2010 Second International Conference on Evolving Internet

A Probabilistic Marking Scheme for Fast Traceback

Hongcheng Tian, Jun Bi, Xiaoke Jiang and Wei Zhang
Network Research Center
Tsinghua University
Beijing, China
tianhc08@mails.tsinghua.edu.cn, junbi@tsinghua.edu.cn, justok06@gmail.com, zhangwei734@gmail.com

***Abstract*—For existing probabilistic marking technologies for IP traceback, such as Probabilistic Packet Marking (PPM), TTL-based Packet Marking (TPM) and Dynamic Probabilistic Packet Marking (DPPM), it is difficult to reconstruct attack path(s) fast and defend against spoofed marks. In this paper, we present Adaptive Probabilistic Marking scheme (APM), where the TTL value of each packet is set to a uniform number at the first-hop router, and each router deduces the distance that each packet has already traveled, and then adaptively marks the packet with the probability inversely proportional to the distance. We theoretically prove that, in APM, the victim requires the fewest packets for a successful traceback, the effect of spoofed marks can be eliminated. NS2 experiments show, in APM, the time for the victim to collect all the obligatory marks for the path reconstruction is reduced by more than 20% compared with existing schemes, and spoofed marks cannot reach the victim.**

Keywords-traceback; adaptive; probability; marking

I. 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 is routed dependent entirely on its destination address. Thus, attackers may spoof source addresses to attack remote hosts or nets, but it is difficult for victims to block the attack in real time, precisely locate attacker(s) and pursue legal actions.

To identify the machines that directly generate attack traffic and the network path this traffic subsequently follows is called traceback problem [1]. 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. [1] have originally implemented Probabilistic Packet Marking (PPM), where each router marks each packet with a fixed probability. In PPM, the victim requires many packets for the path reconstruction, slowing down IP traceback. In addition, it is difficult for PPM to defend against spoofed marks, resulting in uncertainty of the path reconstruction. Paruchuri et al. [2] have proposed TTL-based Packet Marking (TPM) and Liu et al. [3] have proposed Dynamic Probabilistic Packet Marking (DPPM), attempting to solve the problems introduced by 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), where the TTL value of each packet is modified uniformly at the first-hop router, and each router deduces the traveling distance (in hops) of each arriving packet from its source, and then adaptively marks it with the probability inversely proportional to its traveling distance. APM can be incorporated in other probabilistic marking techniques, such as PPM [1] and the randomize-and-link approach [11]. APM has the following two advantages: Firstly, the victim requires the fewest packets for the path reconstruction, speeding up IP traceback. Thus, subsequent actions – such as packet filtering and traffic constraint along the attack path(s) – can be taken in time against attack(s). Secondly, APM can eliminate the effect of spoofed marks on the victim. NS2 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.

The rest of the paper is organized as follows: In Section II, we review traceback literature. Section III presents APM and its implementation issues. Section IV and Section V compare the performance of PPM, TPM, DPPM and APM from theory and experiment, respectively. And in Section VI, concluding remarks and our future work are given.

II. RELATED WORK

Researchers have proposed various approaches to trace attacking packets back to attacker(s).

A. A General Background

Input debugging and controlled flooding [4] belong to real-time approaches for IP traceback. 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 hop by hop. Controlled flooding [4] floods each link of a router with large bursts of traffic and observes changes in the rate of invading packets. When the rate of invading packets is reduced, the link the attacking packets come from can be deduced.

978-0-7695-4185-3/10 $26.00 © 2010 IEEE
DOI 10.1109/INTERNET.2010.32

The following approaches can traceback not only in real time but also post mortem. ICMP traceback message [5] 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 attack path(s). A weakness of this approach is that the ICMP traceback message may be filtered in the Internet and cannot arrive at the victim. The packet marking schemes [1][10][11] are for routers along attack path(s) to mark packets with partial path information, and for the victim to extract path information to reconstruct attack path(s). The shortcomings of packet marking schemes are high false positives and computing overhead at the victim. Packet logging schemes [6][13] are to log packet digests at intermediate routers, and can trace the origin of a single IP packet by recursive queries. Packet logging schemes suffer from the high computing and storage overhead at routers. Hybrid IP traceback [7][8][14] is designed to make use of advantages of packet marking and packet logging schemes and alleviate their weaknesses, but relevant management mechanism is complicated.

B. *Probabilistic Marking Schemes*

Probabilistic marking schemes have been much studied [1][2][3][9][10][11][12]. 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 having been marked by intermediate routers. And crafty attackers may send packets with spoofed marks to compromise the traceback. Paruchuri et al. [2] and Liu et al. [3] attempted to solve the problems introduced by the nature of probability.

TPM has been proposed by [2], 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 [3], attempting to precisely pinpoint the attacker(s) even under spoofed marking attacks. But DPPM doesn't work well enough when attackers cunningly set TTL values and spoof marks in packets, in which case the path reconstruction process may be significantly confused.

III. ADAPTIVE PROBABILISTIC MARKING SCHEME

In this section, we present APM, an adaptive marking scheme based on the TTL field, in order to minimize the number of packets required for the attack 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.

A. *Definitions and Lemma*

Fig. 1 describes an attacking scenario aimed at a victim V. V may be a host, a NAT or a firewall. Attacker(s) may be a host, or a group of hosts distributed in different sites. *Attack path* is defined as an ordered list of routers from an attacker to a victim. For example, in

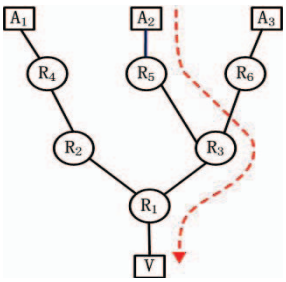

Figure 1. An attacking scenario aimed at a victim V

Figure 2. Attack path ξ

Fig. 1, from A_2 to V, the attack path is (R_5, R_3, R_1), as shown by a dashed line. When the victim faces DDoS attacks, different attack paths from different attackers form an attack tree rooted at the victim, in which each attacker is a leaf node. Assume that an attacking packet traverses d routers from A to V, the attack path ξ is (R_1, R_2, …, R_i, …, R_d) ($1 \le i \le d$), as described in Fig. 2. In the following, we will carry out our research on the attack path ξ.

Let p_i represent the marking probability of router R_i for an attacking packet. In the attack path ξ, 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 ξ 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

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

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 ξ, which is expressed as

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

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

Proof: Note that the attack path ξ 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

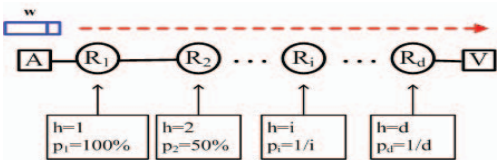

Figure 3. When an attacking packet w travels from the attacker A to the victim V, its TTL value is set to a uniform number (such as 128) at the first-hop router R_1, and each router deduces its traveling distance (in hops) from its source, and then adaptively marks it with the probability inversely proportional to its traveling distance. Let h and p_i represent the traveling distance of w and the marking probability of R_i for w, respectively.

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 [15]. 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.

B. Design Goals

If a number of packets reach the victim unmarked, the victim will take additional time to receive more packets so as to obtain enough marking information for the path reconstruction, and countermeasures (such as packet filtering and traffic constraint along the attack path ξ) cannot be taken in time to alleviate the effect on the victim.

On the other hand, Park and Lee [16] have showed 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 attacker cannot be identified precisely. Furthermore, the uncertainty is significantly amplified under DDoS attacks. In TPM and DPPM, when attackers 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 attack path against the attack.
- Security. Spoofed marks as few as possible can reach the victim, lessening the effect on the path reconstruction.

C. 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 128) at the first-hop router, and each router deduces the traveling distance (in hops) of each arriving packet from its source, and then adaptively marks it with the probability inversely proportional to the distance. For the given attack path ξ, let h ($1 \le h \le d$) be the traveling distance of an attacking packet from the attacker A to router R_i ($1 \le i \le d$). Obviously, h=i, router R_i adaptively chooses the probability

$$p_i = \frac{1}{h} = \frac{1}{i} \qquad \text{for} \quad 1 \le i \le d. \qquad (3)$$

to mark the packet. The marking procedure is described in Fig. 3.

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

$$r_i = \frac{1}{d} \qquad \text{for} \quad 1 \le i \le d. \qquad (4)$$

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

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

$$U_{APM} = \prod_{k=1}^{d}(1 - p_k) = (1-1)\ldots(1-\frac{1}{d}) = 0. \qquad (5)$$

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

Thus, APM meets the two design goals mentioned in Section III.B. 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 hops) of an arriving packet from its source? We will answer this in the following subsection.

1) 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, firstly, it is meaningless that different operating systems and protocols use different Initial TTL values for the same Internet; secondly, 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; thirdly, a malicious attacker 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 router can identify the traveling distance of the packet from its source by calculating the difference between 64 (128 or 255) and the TTL value of the packet, and then mark the packet with the probability inversely proportional to the traveling distance.

How does 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 so as 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 attack path ξ can deduce the traveling distance of each packet from its source, and then adaptively choose marking probability.

2) APM marking algorithm:

The APM marking algorithm is shown in Fig. 4, 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 =128. 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 attack path ξ, the sequence of the marking probabilities for intermediate routers is 1, 1/2 …… 1/d.

```
for each packet
  if it is at a first-hop router
    t ← t_u;
  h ← t_u - t+1;
  t ← t - 1;
  let r be a random number in [0,1);
  if r <= 1/h
    mark the packet
```

Figure 4. APM marking algorithm

IV. Performance Comparison

PPM [1], TPM [2], DPPM [3] and APM are compared in terms of the reaching probability and the unmarked probability in Table I.

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 [1]. When TPM is deployed in the Internet, the reaching probability for each router is unequal according to [2]. In DPPM, when the spoofed initial TTL values of packets are smaller than 32, the reaching probability for each router is equal according to [3], but is not 1/d. For APM, no matter how an attacker 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 [1]. 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 [2]. 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 [3]. In APM, no matter how an attacker 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 [17], the optimal marking probability is 1/25 for PPM [1], and we choose 17 as the spoofed initial TTL value. Thus, U_{PPM}, U_{TPM} and U_{DPPM} are around 52%, 30% and 48%, respectively. But according to Eq. (5), U_{APM} is 0.

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

V. Simulation Experiments

We carried out the simulation experiments using NS2 and BRITE. 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 probability. Especially, for PPM, we chose 2 marking probabilities of 0.04 and 0.1 to make the experiments, respectively. We considered, firstly, attackers could spoof initial TTL values of packets randomly; secondly, 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, 100 experiments were made in each network

TABLE I. Performance Comparison

	PPM	TPM	DPPM	APM
Reaching probability	unequal	unequal, when deployed in the Internet	equal, but is not 1/d, when an attacker craftily spoofs initial TTL values of packets	1/d
Unmarked probability	> 0 (e.g., 52%)	greater than 0 (e.g., 30%), when an attacker craftily spoofs initial TTL values of packets	greater than 0 (e.g., 48%), when an attacker craftily spoofs initial TTL values of packets	0

topology to average the collecting time and the unmarked probability.

Fig. 5 describes the collecting time for each approach in each network topology. Especially, for TPM, it is very difficult for the victim to collect all the obligatory marks in each network topology. The collecting time for TPM is longer than other approaches. Thus, the collecting time for TPM is not listed in Fig. 5. The experimental results show that, the collecting time for APM is the shortest, which is reduced by more than 20% compared with other approaches.

Fig. 6 describes the unmarked probability for each approach in each network topology. The results show that, for PPM-0.04 and PPM-0.1, the unmarked probability is great, for TPM and DPPM small, and for APM zero. So APM is the best for the ability to prevent against spoofed marks.

VI. Conclusions and Future Work

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 other approaches, and spoofed marks cannot reach the victim. APM can be incorporated in other packet marking techniques, such as PPM [1] and the randomize-and-link approach [11]. Our future work is to explore the quantitative relation among factors which affect the speed of the attack path reconstruction (such as the marking algorithm, the reconstruction algorithm and spoofed marks) and to apply APM to hybrid IP traceback [7][8].

Acknowledgment

This work was supported by the Specialized Research Fund for the Doctoral Program of Higher Education of China under Grant No.20090002110026; the National Science and Technology Support Program of China under Grant No. 2008BAH37B02.

References

[1] S. Savage, D. Wetherall, A. Karlin, and T. Anderson, "Network support for IP traceback," IEEE/ACM Transactions on Networking 20 (2), 226–237(2001).

[2] V. Paruchuri, A. Durresi, and S. Chellappan, "TTL based Packet Marking for IP Traceback," Proceedings of IEEE GLOBECOM (2008).

[3] J. Liu, Z.J. Lee, and Y.C. Chung, "Dynamic probabilistic packet marking for efficient IP traceback," Computer Networks 51(3), 866-82 (2007).

[4] H. Burch and B. Cheswick, "Tracing anonymous packets to their approximate source," Proceedings of 14th Systems Administration Conference (2000).

[5] S.M. Bellovin, "ICMP Traceback Messages," Internet Draft, draft-ietf-itrace-04.txt (2003).

[6] A.C. Snoeren, C. Partiridge, L.A. Sanchez, C.E. Jones, F. Tchhakountio, S.T. Kent, and W.T. Strayer, "Hash-Based IP TraceBack," Proceedings of ACM SIGCOMM (2001)

[7] B. Duwairi and G. Manimaran, "Novel Hybrid Schemes Employing Packet Marking and Logging for IP Traceback," IEEE Transaction on Parallel and Distributed Systems 17(5), 403-418 (2006).

[8] C. Gong and K. Sarac, "A more practical approach for single-packet IP traceback using packet logging and marking," IEEE Transaction on Parallel and Distributed Systems 19(10), 1310-1324(2008).

[9] M. Adler, "Trade-offs in probabilistic packet marking for IP traceback," Journal of the ACM 52(2), 217-244(2005).

[10] A. Yaar, A. Perrig, and D. Song, "FIT: Fast Internet Traceback," Proceedings of IEEE INFOCOM (2005).

[11] M.T. Goodrich, "Probabilistic packet marking for large-scale IP traceback," IEEE/ACM Transactions on Networking 16(1), 15-24(2008).

[12] Y. Xiang, WL. Zhou, and MY. Guo, "Flexible Deterministic Packet Marking: An IP Traceback System to Find the Real Source of Attacks," IEEE Transaction on Parallel and Distributed Systems 20(4), 567-580(2009).

[13] T. Korkmaz, G. Chao, S. Kamil, and S.G. Dykes, "Single packet IP traceback in AS-level partial deployment scenario," International Journal of Security and Networks 2(1-2), 95-108(2007).

[14] MH. Sung, J. Xu, J. Li, and L. Li, "Large-Scale IP Traceback in High-Speed Internet: Practical Techniques and Information-Theoretic Foundation," IEEE/ACM Transactions on Networking 16(6), 1253-66(2008).

[15] P. Neal, "The Generalised Coupon Collector Problem," Journal of Applied Probability 45(3), 621-629 (2008).

[16] K. Park and H. Lee, "On the effectiveness of probabilistic packet marking for IP traceback under denial of service attack," Proceedings of IEEE INFOCOM (2001).

[17] University of Oregon Route Views Project, http://www.routeviews.org/. last accessed in June, 2010.

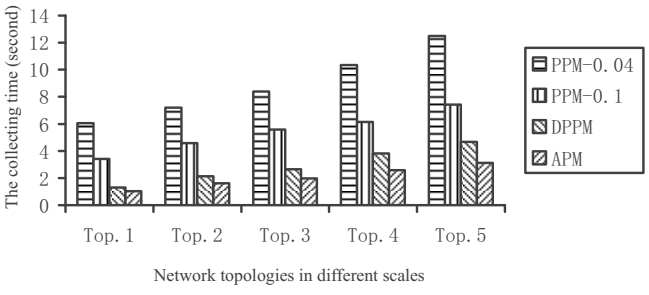

Figure 5. The collecting time

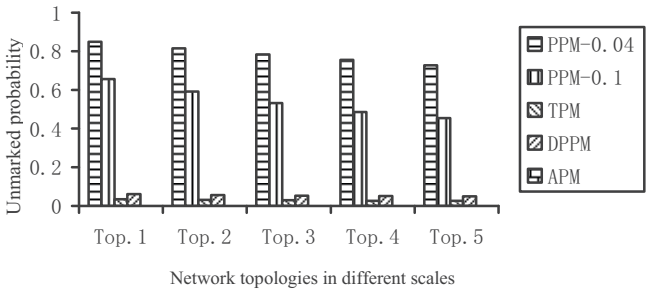

Figure 6. Unmarked probability