp. 1253 – 1266, Dec. 2008.

[27] R. Stone, "Centertrack: an IP overlay network for tracking DoS floods," in *Proc. 9th Conf. USENIX Security Symp.*, Berkeley, USA, 2000.

[28] P. Neal, "The generalised coupon collector problem," *J. Appl. Prob.*, vol. 45, no. 3, pp. 621 – 629, Sept. 2008.

[29] K. Park and H. Lee, "On the effectiveness of probabilistic packet marking for IP traceback under denial of service attack," in *Proc. IEEE INFOCOM 2001*, Anchorage, USA, vol. 1, pp. 338 – 347.

[30] J. Postel, "Internet Control Message Protocol", RFC 792, Sept. 1981.

[31] M. Cotton, L. Vegoda, and D. Meyer, "IANA Guidelines for IPv4 Multicast Address Assignments ", RFC 5771, Mar. 2010.

[32] University of Oregon Route Views Project, Available: http://www.routeviews.org/. Accessed Jun. 2010.

[33] BRITE Topology Generator, BRITE version 2.1b, Available: http://www.cs.bu.edu/fac/matta/Research/BRITE.

[34] CNGI-CERNET2, China Next Generation Internet project — China Education and Research Network (CERNET), Available: http://www.cernet2.edu.cn/index_en.htm.