A Proposal for Cyber-Attack Trace-back Using Packet Marking and Logging

Li, Pengfei
Graduate School of Information Science and Electrical Engineering, Kyushu University

Feng, Yaokai
Faculty of Information Science and Electrical Engineering, Kyushu University

Kawamoto, Junpei
九州大学大学院システム情報科学研究院

Sakurai, Kouichi
Faculty of Information Science and Electrical Engineering, Kyushu University

出版情報：Proc. of the 10th International Workshop on Advances in Infromation Security, 2016-07-06
バージョン：
権利関係：

A Proposal for Cyber-Attack Trace-back Using Packet Marking and Logging

Pengfei Li † , Yaokai Feng ‡ , Junpei Kawamoto ‡ , Kouichi Sakurai ‡
† Graduate School of Information Science and Electrical Engineering, Kyushu University
‡ Faculty of Information Science and Electrical Engineering, Kyushu University
744 Motooka, Nishi-ku, Fukuoka 819-0395 Japan

***Abstract*—Cyber-attack incidents have become more and more frequent and serious. As a countermeasure against cyber-attacks, the technology of (IP address etc.) trace-back to the attackers is essential. Although many methods have been proposed for this purpose, the existing techniques suffer from the following problems. Only the specific attacks can be traced back. The tracing back is too time-consuming and correct traffic-path reconfiguration cannot be guaranteed. In this study, we propose a new method to discover attackers quickly and correctly. By using simulation data, its performance is demonstrated.**

Keywords—cyber-attack, trace-back, packet marking, logging.

I. Introduction

To fight against cyber-attacks, only defending is no longer enough. Technologies to find the attackers out and to pursue the responsibility of the attackers by law are also required. Of those technologies, IP trace-back is thought the most important. Several methods for IP trace-back have been proposed: packet marking [1][2], logging [2], ingress filtering [3], sleep watermark tracing [4], ICMP trace-back [5] and so on. Furthermore, hybrid methods called Hybrid Single-Packet IP Trace-back [6] and precise and practical IP trace-back approach [7] also have been proposed using both packet marking and logging. The most popular technologies in existing methods are packet marking and logging. However, existing trace-back methods are too time-consuming and only can trace back some certain attacks. Obviously, not only tracing back accurately, but also tracing back immediately is important to a real trace-back system. If not be traced back quickly, attackers may erase the evidence and cause other damages and problems.

In order to find the attackers out as soon as possible, in this paper, we propose a new method for IP trace-back. In our proposal two marks are introduced to every packet. During the packet flows in the network, each router these packets passed records (logs) the IP address of the previous router and, at the same time, write the IP address of its own to the second mark of this packet. In this way, the packet can be traced back with small amount of calculation.

The rest of this paper is organized as follows. In Section 2, the relevant existing works are introduced. In Section 3, the proposed method and its benefits will be described in detail. In Section 4, based on a simulation experiment, performance of our proposal will be described. In Section 5, the paper will be summarized. Finally, in Section 6, deployment issue of our proposal and future work will be discussed.

II. Related Work

To trace the attackers back, evidences of the attacks are needed. In order to obtain evidences, the most typical way is to record the communication path, and reconstruct communication path using the records when being attacked. To record the necessary information for tracing back, two methods including writing some information to the packet (called Packet Marking) and recording some information to the routers (called Logging) are used. Here we will introduce some of existing trace-back methods.

A. Packet Marking

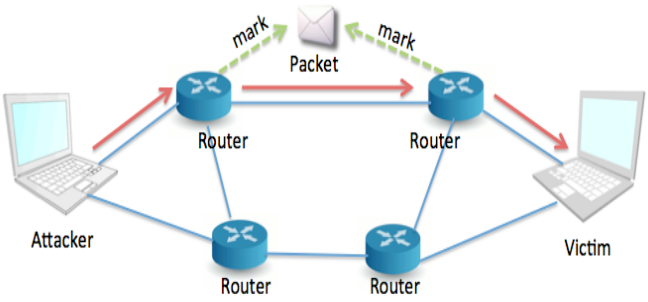

Fig. 1. Packet Marking

Packet marking is a technique to mark the IP address of the router on the packets when they pass a router. Fig. 1 shows an example. In this example, the router writes its information (such as its IP address) into the packets passing through it. According to the information marked in the packets received by the terminal, the attack path could be estimated. There exist two types of Packet Marking, PPM (Probabilistic Packet Marking) method [8] that marks packets at a certain probability and DPM (Deterministic Packet Marking) method [9] that always marks packets. In PPM, routers mark packets at a certain probability with the necessary information of the router. PPM is being paid much attention to because it can reduce the

cost using lower probability (in existing studies, the probability is often set to 1/25). In DPM, when packets pass through an edge router [10], the edge router must mark the packets with the information of the edge router.

However, each Packet Marking method has its own problems. PPM method needs a large amount of calculation when reconstructing the attack path and thus, trace-back is slow [11]. Since the packets are marked in every router on the way, the marks attached to the packet will probably increase rapidly. When reconstructing the attack path, the addresses (marks) have to be calculated one by one. On the other hand, performance of DPM depends on the performance of the edge router.

B. Logging

Logging is a technique that every router records the information such as IDs of the packets passing the router [2]. When packets pass through a router, the router records information of the packets in a certain format. Not only flooding attacks sending a large number of packets, also the attacks even sending only a single packet can be traced by Logging. In this method, a large amount of packet information has to be logged. Thus, the burden on the routers will be increased, which is a big weakness of this method. If a router tries to record the information of all packets, a high-performance router is needed in both of computation speed and storage size, which makes the cost very high. In fact, even for high-performance routers, it is also impossible to record the information of all the packets in the cases of high-speed and large amounts of attacks such as DDoS attacks.

C. Ingress Filtering

Ingress Filtering [3] is one approach used by many Internet service providers to try to prevent the forwarding of IP packets that spoofing the source addresses of IP packets. For example, there are several ways to spoof the source addresses of the IP packets when doing a denial of service (DoS) attack on the Internet. In order to prevent such kind of attacks, Ingress Filtering is recommended. Ingress Filtering can filter the packets that are going into the router from a network. When forwarding a packet, validate whether the source address of a packet is assigned to the network or not. If it is assigned forwards the packet. If not, reject the packet. If Ingress Filtering can be set correctly in the network devices such as routers, the attacks by spoofing the source address can be prevented.

D. Sleep Watermark Tracing

Sleep Watermark Tracing [4] is an active network-based intrusion-response tracing framework for the real-time tracing that when no intrusion is detected in the network, the function of sleep watermark tracing except detection will not start to work. And when an intrusion is detected, the other parts of sleep watermark tracing will be waked up. The core of sleep watermark tracing consists of three interacting components: sleepy Intrusion Response for accept tracing request from intrusion detection system, Watermark Correlation for correlating incoming and outgoing connections through watermarks and Active Tracing for tracing incoming path and source of intrusions. In fact, this method did not propose an effective detection for sleep watermark tracing. So, the disadvantage of sleep watermark tracing is the core of sleep watermark tracing may not be waked up when intrusion comes, the worse is trace-back even not starts.

E. ICMP Trace-back

ICMP, the full name is Internet Control Message Protocol, is one of the protocols used in the Internet process, and is used for notifying the information about Internet communication and errors in datagram processing of Internet Protocol. ICMP trace-back [5] is a method that for every router to sample with low probability, when the router forwarding a packet, the method will copy the contents into a special ICMP trace-back message including information about the adjacent routers along the path to the destination. During a flooding-style attack, the path back to the attacker can be reconstructed with these messages. However, the problem of ICMP trace-back is that the structure of ICMP traffic is different with normal traffic, so the capacity of ICMP trace-back may be limited when tracing back in a normal traffic.

F. HIT and PPIT

HIT [6], the full name is Hybrid Single-Packet IP Trace-back, is proposed by Chao Gong, et al. HIT is a kind of probabilistic packet marking based on both Packet Marking and Logging. In HIT, there are two operations each trace-back enabled router should commit: packet marking and logging. When forwarding a packet, routers decide to mark or log the packet depending on whether the marking field of the packet has enough space available. If it has, routers mark the packet; otherwise, routers log the packet and clear the marking field. And PPIT [7] is a method based on HIT proposed by Dong Yan, et al. PPIT is a development of HIT, at first, PPIT can trace attacks with a large number of packets such as DDoS. And it solves the problem of HIT rewriting the Time to Live (TTL) in the IP header to reconstruct the path that HIT neglects the accuracy of path reconstruction, may lead to the failure of trace-back. At last, PPIT can trace back with more economical storage overhead than HIT. However, PPIT has same problems with HIT. The burden of some certain routers is heavy and much computation of reconstruction is needed. And too much work done on the whole path, the speed of trace-back will be decreased while the risk of miss trace will be increased.

III. Cyber-Attack Trace-Back Using Packet Marking and Logging

The basic idea of our proposal is as follows. Two marks are introduced to each packet, called M1 and M2 (see Fig.2). At first, the two marks are empty. When the packet reach the first router, its M1 and M2 are marked into the IP address of the first router (at that time, M1=M2). Then, when the packet passes the other routers, two operations are needed in the router: one is logging M2 to the current router and writing the IP address of the current router as new M2. And the other one is recording a part of the information of the packet to the

current router (logging). See Fig. 2. In the first router R_1 that the packet P passed, the IP address of R_1 was written to P at a constant probability p as M1 and M2, while logging the source IP of P to R_1. When the packet passed the next router R_2, R_2 checks M1 of P, if R_1 didn't marked P, the following routers also don't mark the packet. If R_1 marked P, R_2 also marks the packet. That is, marking M2 of P with the IP address of R_2. At the same time, the information such as M2 (IP address of R_1) is logged on R_2. In this way, the router R_n logged M2 of P (the IP address of router R_{n-1}) and re-marked M2 of P with the IP address of R_n. In other words, M2 in P was changed from the IP address of R_{n-1} to the IP address of R_n. In this way, the contents of M2 in P would continue to change, and in each router the IP address of the previous router was logged. Finally, the information of the first and last of the router addresses and the entire path (router list) of each packet would be recorded.

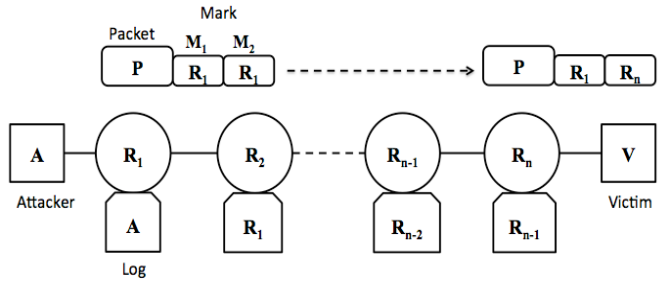

Fig. 2. The basic idea of our proposal

The reason of marking with a certain probability p in R1, is attackers (especially, DDoS attackers) are assumed to send a large number of packets, it is possible to fully trace even if only a few of packets have been marked. In this way, we can reduce the burden on each router. The probability p could be adjusted. Like the existing PPM method, the larger the amount of packets the assumed attackers send, the lower the probability p would be. In that case, even there is something of trouble in the R1 that cause R1 could not mark, because the next router would mark the packets, when tracing back, the first router could also be found out when we trace back.

There are five benefits in the proposed method,

- Fast tracing

  With the information of the first mark area of the packet, we could ignore the detail of attack path, could find out the first router and the source IP address the attack has been logged in the first router.

- Strong against IP spoofing

  In our proposal, the entire attack path is distributed recorded in the log of each router, even if the first step is forged by attacker, the entire path could not be able to forged, therefore this proposal is also against to IP Spoofing.

- Small amount of calculation needed to reconstruct attack path

  For marking, no complex calculations are needed. And for path reconstruction, complex calculations are also not required.

- Light burden on packets

  Only two mark areas in the packet, compared with other Probabilistic Packet Marking, it is very light.

- Distributed stored attack path

  This makes our proposal robust. The roles of the routers are being shared, the risk are reduced. The roles of the routers are totally same. And the most important mark is IP address of router R_1, but not essential if IP addresses of other routers are marked correctly. Other marks are also not essential too if IP address of router R_1 are marked correctly.

IV. Simulation

We design a simulation experiment for determining our packet tracing proposal to know what would effect the success rate of trace-back and how much be effected to while the value of the parameters changed. Our simulation is executed 10000 times. In our proposal, if the information of a router were not marked into a packet, the information of packets would also be not logged into that router. Furthermore, even if the whole path of only one packet was marked successfully at one attack, the attackers can also be traced back. Thus, success rate of trace-back is determined by whether such a packet exists or not that it's whole path was successfully marked.

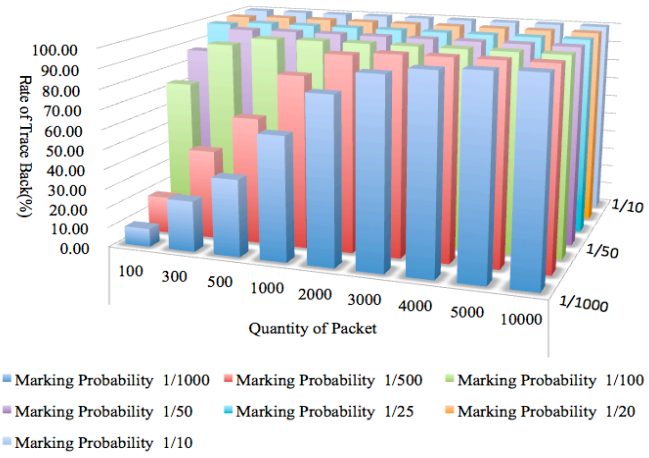

Fig. 3. The average result of 10000 simulations *(Diagram)*

In our simulation, when the number of packets is few, setting marking probability to 100%, and as the number of packets increases, marking probability will decrease. We increased the number of packets from 100 to 10000, and we executed the simulation for monitoring the number of marks when the number of the packets is 100, 300, 500, 1000, 2000, 3000, 4000, 5000, 10000 in different probabilities from 1/1000 to 1/10. Figures 3 and 4 shows the simulation result. When the marking probability is 1/10, the number of packets is increased from 100 to 10000. And the simulation was executed 10000 times, the attack can always be traced every times. This

indicates that 1/10 is a high-enough marking probability even for a relatively small number of packets. In the case of Probability is 1/20, when the number of communication packets is 100, the success rate of trace-back got down to 99.37%. And while the number of packets is increased to 300, the success rate was back to 100%. This indicates that 1/20 is also high enough for not very small number packets. When probability is 1/50, the success rate can be sustained to 100 percent after the number of packets is more than 1000. In addition, when the probability is 1/500 or 1/1000, the success rate can be sustained to 100% only when the number of packets reached 10000.

#Packet	Marking Probability						
	1/10	1/20	1/25	1/50	1/100	1/500	1/1000
100	10000	9937	9854	8733	7356	1867	918
300	10000	10000	10000	9976	9544	4518	2579
500	10000	10000	10000	9999	9955	6381	3907
1000	10000	10000	10000	10000	10000	8673	6264
2000	10000	10000	10000	10000	10000	9821	8398
3000	10000	10000	10000	10000	10000	9981	9486
4000	10000	10000	10000	10000	10000	9997	9817
5000	10000	10000	10000	10000	10000	9998	9934
10000	10000	10000	10000	10000	10000	10000	10000

Fig. 4. The number of times in 10000 simulations that trace-back was successful

We can know from the simulation results shown in Fig. 3 and Fig.4 that in each probability, the fewer the quantity of packets is, the lower success rate will be. We suppose that the probability should be adjusted automatically according to the real situations. This will be done in our future work.

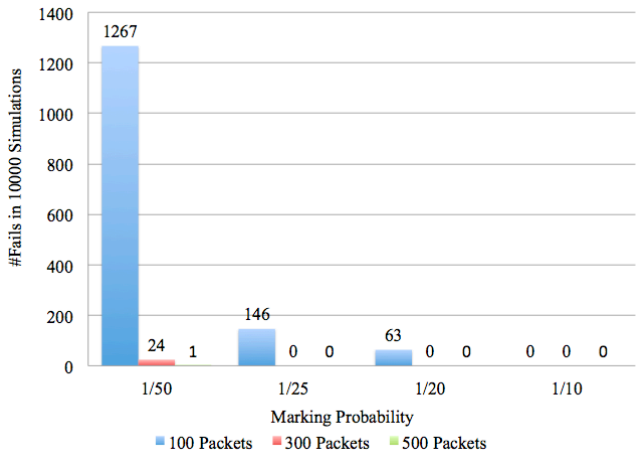

Fig. 5. The times of fails in 10000 simulations

Figure 5 shows the result of another simulation that the number of packets increases from 100 to 500 in probability 1/10, 1/20, and 1/25 and 1/50. As it can be seen from the results shown in Fig. 5, in order to sustain the high success rate, when the number of packets is 100 or fewer, the marking probability should be 1/10 or higher. When the number of packets is 300, even if marking probability is down to 1/25, attack could still be traced. When the number of packets is 500, the marking probability can be down to 1/50. In Figure 6, while increasing the number of packets from 100 to 10000, it shows the number of packets successfully marked at each probability. From the simulation result, we knew that, as the number of packets increases, if the marking probability is fixed, the burden on router is increased. In addition, the larger marking probability is, the heavier the burden on router is significantly.

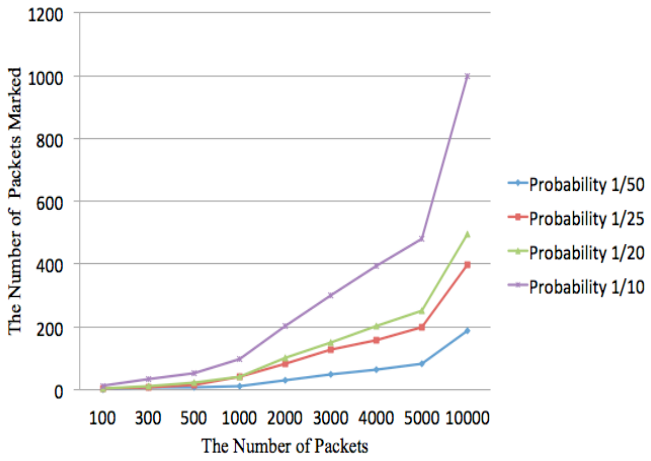

Fig. 6. The number of packets marked successfully vs. the number of total packets.

By the above simulation, we know that the success rate of trace-back is determined by two parameters - the number of packets in an attack and the marking probability. If the number of packets is smaller, higher marking probability is needed, while the number of packets is increased to a certain degree, in order to reduce the burden on the router, marking probability should be cut down.

V. Conclusion and Future Work

In order to solve the problem in existing trace-back technologies that only can trace back some certain attacks, need much calculation for reconstructing the communication path. We proposed a new method and confirmed the effectiveness of the proposed method by simulation.

In our proposal, less calculation is needed for path reconstruction and thus, the trace-back is fast. In order to distribute the risk of trace-back being failed, the burden is distributed on the routers. And only two marks are needed in each packet. Finally, as a result, in the case of IP spoofing attacks, the performance of trace-back is a little affected. Basically, our proposal is strong against to IP Spoofing.

In existing trace-back technologies, marking probability is always a constant (in general, the probability is set to 1/25). In existing trace-back technologies, marking probability is always a constant (in general, the probability is set to 1/25). And in some papers, marking probability can be changed when the

length of attack path changed. Such methods are named Dynamic Probabilistic Packet Marking [12]. In order to reduce burden on routers, in our future study, marking probability will be adjusted automatically according to the real situations.

REFERENCES

[1] H. Burchand and B. Cheswick, "Tracing anonymous packets to their approximate source", in Proceedings of the 14th USENIX Conference on System Administration, 2000.

[2] R. Jain and A. Meshram, "A Survey on Packet Marking and Logging", (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 4 (3), 2013.

[3] P. Ferguson and D. Senie, "Network Ingress Filtering: Defeating Denial of Service Attacks Which Employ IP Source Address Spoofing", RFC 2267, Jan. 1998.

[4] X. Y. Wang, D. S.Reeves, S. F. Wu and J. Yuill, "Sleepy watermark tracing: An active network-based intrusion response framework", In Trusted Information. Springer US, p. 369-384, 2002.

[5] S. M. Bellovin, "ICMP traceback messages", Network Working Group Internet Draft, March 2000.

[6] C. Gong and K. Sarac, "A more practical approach for single-packet IP traceback using packet logging and marking", IEEE Transactions on Parallel and Distributed Systems, Vol. 19, 2008.

[7] Y. Dong, Y. Wang, S. Su, and F. Yang, "A Precise and Practical IP Traceback Technique Based on Packet Marking and Logging", Journal of Information Science and Engineering March 2012.

[8] S. Savage, D. Wetherall, A. Karlin and T. Anderson, "Practical Network Support for IP Traceback", ACM/IEEE Trans. Networking, vol. 9, no. 3, pp. 226-237, 2001.

[9] A. Belenky and N. Ansari, "IP Traceback With Deterministic Packet Marking", IEEE Communication Letters, Vol.7, No.4, Apr.2003.

[10] F. Xue and S. J. Ben Yoo, "Self-similar traffic shaping at the edge router in optical packet-switched networks", Communications, 2002. ICC 2002. IEEE International Conference on. Vol. 4. IEEE, 2002.

[11] X. Yang, W. Zhou and M. Guo, "Flexible Deterministic Packet Marking: An IP Traceback System to Find the Real Source of Attacks", IEEE Trans. on Parallel and Distributed Systems. Vol. 20, No. 4, Apr. 2009.

[12] Y. Qiao, X. He, and T. Ning, "An Improved dynamic probabilistic packet marking for IP traceback. " International Journal of Computer Network and Information Security, 2010.