I.J. Wireless and Microwave Technologies, **2011, 5, 23-30**
Published Online October 2011 in MECS (http://www.mecs-press.net)
DOI: 10.5815/ijwmt.2011.05.04

Available online at http://www.mecs-press.net/ijwmt

An Improved Dynamic Probabilistic Packet Marking Algorithm

Qiao Yan[a], Xiaoming He[b],Tuwen Ning[b]

[a]*Department of Computer and Software, Shenzhen University, Shenzhen, China*
[b]*Department of Information Engineering, Shenzhen University, Shenzhen, China*

Abstract

An improved dynamic probabilistic packet marking algorithm named IDPPM is presented, which not only can reduce the marking overhead of routers near the attackers, but also can locate attack source rapidly and accurately. The challenge of weakest node and weakest link is solved with the price of a little more numbers of packets to reconstruct the attack path. Theoretical analysis and NS2 simulation demonstrate the effectiveness of the algorithm.

Index Terms: Distributed denial of service(DDoS);IP traceback; Dynamic Probabilistic Packet Marking(DPPM)

1. Introduction

Denial-of-service (DoS) attacks pose an increasing threat to the network systems in recent years as they are simple to implement, hard to prevent, and difficult to trace. In particular, Distributed-denial-of-service attacks (DDoS) become a major threat for the Internet because cohorts of malicious or compromised hosts coordinate to send a large volume of aggregate traffic to a victim. And the attackers often use spoofed IP address to disguise the true IP address because of the flaws of IP protocol. So IP source tracing on network is one of the most pressing tasks for network security researchers.

A variety of IP traceback techniques have been proposed and assessed. The idea of encoding the address of the routers into attacking packets was first presented by Burch and Cheswick [1]. Savage, et al [2] proposed the famous traceback scheme, probabilistic packet marking (PPM), for practical IP traceback. Song and Perrig have an improved packet marking scheme that copes with multiple attackers IP traceback problem [3]. The technique features low router overhead, supports incremental deployment, and provides efficient authentication of routers' markings. In 2009,the work of Feng bo, et al present a new packet marking algorithm to improve the effectiveness of PPM by using dynamic probability and fragment-reassembly[4], which significantly solves the problems of the lost of marking information and the difficulties to reconstruct the attack path. In 2006, a subtle

* Corresponding author.
E-mail address: yanq@szu.edu.cn, hfreeman2008@126.com

approach, called dynamic probabilistic packet marking (DPPM), was presented [5]. Instead of using a fixed marking probability, DPPM deduces the traveling distance of a packet and then choose a proper marking probability. And DPPM may completely remove uncertainty and enable victims to precisely traceback the attacking origin even under spoofed marking DoS attacks. And Gao dapeng, et al proposed a new approach of composed packet marking method [6]. Compare with the DPPM algorithm, the marking probability of border router decreases from 1 to 0.5 in this new proposal.

2. Dynamic Probabilistic Packet Marking

We assume that the attack path $\psi = a, r_1, r_2, ..., r_D, v$ is comprised of D routers, where a and v denote the attacker and the victim of a DoS occurrence, and r_i (i=1, 2,…, D) indicate D routers in the attack path. Let p_i represent the marking probability of router r_i. Define the leftover probability for router r_i, denoted by α_i, to be the probability that an attacking packet has lastly been marked at router r_i and nowhere further down the path [5].For victim v, α_i is the probability that allows v to learn that router r_i is on the attack path by examining this arriving packet [5].

It can be seen that:

$$\alpha_i = \begin{cases} p_i * \prod_{j=i+1}^{D} (1-p_j) & for \quad 1 \le i < D \\ p_D & for \quad i = D \end{cases} \tag{1}$$

According to dynamic probabilistic packet marking (DPPM), the marking probability of router r_i is chosen $p_i = 1/i$ to mark packet.

It can be shown that

$$\alpha_i = \frac{1}{D} \quad for \quad 1 \le i \le D \tag{2}$$

So each router along the attack path has the same leftover probability. By the theory of Coupon collector [7], we know that the victim needs the minimal number of packets to reconstruct the attack path successfully.

3. An Improved Dynamic Probabilistic Packet Marking Algorithm

A. The Basic Idea of The Improved Dynamic Probabilistic Packet Marking

When the value of i is smaller, which means that the router is closer to the attacker, the marking probability of the router is greater. Especially, as the i is 1,2,3, the marking probability of r_1, r_2, r_3 is up to 1,1/2,1/3 respectively, which will result in an excessive burden on the router, and even the service is paralyzed. The above-mentioned is the biggest drawback of DPPM [5]. And from a purely sampling point-of-view, edge (a, r_1) is the "weakest link" and node a is the "weakest node" requiring the most samples for path reconstruction because the packet's marking information will be overwritten[8].

In this paper, we present a technique which make an improvement on DPPM by using 2bits field (F0F1) in packet header to solve these problems, which is named IDPPM. The following is basic idea of IDPPM. We initialize the value of F0F1 to (00). A router checks the value of i .If the value of i is equal to (1,2,3),it denotes

that the distance between the router and the attacker is 1,2,3 respectively, the router will mark the packet with probability p1,p2,p3,and set the value of F0F1 to (01,10,11) separately. When a router marks packets, it must first check the value of F0F1.If F0F1= (00), the router marks the packet with the DPPM algorithm. If F0F1= (01, 10, 11), this means that this packet is marked before, the router does not mark this packet in order to avoid overwritten.

We choose the value of p1, p2, p3 to1/D. This can not only greatly reduce the marking probability of routers which are near the attackers, but also ensure that the leftover probability of the router (r_1 , r_2 , r_3) are the same as DPPM. Furthermore, IDPPM can resolve the "weakest link" and "weakest node" puzzle by using 2bits field to avoid overwritten during the attacker path reconstruction.

B. Marking Field Selection and Encoding Issues

According to [9], since less than 0.25% of all Internet data packets will use the "identification" (16-bit), we think that the path information is overloaded into this field is appropriate. The TOS field is an 8bits field in the IP header. And the field has been little used in the past. Reference [10] shows that setting this field arbitrarily makes no measurable difference in packet delivery. As shown in Fig. 1, we choose to use ID field (16-bit) and 2 bits out of the TOS field as marking field for IDPPM algorithm.

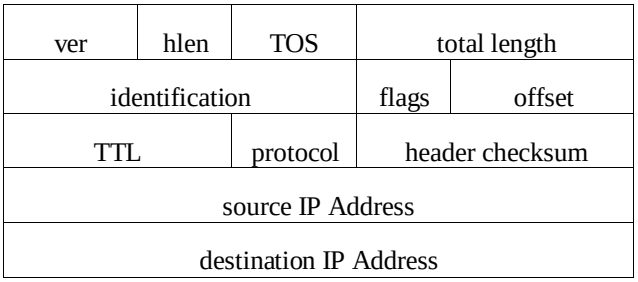

Figure 1. The IP header(darkened areas represent underutilized bits)

There are just only 18bits field available for use in each packet. So we use the Compressed Edge Fragment Sampling scheme to encode the edge fragments into the IP marking Field. The marking field encoding format is shown in Fig. 2:

offset	distance	edge fragment	F0	F1
3-bit	5-bit	8-bit	1-bit	1-bit

Figure 2. The marking field encoding format

F0F1: value set is (00, 01, 10, 11), mainly used to mark the router r_1 , r_2 , r_3 .

C. The Algorithm of IDPPM

Marking procedure at router R:

let R'=BitIntereave(R,Hash(R))

let k be the number of non-overlapping fragments R'

for each packet w

let x be a random number from[0,…,1)

if F0F1==(00) then

```
        let o be a random integer from [0,…,k-1]
        let f be the fragment of R' at offset o
        if 3<i and x<1/i then
           Mark_packet()
        else if i==1 and x<p1 then
           write 01 into w.F0F1
           Mark_packet()
        else if i==2 and x<p2 then
           write 10 into w.F0F1
           Mark_packet()
        else if i==3 and x<p3 then
           write 11 into w.F0F1
           Mark_packet()
    else
       if w.distance=0 then
       let f be the fragment of R' at offset w.offset
           write f ⊕ w.frag into w.frag
           increment w.distance

Mark_packet():
    write 0 into w.distance
    write o into w.offset
    write f into w.frag
```

4. Performance Analysis

D. Overhead on Routers

Each marking poses some cost to a router. We now proceed to compare the overhead of DPPM and IDPPM. For simplicity, we use number of markings performed as our measurement for overhead on router [5]. Let us consider a DoS attack with N packets sent from α to ν. And let o_{dppm} and o_{idppm} denote individual overhead of DPPM and IDPPM in a route along the attack path, respectively. Let O_{dppm} and O_{idppm} denote the total overhead summed over all D routers of DPPM and IDPPM, respectively.

$$o_{dppm} = N / i \tag{3}$$

$$o_{idppm} = \begin{Bmatrix} N * \frac{1}{D} & i \le 3 \\ N / i & i > 3 \end{Bmatrix} \tag{4}$$

$$O_{dppm} = N\left(\frac{1}{1} + \frac{1}{2} + \frac{1}{3} + \mathrm{L} + \frac{1}{D}\right) \tag{5}$$

$$O_{idppm} = \begin{Bmatrix} N * \frac{i}{D} & i \le 3 \\ N\left(\frac{3}{D} + \frac{1}{4} + \mathrm{L} + \frac{1}{i}\right) & i > 3 \end{Bmatrix} \tag{6}$$

Fig. 3 compares the total overhead of DPPM and IDPPM with different D. It is clear that O_{idppm} is less than O_{dppm} significantly on all routers. This means that the routers under IDPPM suffer a low total overhead. This is mainly due to the fact that IDPPM algorithm reduces the marking probability of routers (r_1, r_2, r_3) greatly by using the F0F1 field. So $o_{idppm} < o_{dppm}$, $O_{idppm} < O_{dppm}$.

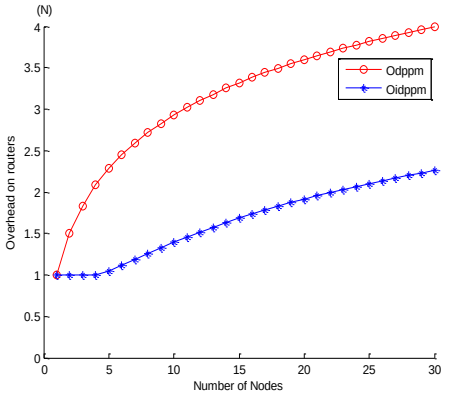

Figure 3. A comparison of total overhead by DPPM and IDPPM

E. False Positive

We would like to note that a path reconstruction mechanism will suffer from false positives. The main reason is that it is difficult to prove whether the path is reconstructed completely or partly [11].

As the IDPPM algorithm marks the edge router of r_1 with the value of F0F1 (01). If F0F1=01, it denotes that this packet is marked by the edge router (r_1), and the path is reconstructed completely. So this can reduce the false positives obviously.

F. Expected Value of Minimal Number of Packets for Reconstruction

To satisfy the requirement of at-least-one-marking per router, a victim needs to collect a certain number of packets [5]. The expected value of minimal number of packets required for a successful traceback by both DPPM and IDPPM, denoted by $E(N_{dppm})$ and $E(N_{idppm})$, respectively, depends on the leftover probability.

We learned from in (2) that the leftover probability of all routers on the attack path is 1/D.Therefore, we conclude that

$$E(\mathrm{N}_{\mathrm{dppm}}) = \mathrm{D}*\ln\mathrm{D} \tag{7}$$

For IDPPM algorithm,

$$\alpha_{\mathrm{i}} = \begin{cases} \dfrac{1}{\mathrm{D}} & i = 1,2,3 \\ \dfrac{1}{D} * (\dfrac{\mathrm{D}-1}{\mathrm{D}})^3 & i > 3 \end{cases} \tag{8}$$

Therefore, we can obtain the value of $E(\mathrm{N}_{\mathrm{idppm}})$:

$$E(N_{idppm}) = D*(\frac{D}{D-1})^3 * \ln D \tag{9}$$

It can be seen that IDPPM needs a little more numbers of packets for a successful traceback than DPPM from(7) and (9).Table 1 displays some numerical values of $E(N_{dppm})$ and $E(N_{\mathrm{i}dppm})$ .The table clearly shows that the difference between DPPM and IDPPM decreases gradually with the value of D increasing. $E(N_{\mathrm{i}dppm})$ is 1.95 times as much as $E(N_{dppm})$ when D=5,but it is acceptable for that the amount of $E(N_{\mathrm{i}dppm})$ and $E(N_{dppm})$ is extremely small. And $E(N_{\mathrm{i}dppm})$ is in close proximity to $E(N_{dppm})$ with D =25, 30.

TABLE I. COMPARISION OF $E(N_{dppm})$ AND $E(N_{\mathrm{i}dppm})$

	D					
	5	***10***	***15***	***20***	***25***	***30***
$E(\mathrm{N}_{\mathrm{dppm}})$	9	24	41	60	81	103
$E(\mathrm{N}_{\mathrm{idppm}})$	16	32	50	70	91	113
$\mathrm{D}^3/(\mathrm{D}-1)^3$	1.95	1.37	1.23	1.17	1.13	1.11

5. Simulation

To test the performance of the IDPPM algorithm, we choose to use NS-2.33 to simulate.

And we need to expand the NS2 to evaluate the effectiveness of the PPM, DPPM and IDPPM algorithm. First, the offset (3-bit), the distance (5-bit) and the edge fragment (8-bit) are added into the IP header to be used as marking field. Second, we use the default address format (a 32-bit integer node-id) to identify the node itself. And the PPM, DPPM and IDPPM marking algorithm are injected into the "recv" function in the "trace.cc" file. Then the marking information of each packet is output into the trace file by modifying the "format" function in the "trace.cc" file so as to process all data at centralized locations. Calling the Tcl scripts generates the trace files. Finally, calling awk documents written with the attack path reconstruction algorithm processes the trace file to locate attack sources. The result of simulation is shown in Fig. 4.

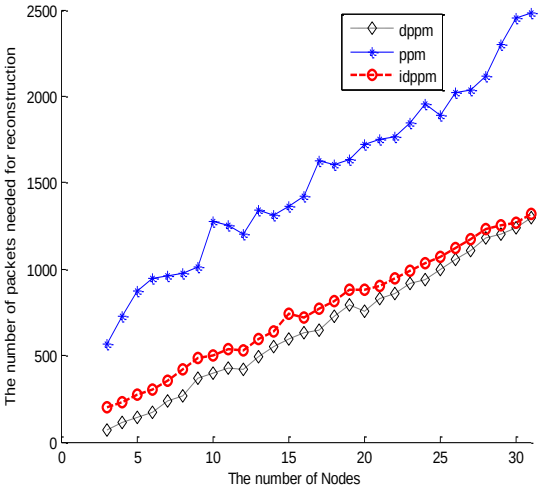

Figure 4. A comparison of numbers of packets required by PPM, DPPM and IDPPM

It can be seen that the IDPPM, just as DPPM, requires obviously much less packets than PPM to reconstruct the attack path. Although the IDPPM algorithm needs a little more packets to traceback than DPPM algorithm, its individual overhead on the routers close to the attacker and the total overhead summed over all routers are less than DPPM algorithm dramatically. This means that the IDPPM algorithm features fewer packets to reconstruct the attack path compared with the PPM algorithm and lower overhead on router compared with the DPPM algorithm.

6. Conclusion

In this paper, we introduce the IDPPM algorithm to locate the packet flooding attacks source in the Internet. The IDPPM algorithm not only can reduce the marking overhead of routers near the attackers, but also can locate attack source rapidly and accurately. And the challenge of weakest node and weakest link is solved with the price of a little more numbers of packets to reconstruct the attack path. The rate of false positive is reduced obviously with the value of F0F1 (01).

Acknowledgment

This paper is benefited greatly from the help of many different people-far more than can be listed completely here. Still, we would like to thank to LIU Ling for her suggestion to the simulation. This work was supported by The National Natural Science Foundation of China (No.60972011)

References

[1] H. Burch, and B. Cheswick, “Tracing Anonymous Packets to Their Approximate source,”Proc. the 14th USENIX conference on System administration,USENIX Association Press, pp. 319-328,Jul. 2000

[2] S. Savage, D. Wetherall, and A. Karlin, “Network Support for IP Traceback,” Proc. IEEE/ACM Transactions on Networking, IEEE Press, pp. 226-237, June 2001

[3] D. Song,and A. Perrig, “Advanced and Authenticated Marking Schemes for IP Traceback,” Proc. the IEEE INFOCOM, IEEE Press,pp. 878-886,2001

[4] F. Bo,G. Fan,and Y. Min, "Dynamic Probabilistic Packet Marking Based On PPM,"Proc. WMWA 09. Second Pacific-Asia Conference, pp. 289-292 ,June 2009

[5] L. Jenshiuh,L. Zhi-Jian,and C. Yeh-Ching, "Dynamic probabilistic packet marking for efficient IP traceback," Proc. the International Journal of Computer and Telecommunications Networking, Elsevier North-Holland Press,Feb 2007,pp. 866-882, doi: 10.1016

[6] G. Dapeng,Y. Shicai,and Y. Wenzhi, "Research on Composed Packet Marking for IP Traceback Algorithm,"Computer Engineering, Vol . 35, pp. 115-117 ,May 2009(In Chinese).

[7] A. Boneh,and M. Hofri, The Coupon Collector Problem Revisited Commun[J].Static Stochastic Models,pp. 39-66,1997

[8] K. Park,and H. Lee, "On the effectiveness of Probabilistic Packet Marking for IP Traceback under denial of service attack,"Proc. IEEE INFOCOM 2001, IEEE Press ,pp. 338-347,2001

[9] L. Stoica,and H. Zhang, "Providing Guaranteed Services Without Per Flow Management," Proc. the conference on Applications,technologies,architectures,and protocols for computer communication,ACM Press,pp. 81-94,1999

[10] D. Drew,F. Franklin, S. Adam, "An Algebraic Approach to IP Traceback," Proc. the ACM Transactions on Information and System Security, ACM Press ,pp. 119-137, May 2002

[11] V. Kuznetsov, A. Simkin,and H. Sandstom, "An evaluation of different IP traceback approaches,"Proc. the 4th International Conference on Information and Communications Security,pp. 37-48,2002