Probabilistic Packet Marking With Non-Preemptive Compensation

Yu Kuo Tseng, Hsi Han Chen, and Wen Shyong Hsieh, *Member, IEEE*

***Abstract*—A new scheme in probabilistic packet marking (PPM) for IP traceback against denial-of-service attack is presented. Non-preemptive PPM is performed while a marked packet is coming, but compensates the reduction of marking probability in marked-free packets. The nonpreemptive compensation makes the probability of each marked packet arrived at the victim is equal to its original marking probability. This scheme efficiently improves the convergent amount of marked packets required for reconstructing the complete attack path.**

***Index Terms*—Computer network security, denial-of-service (DoS), IP traceback, probabilitic packet marking (PPM).**

I. Introduction

IP TRACEBACK is a technique for identifying the source of the anonymous flood-type denial-of-service (DoS) attack, which consumes the resource of the victim with hundreds of thousands of spoofed packets to obstruct services to legal users, to make the attacker accountable. Numerous approaches have been proposed, such as ingress filtering, packets logging, link testing, and additional Internet Control Message Protocol (ICMP) messages, etc., [1]. Generally, probabilitic packet marking (PPM) [1]–[3] is still better than others, because it is simple to incrementally implement, does not need any additional bandwidth or storage, and can be performed "post mortem." In this letter, we first review S. Savage's PPM [1], and then introduce and analyze our scheme on improving the PPM's convergence.

II. Probabilistic Packet Marking (PPM)

Each router with PPM marks packets probabilistically, and the victim reconstructs the attack path by collecting enough marked packets after a DoS attack is detected. Because 50% of DoS attacks have at least 1000 packets per second, and most attacks last at least 10 min [4], then the victim may obtain enough marked packets. The algorithm is shown as Fig. 1 [1].

In PPM, there are four important criteria: 1) the convergent amount of marked attack packets; 2) the computing overhead for reconstructing the attack path; 3) the robust against the false positive/negative,; and 4) the deployment cost. In this letter, our main objective is to improve the convergent amount.

Marking procedure at router R with the marking probability p:

```
for each packet w with a tuple (w.start, w.end, w.distance) in the IP Header
/* Both w.start and w.end are assigned the address of routers at each end
   of one link respectively.
   w.distance is the hops from a PPM router to the victim. */
{
 get a random number x, and x is in [0..1]
 if x < p then
     write R into w.start and 0 into w.distance
 else
  {
   if w.distance = 0 then
      write R into w.end
   increment w.distance
  }
}
```

Path reconstruction procedure at victim v:

```
let G be a tree with root v
let edges in G be tuples (start, end, distance)

for each packet w collected by v
{
 if w.distance = 0 then
  insert edge (w.start, v, 0) into G
 else
  insert edge (w.start, w.end, w.distance) into G
}
remove any edge (x, y, d) with d ≠ distance from x to v in G
extract path (Ri..Rj) by enumerating acyclic paths in G
```

Fig. 1. PPM algorithm.

If each router has a disjoint equal marking probability p, the probability of receiving a marked packet from a router, d hops away from the victim, will be reduced to $p(1-p)^{d-1}$. PPM conservatively assumes that marked packets from all of d routers have the same likelihood as the furthest router, and the number of packets required for reconstructing a path of d routers has the following bounded expectation [1]:

$$E(X) \approx d(\ln(d)+\gamma)\cdot\frac{1}{dp(1-p)^{d-1}} = \frac{\ln(d)+\gamma}{p(1-p)^{d-1}} \quad (1)$$

where $d(\ln(d)+\gamma)$ is the expected trial number of obtaining at least each kind of d equiprobable marked packets from d routers respectively according to the coupon collector problem [5], and $\gamma \fallingdotseq 0.58$. By differentiating (1), the optimal p is $1/d$.

III. PPM With Non-Preemptive Compensation (PPM-NPC)

We propose a simple and efficient improvement in PPM, trying to make each router along the attack path can fully nonpreemptively compensate the reduced marking probability by utilizing lots of marked-free packets, so that the probability that a marked packet is received by the victim is equal to its

Manuscript received November 13, 2003. The associate editor coordinating the review of this letter and approving it for publication was Prof. C.-K. Wu.

Y. K. Tseng and H. H. Chen are with the Department of Computer Science and Engineering, National Sun Yat-Sen University, Kaohsiung 804, Taiwan, R.O.C. (e-mail: d9034812@student.nsysu.edu.tw; clare@dagger.com.tw).

W. S. Hsieh is with the Department of Computer Science and Information Engineering, Shu Te University, Kaohsiung 824, Taiwan, R.O.C. (e-mail: wshsieh@mail.stu.edu.tw).

Digital Object Identifier 10.1109/LCOMM.2004.831336

PPM-NPC marking procedure at router R with marking probability p:

```
for each incoming packet w with a tuple (w.start, w.end, w.distance) in the IP Header
    get a random number x. x is in [0..1]
    let C be a compensation counter in this router
    if x < p then
      {
       if w.start <> 0 then
          {
           C= C + 1
          }
       if w is a marked-free packet then
          w.start = R,    w.distance = 0
      }
     else
      if w is a marked-free packet and C is not zero then
       {
        w.start = R,    w.distance = 0
        C= C - 1
       } /* end of " if x<p" */

     if ( w.disatnce = 0 and w.start <> R ) then
        w.end = R
     increment w.distance
```

Fig. 2. PPM-NPC marking algorithm.

marking probability p. The PPM-NPC marking procedure is shown as Fig. 2.

PPM-NPC is similar with PPM, but when a router receives a marked packet and $x < p$, it will increase its compensation counter instead of re-marking this packet. While $x \geqq p$, the reduced probability of nonpreemptive marking will be compensated by marking a marked-free packet, and the compensation counter will be decreased.

IV. Analysis

In this section, we analyze the performance of PPM-NPC from two important properties and the compensation counter's efficiency.

A. Properties of PPM-NPC

Fig. 3 is the decision tree in R_1, R_2, and R_3. A router labeled as R_i indicates that the distance between the attack source and this router is i hops, and $R_{\max(i)}$ is the attached router of the victim. We define two parameters: (i) $p_{j,i}$ is the probability that a packet is marked by R_j, and this marked packet will arrive at the successive router R_i, where $i > j$. (ii) E_i is the probability that outgoing packets from R_i are marked-free.There are two properties as follows:

1) $p_{j,i} \leqq p$. As the nonpreemptive nature of PPM-NPC, every packet marked by R_j will arrive at the successive routers without being altered, so that $p_{j,i} \leqq p$ while $i > j$. The reduction of marking probability caused by the nonpreemptive scheme may be compensated in lots of marked-free packets. If the volume of marked-free packets is greater than the reduction, $p_{j,i}$ may be equal to p, which is the marking probability in any router.
2) $E_i \geqq (1 - i \cdot p)$, for all $i \geqq 1$.

Proof: The general form of E_i is as follows:

$$E_i = E_{i-1} \cdot (1-p) - \left(\sum_{j=1}^{i-1} p_{j,i} \right) \cdot p. \tag{2}$$

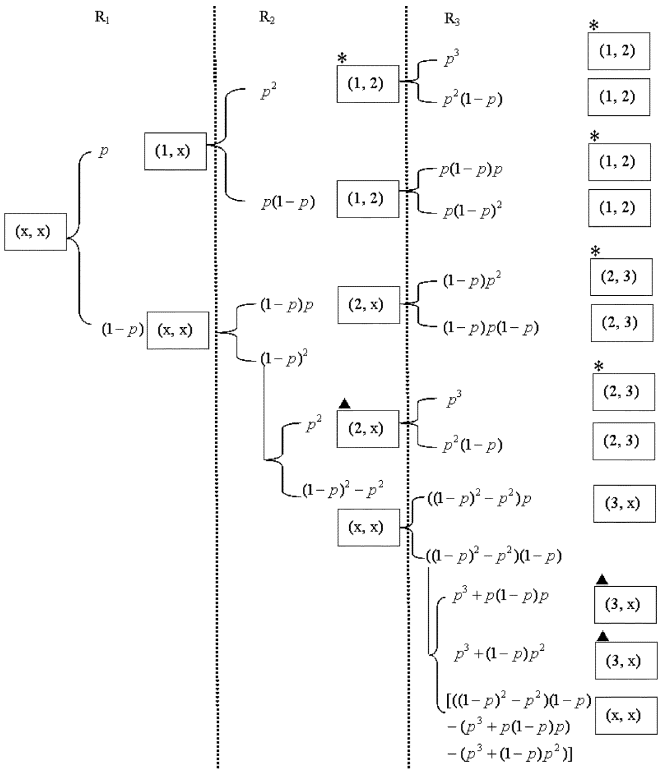

Fig. 3. PPM-NPC decision tree from router R_1 to router R_3 (x, x)—marked-free situation $*$—nonpreemptive situation ▲—compensated situation.

The first term in (2) is the probability of marked-free packets, and the second term is the reduction of marking probability in R_i. We know that $p_{j,i} \leqq p$, so that

$$\begin{aligned} E_i &= \left[E_{i-1} \cdot (1-p) - \left(\sum_{j=1}^{i-1} p_{j,i} \right) \cdot p \right] \\ &\geq E_{i-1} \cdot (1-p) - (i-1) \cdot p^2. \end{aligned} \tag{3}$$

Now, we prove $E_i \geqq (1 - i \cdot p)$. When $i = 1$, $E_1 = 1 - p$. When $i = 2$, $E_2 \geqq (1-p)(1-p) - p^2 = 1 - 2p$. Assume that $E_n \geqq (1 - n \cdot p)$ is true while $i = n$

$$\begin{aligned} E_{n+1} &= \left[E_n \cdot (1-p) - \left(\sum_{j=1}^{n} p_{j,n+1} \right) \cdot p \right] \\ &\geq E_n \cdot (1-p) - n \cdot p^2 \\ &\geq (1-np)(1-p) - np^2 = 1 - (n+1)p \end{aligned} \tag{4}$$

so that $E_i \geqq (1 - i \cdot p)$ is true for all $i \geqq 1$

Property 2 shows that E_i is a monotonously decreasing series, and $E_i > E_j$ while $j > i$. If the distance from the attacker to the victim is $\max(i)$ hops, and we set p as $1/\max(i)$, then we can get $E_i \geqq 0$ for all $i \geqq 1$. This result shows that the volume of marked-free packets is always equal to or greater than the reduction of marking probability, so that we can get $p_{j,i} = p$ in all j and $i \geqq j$.

In PPM, the field of distance is 5 bits, this means that the longest path from the attacker to the victim is 31 hops, the value is reasonable for a practical network [1]. Therefore, if we set p as 1/31, then $E_j \geqq 0$, and $p_{j,i} = p$ is true for all j and $i > j$.

$p_{j,victim} = p$ implies the marked packets marked by any router will arrive at the victim without being altered.

B. The Consideration of Compensation Counter

In programming language C, the size of normal integer variable is 2 bytes. For a 2-byte unsigned integer variable, the counter could have 655 356 unsigned integers. Furthermore, the probability required for being compensated in R_n is C_n

$$C_n = \left(\sum_{j=1}^{n-1} p_{j,n} \right) \cdot p = (n-1) \cdot p^2. \tag{5}$$

As the discussion in PPM-NPC properties, we will set p as $(1/(2^5 - 1)) \leqq p \leqq (1/\max(i))$. We can conservatively assume that the probability required for being compensated, in worst case, is $(2^5 - 2) \cdot p^2$ for each router. If the total number of packets is M, each router's counter must handle $[M \cdot (2^5 - 2) \cdot p^2]$ at most. While the number of packets required for convergence in PPM-NPC is $[(d \cdot (\ln(d) + \gamma))/(d \cdot p)]$, the necessary counter size of a linear topology is:

$$\left[\frac{\ln(d) + \gamma}{p} \right] \cdot \left(2^5 - 2\right) \cdot p^2 = \left(2^5 - 2\right) \cdot p \cdot (\ln(d) + \gamma). \tag{6}$$

While d is 10 and p is 1/25 as in [4], the average counter size of each router is approximately 4. Therefore, for a 2-byte integer variable, we can cope with (65536/4) different attack paths, especially in the distributed DoS.

V. Simulation Result

We use the real-time NS-2 simulator [6] to verify our scheme. The experimental linear topology consists of 1 attacker, 1 victim, and 10 intermediate routers. The attack traffic rate is 200 User Datagram Protocol (UDP) packets per second. The marking probability is 0.1. After 15 simulation runs, the mean value is roughly 30, and the standard deviation is 7.8. Because the idea result is $(\ln(10) + 0.58)/0.1 \fallingdotseq 29$, our simulation result also proves our PPM-NPC scheme can approach the optimal situation.

VI. Conclusion

With nonpreemptively compensation, the probability of each marked packet arrived at the victim is equal to its original marking probability. Therefore, the PPM-NPC will efficiently achieve the optimal convergent situation by simply utilizing a 2-byte integer counter.

References

[1] G. S. Savage, D. Wetherall, A. Karlin, and T. Anderson, "Network support for IP traceback," *IEEE/ACM Trans. Networking*, vol. 9, pp. 226–237, June 2001.

[2] T. Peng, C. Leckie, and K. Ramamohanarao, "Adjusted probabilistic packet marking for IP traceback," in *Proc. Networking*, May 2002, pp. 697–708.

[3] M. Adler, "Tradeoffs in probabilistic packet marking for IP traceback," *Proc. 34th ACM Symp. Theory of Computing (STOC)*, pp. 407–418, May 2002.

[4] D. Moore, G. M. Voelker, and S. Savage, "Inferring internet denial-of-service activity," in *Proc. 10th Usenix Security Symp.*, 2001, pp. 9–22.

[5] W. Feller, *An Introduction to Probability Theory and Its Applications*, 2nd ed. New York: Wiley, 1996, vol. 1.

[6] LBNL Network Research Group. UCB/LBNL/VINT Network Simulator—ns (version 2). DARPA: VINT project. [Online]. Available: http://www.isi.edu/nsnam/ns