Scientific Research

Research of Improved Probabilistic Packet Marking Algorithms

Yan WANG, Lingdi PING

[1]*Dept. of Information Technology and Management Zhejiang Police Vocational Academy*

[2]*Visiting scholar, College of Computer Science Zhejiang University Hangzhou Zhejiang, China;*

College of Computer Science and Technology Zhejiang University Hangzhou Zhejiang, China

Email: wangyan@zjjy.com.cn, ldping@cs.zju.edu.cn

Abstract: IP traceback is the technology to control Internet crime. A promising solution to IP traceback is probabilistic packet marking (PPM). In this paper we present a novel and practical IP traceback algorithm which can improve the PPM convergency and computational overhead. This scheme may also reduce the deployment overhead without requiring the participation of all routers on the attack path. It has been used not only to trace Distributed Denial of Service (DDoS) attacking packets but also to enhance filtering attacking traffic. It has wide applications for other security systems. To the best of our knowledge, this thesis is the first of its kind considering the impact of packet loss in designing packet marking scheme.

Keywords*:* DoS; DDoS; IP traceback; spoofing

1 Introduction

DoS/DDoS attacks had caught international attentions to the vulnerability of the Internet [1]. Such attacks are among the hardest security problems to address because they are simple to implement, difficult to prevent, and very difficult to trace [2]. Accordingly, the research motivation of this dissertation is to explore and conquer a representative network security problem, i.e. DoS/DDoS problem, with IP traceback, and to design an optimal mechanism among many present IP traceback methods as well.

A simple method to solve the problem about forging the IP address of packet is to use router to filter the packets. For the router, the common method is to write down the local address of the router in the packet. This motion is called packet parking. The schemes of marking are roughly divided into two categories at present: Probability Packet Marking (PPM) and Deterministic Packet Marking (DPM).

In order to let the legitimate users get the service from service providers, we must prevent the attacks mentioned above. There are many defense mechanisms [3]. Yet, merely defending the attack may not be enough. It is also important to find the attacker and let it stop attacking or let the victim drop the requests from the attacker. The major purpose at present is to find the attack path leading to the router closest to the attacker. We wish to find out a better method in order to accurately point out the location of attacker in this thesis.

2 Related Work on PPM

2.1 PPM Overview

Project support: Zhejiang Police Vocational Academy Project.

We give a general introduction of PPM [4, 5, 6, 7, 8] in the following. PPM was first proposed in [8]. The paper discussed a method based PPM and its performance regarding the convergence time and effectiveness against packet spoofing. It also covered discussion on the multiple sourced packet tracing. As mentioned previously, PPM trades off the amount of marking information and the convergence time. The packets are marked by routers, which the packets pass through, in probability p. Every router generates a random number x *in [0, 1]* as it receives a packet, if $x<p$ then the packet is marked. It is possible that marked packets are overwritten by the subsequent routers. The victim collects the marked packets to construct the attack path. Packets are marked with a marking probability p (which is suggested to be $p=0.04$ in [3]) when they pass through a router in this scheme.

Due to the property of the node based PPM scheme, the number of packets marked by the closest router to the end host is the greatest among all the marked packets. In fact, the end host reconstructs the path from itself to the source by sorting the numbers of packets marked by various routers in descending order. The sorted router sequence represents the path of interest. The expected number of packets required to construct the path is computed as

$$E(N) = O(\frac{\ln(d)}{p(1-p)^{d-1}}) \quad (1)$$

where N denotes the number of packets required for attack path construction, d is the hop count of the path, and p is the marking probability. However, a node based PPM is vulnerable to packet spoofing and packet loss. It is recognized that the great number of packets required for attack path calculation is the major drawback of PPM.

2.2 EPPM and CEPPM Algorithm

978-1-935068-10-5

The method of edge based PPM (EPPM) was proposed in [8, 9]. It allows the routers to mark the information of the two vertices of a hop into the packets. The ro12uter of this scheme will hash its local address and then combine it with that of its immediate predecessor to form the marking information. And, the scheme will divide the marking information into eight parts and attach them to eight packets going to the destination. To reconstruct the path, the end host will carry out array insertion instead of sorting. An edge will be captured as long as one set of its associated marked packets are successfully received by the end host. Hence, it is more robust against packet loss. However, its convergence time remains the same as that of the node based PPM. In addition, complicated hashing computation in routers is advocated by some variations of the EPPM scheme. Fig.1 shows the numbers of packets needed in various settings. This is an imitation to the simulation done [9]. As the figure presents, hundreds and even thousands of packets are needed to reconstruct the path.

3 Improved PPM Algorithm

An anonymous flooding-type DoS/DDoS attack is to occupy victim's resources with considerable spoofed packets to block services from normal users. A better way to stave DoS/DDoS attacks is to detect and isolate the attack origin, rather than just to alleviate the disaster. Forenamed IP trackback is one such technique for identifying the true attack source to make the attacker accountable. Despite lots of proposed IP traceback approaches, PPM, DPPM, is still better since it is simple to incrementally implement, doesn't need any additional bandwidth or storage, and can be performed "post mortem". Accordingly, a study on efficient PPM is a good start for defeating DoS/DDoS attacks.

For developing an optimal and robust PPM, there are four basic problem required for overcoming. These challenges/problems are derived from four PPM criteria. In this article, one important method for improving two of four PPM challenges, i.e. the convergency problem and computational overhead, will be illustrated.

3.1 PPM Criteria

In PPM, there are four important criteria required for satisfaction:

- The convergent amount of marked attack packets
- The computational overhead for reconstructing the attack path
- The robustness against the false positive/negative
- The incrementally deployment

In each criterion, there are different subjects required to overcome. These subjects are also good research topics. Accordingly, we would first focus on improving the first PPM criteria: improving the convergent amount.

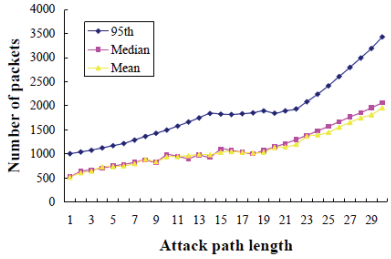

Figure 1. Number of packets required for Compressed EPPM(CEPPM)

3.2 Proposed Methods

We introduce one mechanism, called Compensating Probability Packet Marking (CPPM). CPPM is a heuristic philosophy for achieving the optimal convergent quantity of marked packets and minimizing the path-reconstruction time through compensating every PPM-capable router's marking probability that may be impaired by following PPM-capable routers because of the preemptive property of the PPM-capable routers. CPPM is proposed such that the marked packets can also be fully compensated while they are remarked. In CPPM, a router maintains a compensation table, formatting each entry *e* as (*e.start, e.end, e.distance, e.counter*), to record the information of marked packets, which are remarked by this router.

An incoming packet is re-marked by a CPPM-capable router if $x < p$, where x denotes a randomly generated number between 0 and 1. If this incoming packet is a marked packet barked by some previous CPPM-capable router, the original marked message in the incoming packet will also be recorded in a compensation table owing to being re-marked. The reduced marking probability is compensated according to the compensation table, while each marked-free packet is received. The CPPM marking procedure is shown in Fig.2.

CPPM marking procedure at router R with the marking probability p :

4 Performance and Evaluation of Proposed Algorithm

4.1 Number of Attack Packet Required to Constuct Full Attack Path

Schemes of IP traceback which utilize probability have the same problem. Because the packets received in the victim are marked probabilistically, the number of packets marked by a router will become geometrically smaller when it locates further from the victim. The probability of the victim receiving marked packet from the furthest router is $p(1-p)^{d-1}$, for d is the distance between the router and the victim, and p is the marking probability. If we want to receive one marked packet from the furthest router, we must wait over than

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 T be the compensation table in this router
if $x < p$ then {
if $w.start <> 0$ then {
for each entry e in T if $(w.start, w.end, w.distance) == (e.start, e.end, e.distance)$ then
increment $e.counter$
else
insert $(w.start, w.end, w.distance, 1)$ into T
}
$w.start = R, w.distance = 0$
} else {
if w is a marked-free packet and T is not empty then {
select a entry e by round robin from T
$(w.start, w.end, w.distance) == (e.start, e.end, e.distance)$
$e.counter = e.counter - 1$
remove the entry while $e.counter$ is zero}
if $w.distance = 0$ then {write R into $w.end$ }
increment $w.distance$
}

Figure 2. CPPM marking algorithm

$(p(1-p)^{d-1})^{-1}$ packets. Like the well-known coupon collector problem, required to select one of each d equiprobable items in the number of trials is $d(\ln(d)+O(L))$. Therefore, the number of packet required for the victim to rebuild an attack path of the length d has the following bounded expectation:

$$E(X) < \frac{\ln(d)}{p(1-p)^{d-1}} \quad (2)$$

For example, if $p=0.1$ and path d is 10, then a victim can reconstruct this attack path after receiving 52 packets from the attacker on average. If $p=0.15$ and the attack path d is 10, then a victim can rebuild this attack path become 58 packets. If $p=0.2$ the attack path d is 10, then a victim can rebuild this attack path become 85 packets. Every scheme with probabilistic marking packet has the same assumption. The expected number of packets for compressed edge fragment sampling is similar to the edge marking approach. The expected number of packets required for attack path rebuild is bounded by:

$$E(X) < \frac{k * \ln(kd)}{p(1-p)^{d-1}} \quad (3)$$

where k is the fragment for each edge information of router, and kd is the total number of fragments. For example, if there are 8 fragments per router and, $p=1/10$, and an attack path d is 10, then the victim can rebuild the full attack path after receiving packets. Fig.3 is the simulate results of required packets when rebuild a full attack path for CPPM (Compensating Probabilistic Packet Marking), EPPM (Edge Probabilistic Packet Marking) and CEPPM (Compressed Edge Probabilistic Packet Marking). Where p is 0.04 and d is 30.

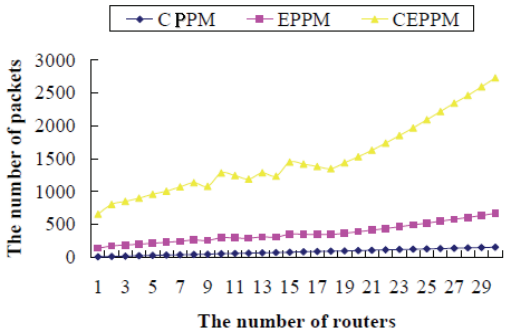

Figure 3. Required Packets for Rebuild Attack Path

4.2 Packet Loss Simulation

The packet loss simulation presented in this section focuses on the required packets to rebuild the attack path given probabilistic packet loss. The packet loss occurs when the network has heavy load. The assumption made here is that the network load is heavy and produces various levels of packet loss. The simulation experiments carried out in this section include the following.

- Evaluating the proposed scheme in different packet loss severity.
- Comparing the performance against that of EPPM and CEPPM.

We start by discussing the assumptions, notations, and parameters used in our simulations. Then, we describe the simulation results.

1) Simulation Method

In each simulation experiment, we assume single attacker and single victim. We performed simulation experiments using the following notations and assumptions.

a) The attack path length, *d*, indicates the number of hops between the attacker and the victim.

b) In EPPM and CEPPM, packets are marked according to the marking probability *p*.

c) The load of the network is heavy and the probability of packet loss is between 3% and 6%.

Each of the following simulation result represents the average of 1000 independent runs.

2) Simulation Results

Fig 4, Fig.5, Fig.6 and Fig.7 show the required number of packets for CPPM, EPPM and CEPPM individually when a victim rebuilds the attack path. *d* is set to 30 while there are 3%~6% chances of packet loss. These figures correspond to single attacker case. It is shown that the higher the packet loss probability the more packets are required for rebuilding the path. Also, it seems EPPM can withstand mild packet loss while CEPPM is the most sensitive if we count the additional packets required as the packet loss ratio increases. How-

978-1-935068-10-5 © 2010 SciRes.

ever, with 1% increment in the packet loss ratio, roughly additional 15% packets are needed for CEPPM. CPPM greatly outperforms these two schemes and is virtually invariant to packet loss in our simulations.

The same real-time NS-2 simulator [10] is used to verify CPPM. The experimental linear topology comprises one attacker, one victim, and ten intermediate routers. The attack traffic rate is 1000 UDP attack packets per second. The marking probability is 0.1. Following ten simulation runs, the mean value is roughly 30, and the standard deviation is 8.6. Because the idea result is about 29, the simulation results presented here also prove that CPPM can approach the optimal situation

5 Discussion and Conclusion

In this thesis, we introduce the Compensating Probability Packet Marking (CPPM). To the best of our knowledge, CPPM is one of the pioneers in considering packet loss and other realistic network conditions. We show that it exhibits extremely fast convergence time, a small fraction of that of the prevailing PPM based schemes. It incurs mild computational overhead in the routers while having comparable complexity in the end host. Unlike its PPM brethren, CPPM can be used to build a much complete path by the end host. It is also demonstrated that it is resilient against packet loss. In this thesis, we made qualitative discussions on random activation, packet loss, and fragmentation.

Other related performance issues concerning dynamic routing changes and partial implementation will be

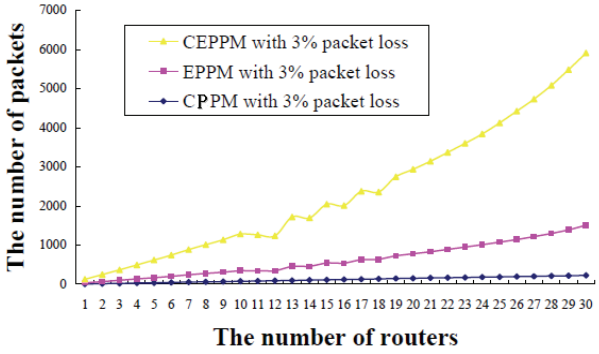

Figure 4. 3% Packets Loss for Various Schemes

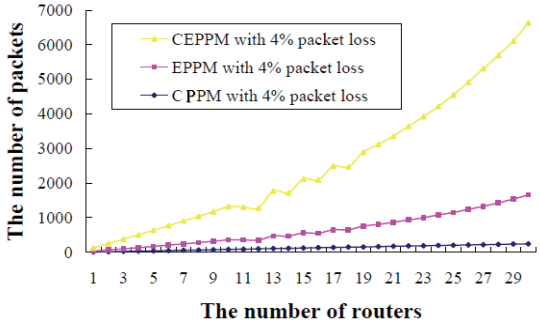

Figure 5. 4% Packets Loss for Various Schemes

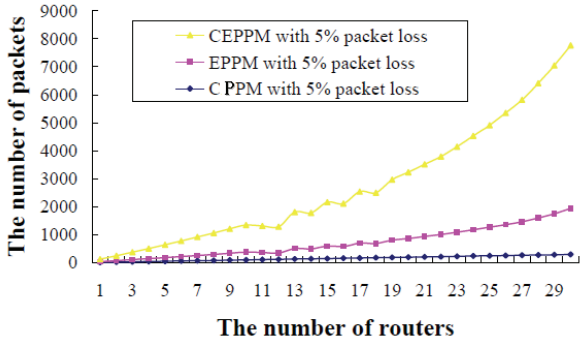

Figure 6. 5% Packets Loss for Various Schemes

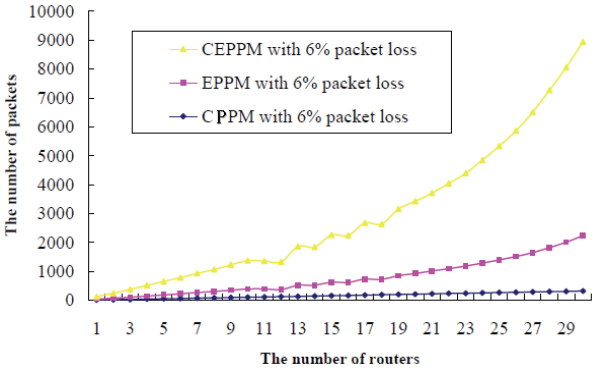

Figure 7. 6% Packets Loss for Various Schemes

further studied in the future research. We will also see how CPPM can be applied to the multiple sourced attacks.

Acknowledgment

We are grateful for helpful conversations with Professor PAN Xue-Zeng, Dr. Li Zhuo and Dr. Li Wen. As a visitor scholar, Professor PING Ling-di and PAN Xue-Zeng give patience directions and good laboratory circumstances to us. FAN Rong and Li Wen provide us much useful information.

References

[1] Harrison, A., (2000). "Cyberassaults hit Bur.com, eBay, CNN and Amazon", computerworld, [online]. Available: http:// www.computerworld.com/news/2000/story/0,11280,43010,00.html

[2] Savage, S., Wetherall, D., Karlin, A., &Anderson, T., (2001). "Network support for IP traceback", IEEE/ACM Trans. Network., Vol.9,No.3, pp.226-237.

[3] A. Belenky and N.Ansari, "On IP traceback", IEEECommunications Magazine. Vol.41, Issue 7,pp.142-153, Jul.2003.

[4] Q. Dong, M. Adler, S. Banerjee, and K. Hirata, "Efficient Probabilistic Packet Marking," in the 13th IEEE International Conference on Network Protocols. Nov. 2005.

[5] K. Park and H. Lee, "On the Effectiveness of Probabilistic Packet Marking for IP Traceback under Denial of Service Attack," IEEE INFOCOM, Vol. 1, pp. 338–347, Apr. 2001.

[6] B. Al-Duwairi and M. Govindarasu, "Novel Hybrid Schemes Employing Packet Marking and Logging for IP Traceback," IEEE Transactions on Parallel and Distributed Systems. Vol. 17, No. 5, May 2006.

[7] C. Gong and K. Sarac, "IP Traceback based on Packet Marking

and Logging," IEEE Communications Magazine, Vol. 2, pp. 1043–1047, May 2005.

[8] M. Sung and J. Xu, "IP Traceback-Based Intelligent Packet Filtering: A Novel Technique for Defending against Internet DDoS Attacks," in the Proc. of IEEE Transactions on Parallel and Distributed Systems, vol. 14, No. 9, pp. 861–872, Sept. 2003.

[9] S. Savage, D. Wetherall, A. Karlin, and T. Anderson "Network Support for IP Traceback," IEEE/ACM Transactions, Vol. 9, Issue 3, pp. 226–237, Jun. 2001.

[10] LBNL Network Research Group, UCB/LBNL/VINT Network Simulator-ns(version 2), DARPA: VINT project. [Online]. Available: htpp://www.isi.edu/nsnam/ns.

978-1-935068-10-5