Router and Interface Marking for Network Forensics

Emmanuel Pilli, Ramesh Joshi, Rajdeep Niyogi

Chapter 16

ROUTER AND INTERFACE MARKING FOR NETWORK FORENSICS

Emmanuel Pilli, Ramesh Joshi and Rajdeep Niyogi

Abstract The primary aim of network forensics is to trace attackers and obtain evidence for possible prosecution. Many traceback techniques exist, but most of them focus on distributed denial of service (DDoS) attacks. This paper presents a novel traceback technique that deterministically marks the interface number and the address of the router from which each outgoing packet entered the network. An analysis against various traceback metrics demonstrates that the technique enhances network attack attribution.

Keywords: Network forensics, traceback, attack attribution

1. Introduction

IP traceback mechanisms attempt to identify the source of attacks, and implicate and prosecute the attackers. The identification of attack hosts and networks is not a major achievement, but the essential clues it provides can help identify the actual attackers. Once realized, IP traceback can be a major component of a network forensic investigation.

Several techniques exist for performing a traceback [8]. However, TCP/IP limitations facilitate IP spoofing, which manipulates the source address in the IP header. Since the routing infrastructure of the Internet is stateless and packet routing decisions are based on the destination, there is no entity responsible for ensuring the correctness of the source address. As such, attackers can generate malicious IP packets with arbitrary source addresses. This makes the reconstruction of the path back to the attack origin a challenging task.

This paper presents a traceback technique involving deterministic router and interface marking (DRIM). The DRIM technique deterministically marks the interface number and the address of the router from

which each outgoing packet entered the network. Every outbound packet is marked at the first ingress edge router. Inbound packets are not marked. Once a packet is marked, other routers do not mark the packet. The marking enables traceback to the ingress router closest to the attacker and identifies the attack path to the source via the interface number. This traceback technique is the first to use both deterministic packet marking and interface marking.

2. IP Traceback for Network Forensics

Network forensics deals with capture, recording and analysis of network traffic. The network forensic process analyzes network log data to characterize attacks and identify the perpetrators. It involves monitoring network traffic, determining if anomalies are present and ascertaining if the anomalies indicate an attack. The ultimate goal is to obtain evidence to identify and prosecute the perpetrators [16].

2.1 Classification of Network Forensics

Network forensic systems are classified into various types based on their characteristics [11]. This classification is useful to identify the set of requirements and make assumptions for traceback in the context of network forensic analysis.

- **Purpose:** General network forensics focuses on enhancing security by analyzing network traffic to discover attack patterns. Strict network forensics involves rigid legal requirements, as the results are used as evidence in court.

- **Packet Capture:** Catch-it-as-you-can systems capture and store packets passing through a particular traffic point. Stop-look-and-listen systems analyze packets in memory as they pass and store limited information about the packets.

- **Platform:** A network forensic system can be a hardware appliance or it can be a software system that is installed on a host to analyze stored packet captures or netflow records.

- **Time of Analysis:** Commercial network forensic systems involve real-time network surveillance, signature-based anomaly detection, data analysis and forensic analysis. Many open source software tools exist to perform *post mortem* investigations of packet captures. The tools perform packet analysis of data captured by sniffer tools.

- **Data Source:** Flow-based systems collect statistical information about network traffic as it passes through a capture platform. The network equipment collects the data and sends it to a flow collector, which stores and analyzes the data. Packet-based systems capture full packets for subsequent deep packet inspection.

2.2 Assumptions

Packet-based systems can provide detailed information about attackers while requiring less resources in *post mortem* investigations. This paper focuses on *post mortem* packet-based network forensics. The following assumptions are made regarding traceback:

- Attackers can generate and send any packet.
- Attackers are aware of the traceback ability.
- Routers possess limited processing and storage capabilities.
- Not all routers participate, but the host router in the attacker's network must participate.
- Routes between hosts are stable, but packets can be reordered or lost.
- An attack packet stream may only comprise a few packets, but an investigation must be conducted despite the limited evidence.

2.3 Requirements

The indispensable requirement for network forensic traceback is that the routers in the attacker's network must use the marking mechanism. Other requirements for IP traceback include:

- Compatibility with existing network protocols, routers and infrastructure.
- Simple implementation with a minimal number of functions.
- Support for partial deployment and scalability.
- Minimal time and resource overhead (processing, bandwidth and memory).
- Fast convergence of the traceback using only a few packets.
- Minimal involvement of an Internet service provider (ISP).

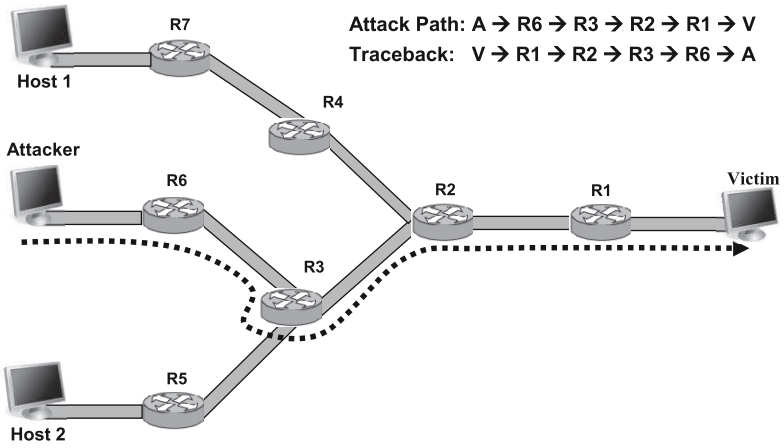

Figure 1. IP traceback mechanism.

- Minimal increase in the packet size due to the traceback mechanism.
- Low potential for evasion by mark spoofing.
- Ability to perform a traceback closer to the attacker than the ingress edge router.

3. Related Work

IP traceback [1, 3, 6] is an important strategy for investigating and attributing network attacks (Figure 1). IP traceback techniques do not prevent and mitigate attacks; instead, they identify the sources of attack packets. IP traceback techniques can be reactive or proactive. Reactive traceback techniques make attack detection decisions while the attack is in progress and require a large amount of traffic (as in a DDoS attack). They use logging, packet marking and hybrid approaches (logging and marking). The techniques fail when attack traffic subsides; therefore, they are not very effective for *post mortem* analysis.

Proactive traceback techniques perform packet marking or interface marking. Packet marking inserts within an IP packet the address of each router along its path. The packets are marked either probabilistically or deterministically. Probabilistic packet marking (PPM) requires many packets for convergence of attacker information. Deterministic packet marking (DPM) techniques need fewer packets for traceback and can be performed *post mortem*.

Belenky and Ansari [2, 4] first proposed DPM, in which only the ingress edge routers mark packets. Each border router marks every

packet with its incoming IP address in the 16-bit ID field as the packet enters the network. Because the IP address requires more than sixteen bits, DPM splits the IP address into two packets and uses the 1-bit reserved flag to indicate the first and second parts of the IP address. Rayanchu and Barua [12] have extended this approach by embedding all the IP information in a single packet. The 16-bit packet ID field is marked with a 16-bit hash of the 32-bit IP address of the edge router. The network maintains a table to identify the IP address based on the packet hash. Lin and Lee [9] have proposed a robust, scalable DPM scheme that uses multiple hash functions to reduce the probability of address digest collisions. Their DPM technique uses three bits to distinguish between eight different hash functions; the remaining fourteen bits carry the hashed address information.

Jin and Yang [7] have proposed a DPM-based redundant decomposition for IP traceback, where the marking field has two sections: information and index. Every ingress edge router decomposes its corresponding IP address into fragments where the neighboring fragments have some redundant bits. The IP ID field is marked with one of the fragments. Xiang, *et al.* [15] have proposed a flexible DPM scheme to identify the source of attack packets. It adopts a flexible mark length strategy for compatibility with different network environments. The scheme also changes the marking rate based on the load of the participating router using a flexible flow-based marking technique.

Router interface marking (RIM) mechanisms consider a router interface (as opposed to the router itself) as an atomic unit for traceback. Chen, *et al.* [5] use a RIM-enabled router to mark each packet with the identifier of the hardware interface that processed the packet. The mark is a locally-composed string of unique router input IDs that serves as a globally-unique path identifier. It uses five bits for distance, six for the XOR value and six for the interface ID. Yi, *et al.* [17] have proposed a DPM technique that marks every packet passing through a router with a link signature (digest of the address information of the two adjacent nodes). Each router participates in marking and the mark changes with each router. The entire path information is available in each packet and single-packet IP traceback is possible.

Peng, *et al.* [10] have proposed an enhanced, authenticated DPM that uses path numbering for traceback. DPM-enabled routers at the edge of a subnet mark each packet based on the incoming interface. PPM-enabled routers are closest to the packet source and mark each packet with path identifiers that represent the path linking them to the DPM-enabled routers. This facilitates attack detection and filtering as well as obtaining accurate information from the authenticated marks.

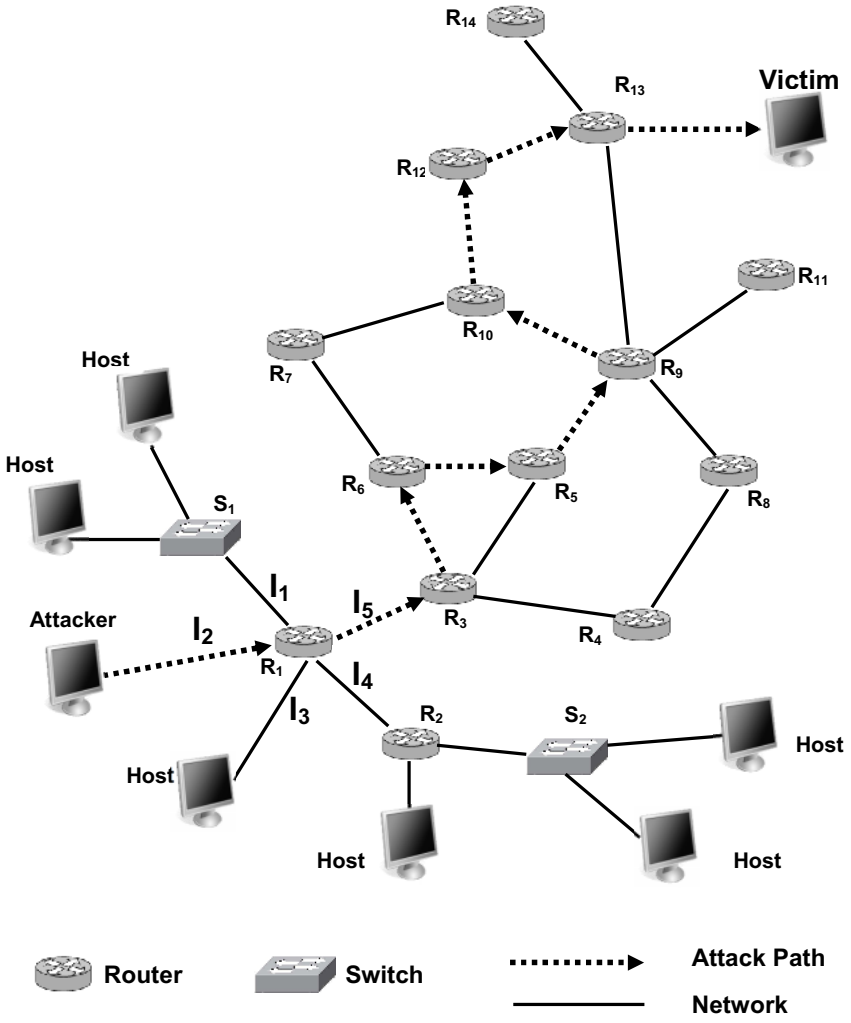

Figure 2. Deterministic router and interface marking.

4. Deterministic Router and Interface Marking

The proposed IP traceback technique deterministically marks each packet with the interface number and the address of the router through which the packet enters the network. Only the first router marks the packet to prevent other routers from overwriting the mark. This makes it possible to perform a traceback beyond the ingress router.

Consider the architecture in Figure 2 with various hosts, switches, routers and interfaces. The attacker is the host that connects to the Internet through ingress edge router R_1. Packets reach the first router

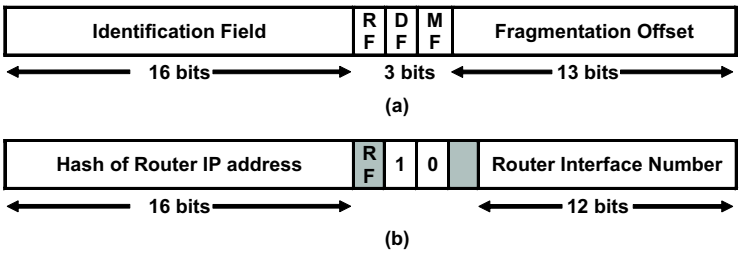

Figure 3. (a) IP header fields; (b) Overloaded fields for marking.

R_1 through interface I_2. The other interfaces I_1, I_3, I_4 and I_5 of the router R_1 are connected to a switch S_1, a host and two routers R_2 and R_3. The interface number I_2 and a hash value of router R_1's IP address are marked deterministically in each packet on the attack path. No other routers (i.e., R_3 to R_{13}) overwrite the mark. Only packets arriving through interfaces I_1, I_2 and I_3 are marked by router R_1 because they belong to the internal network. Packets arriving through I_4 and I_5 connected to routers R_2 and R_3, respectively, are not marked. Each packet is marked only once with two values: the interface number and a hash of the router's IP address.

4.1 Marker Encoding

The marking technique uses 32 bits – corresponding to the 16-bit ID field, 3-bit fragment flag field and 13-bit fragment offset field – in the IP header to store marking information. Figure 3 shows the mapping between the IP header fields and the marking fields. The fragment fields hold information about packet fragmentation. However, fragmented traffic is relatively rare on the Internet (about 0.25 percent of all traffic) [13, 14]. The rarity renders the 32 bits as redundant space in a normal IP header and enables them to be used to store marking information.

As with the technique of Rayanchu and Barua [12], a 16-bit hash value of the 32-bit IP address is embedded in the ID field. An enterprise network grade Cisco router that connects to a maximum number of 4,096 interfaces would use a maximum of twelve bits in the mark. The least significant twelve bits of the 13-bit offset are used to store the interface number. To indicate that the used fields do not contain fragmentation information, the DF bit is set to one and MF bit is set to zero.

Algorithm 1 lists the steps used to encode and mark the IP address and the interface number of the router in each packet.

Algorithm 1 : Marking the address and interface number of router R_i.

for each outbound packet P reaching router R_i through interfaces $I_j \subseteq I_{local}$ **do**
  write $HashIP_{16}(R_i)$ into $P.Identification$
  write I_j into $P.offset[0..11]$
  set $P.DF = 1$
  set $P.MF = 0$
end for

Algorithm 2 : Reconstruction at victim V.

for each attack packet P reaching victim V **do**
  read $HashIP_{16}(R_i)$ from $P.Identification$
  extract IP from $HashIP_{16}(R_i)$
  read I_j from $P.offset[0..11]$
  $IN = I_j$
  return (IP, IN)
end for

4.2 Traceback Operation

The traceback operation (Algorithm 2) is simple because each packet holds the information required to identify the first ingress router and the interface through which the packet reached the router. The 16-bit identification field in the IP header gives the 16-bit hash value of the router's 32-bit IP address. The 12-bit value in the offset field indicates the interface number. The identification of the interface through which the attack packet entered the network places the attacker closer than other traceback techniques that only identify the first ingress edge router. Since each packet has all the marker information, the traceback operation requires only a single packet.

5. Evaluation

The metrics used to evaluate the proposed IP traceback technique are:

- **Number of Packets for Traceback:** Every packet provides information about the attacker. The information includes the ingress router IP address and interface number from which the attack packet entered the network. The technique works for any number of distributed attackers working in coordination.

- **Processing Overhead:** The processing overhead is nominal because the marking operation is a simple function. The overhead

may increase as the router bandwidth reaches its maximum. The overhead can be reduced by precomputing the IP address hash.

- **Storage Overhead:** The technique requires no additional storage beyond the hash value of the router.
- **Infrastructure Changes:** Infrastructure changes are minimal because the technique requires the implementation of only one additional function in the routers. The function to reconstruct the traceback is only required at the victim's end.
- **False Positive Errors:** A false positive error arises when a legitimate client is misidentified as an attacker. Because the routing technique is deterministic, the number of false positive errors is bound by the number of collisions of the hash function.
- **Scalability:** The technique is scalable and can handle multiple attackers because information about the attacker is in each packet.
- **ISP Involvement:** Considerable interaction with an ISP is required to implement the marking function in all routers.
- **Effect of Partial Deployment:** Incremental deployment is limited because the marking is done only once. If the attacker's ingress routers do not perform the marking, then the technique may yield more false positive errors. The assumption that marking occurs in the attacker's network ensures that every packet that reaches the first ingress edge router is marked.

5.1 Comparison with Other Techniques

Table 1 compares the router interface marking (RIM) [5] and deterministic packet marking (DPM) [2] techniques with the proposed deterministic router and interface marking (DRIM) technique. The metrics used for evaluation were originally suggested by Belenky and Ansari [3].

DRIM has several advantages over the other techniques. It can trace the attacker using a single packet and does not require additional memory at the router or at the victim. The marking operation is simple, easily implemented and overcomes mark spoofing. Because the entire marking information is available in a single packet, there are fewer false positive errors. DRIM goes one step beyond other related techniques by identifying the interface from which a packet reached the ingress router. This increases the possibility of tracing an attacker beyond the router, which the other techniques are unable to accomplish.

Table 1. Comparison of RIM, DPM and DRIM techniques.

Metric	RIM	DPM	DRIM
Number of packets	Single packet	Seven packets	Single packet
Processing overhead	Packets probabilistically marked with XOR and interface ID values or XOR value is updated	Packets marked only once with the first or last sixteen bits of the edge router's address	Packets assigned two marks by the first ingress edge router
Storage overhead	Trace table maintains hop count, interface id and XOR value	Table used for matching source and ingress addresses	Hash value of router's address is precomputed and stored
Marking field length	17 bits (handling 64 interfaces)	34 bits in two consecutive packets	31 bits (handling 4,096 interfaces)
Infrastructure changes	One function added to network devices	One function added to network devices	One function added to network devices
False positive errors	Few errors as router interface IDs may not be unique	Two packets carry the router's address and may yield errors	Hashing the router's address yields few errors
Scalability	False positive errors increase with number of attackers	Thousands of attackers can be traced	Any number of attackers can be traced
ISP involvement	Moderate	High	High
Partial deployment	False positive errors decrease with an increase in RIM-enabled routers	Limited	Limited
Mark spoofing	Additional scheme using hash function and offset to prevent mark spoofing	Spoofed marks are overwritten because ingress router determines validity of marks	Spoofed marks are overwritten because interface number determines validity of marks
Extent of traceback	Ingress router closest to the attacker	Ingress router closest to the attacker	Ingress router and interface closest to the attacker

The advantages of RIM are that it can be deployed partially and requires moderate ISP involvement. However, partial deployment is not a disadvantage in the case of DRIM. Partial deployment adversely affects any network forensic technique because it is impossible to attribute the attack to a particular host if the marking mechanism is not in the attacker's network.

6. Conclusions

The deterministic interface and router marking (DRIM) technique can trace an attacker from the ingress edge router usign a single packet, meeting the basic requirement of network forensics. It traces an attacker more closely than other techniques by identifying the interface from which the attack packet arrived at the router. This also overcomes the problem of mark spoofing – the interface number enables the router to overwrite a false mark placed by the attacker. Future research will implement both fragmentation and marking, which will facilitate the incremental deployment of DRIM-enabled routers and reduce ISP interaction.

References

[1] H. Aljifri, IP traceback: A new denial-of-service deterrent? *IEEE Security and Privacy*, vol. 1(3), pp. 24–31, 2003.

[2] A. Belenky and N. Ansari, IP traceback with deterministic packet marking, *IEEE Communications Letters*, vol. 7(4), pp. 163–164, 2003.

[3] A. Belenky and N. Ansari, On IP traceback, *IEEE Communications*, vol. 41(7), pp. 142–153, 2003.

[4] A. Belenky and N. Ansari, On deterministic packet marking, *Computer Networks*, vol. 51(10), pp. 2677–2700, 2007.

[5] R. Chen, J. Park and R. Marchany, RIM: Router interface marking for IP traceback, *Proceedings of the IEEE Global Telecommunications Conference*, 2006.

[6] Z. Gao and N. Ansari, Tracing cyber attacks from the practical perspective, *IEEE Communications*, vol. 43(5), pp. 123–131, 2005.

[7] G. Jin and J. Yang, Deterministic packet marking based on redundant decomposition for IP traceback, *IEEE Communications Letters*, vol. 10(3), pp. 204–206, 2006.

[8] S. Lee and C. Shields, Tracing the source of network attack: A technical, legal and societal problem, *Proceedings of the IEEE Workshop on Information Assurance and Security*, pp. 239–246, 2001.

[9] I. Lin and T. Lee, Robust and scalable deterministic packet marking scheme for IP traceback, *Proceedings of the IEEE Global Telecommunications Conference*, 2006.

[10] D. Peng, Z. Shi, L. Tao and W. Ma, Enhanced and authenticated deterministic packet marking for IP traceback, *Proceedings of the Seventh International Conference on Advanced Parallel Processing Technologies*, pp. 508–517, 2007.

[11] E. Pilli, R. Joshi and R. Niyogi, Network forensic frameworks: Survey and research challenges, *Digital Investigation*, vol. 7(1-2), pp. 14–27, 2010.

[12] S. Rayanchu and G. Barua, Tracing attackers with deterministic edge router marking, *Proceedings of the First International Conference on Distributed Computing and Internet Technology*, pp. 400–409, 2004.

[13] S. Savage, D. Wetherall, A. Karlin and T. Anderson, Network support for IP traceback, *IEEE/ACM Transactions on Networking*, vol. 9(3), pp. 226–237, 2001.

[14] C. Shannon, D. Moore and K. Claffy, Characteristics of fragmented IP traffic on Internet links, *Proceedings of the First ACM SIGCOMM Workshop on Internet Measurement*, pp. 83–97, 2001.

[15] Y. Xiang, W. Zhou and M. Guo, Flexible deterministic packet marking: An IP traceback system to find the real source of attacks, *IEEE Transactions on Parallel and Distributed Systems*, vol. 20(4), pp. 567–580, 2009.

[16] A. Yasinsac and Y. Manzano, Policies to enhance computer and network forensics, *Proceedings of the IEEE Workshop on Information Assurance and Security*, pp. 289–295, 2001.

[17] S. Yi, X. Yang, L. Ning and Q. Yong, Deterministic packet marking with link signatures for IP traceback, *Proceedings of the Second SKLOIS Conference on Information Security and Cryptology*, pp. 144–152, 2006.