An IP Traceback Model for Network Forensics

Emmanuel S. Pilli, R.C. Joshi, and Rajdeep Niyogi

Department of Electronics and Computer Engineering,
Indian Institute of Techology Roorkee, Roorkee, India
{emshudec,rcjosfec,rajdpfec}@iitr.ernet.in, emmshub@gmail.com

Abstract. Network forensics deals with capture, recording, analysis and investigation of network traffic to traceback the attackers. Its ultimate goal is to provide sufficient evidence to allow the perpetrator to be prosecuted. IP traceback is an important aspect in the investigation process where the real attacker is identified by tracking source address of the attack packets. In this paper we classify the various approaches to network forensics to list the requirements of the traceback. We propose a novel model for traceback based on autonomous systems (AS) and deterministic packet marking (DPM) to enable traceback even with a single packet. The model is analyzed against various evaluation metrics. The traceback solution will be a major step in the direction of attack attribution and investigation.

Keywords: network forensics, traceback, DPM, AS, attack attribution.

1 Introduction

IP traceback problem involves identifying the actual source of a packet across the Internet. Many techniques for traceback have been proposed, but all of them are focused on distributed denial of service (DDoS) attacks [1]. Many of the techniques can be slightly modified or extended for handling other attacks as well.

The weaknesses in TCP/IP facilitate *IP spoofing* where source address in the IP header can be manipulated and an address other than the actual can be placed. The routing infrastructure of the Internet is stateless and hence the reconstruction of the path back to the attacker is a non-trivial task. Network address translation (NAT) and stepping-stone attack also complicate the process. IP traceback mechanisms aim at tracking back the source of attacks. If this is realized, IP traceback will be the major part of investigation phase in network forensics.

We propose to use a two level traceback mechanism based on deterministic packet marking (DPM) using an Autonomous Systems (AS). The first level involves marking of each outgoing packet by the first ingress edge router and the second level involves marking each outgoing packet by the AS edge router (ASER). Packet is marked only once at each level. A single packet is sufficient to detect the source and the model is the first of its kind where DPM and AS marking is taken as a combination.

The paper is organized as follows: Section 2 provides a basis for network forensic approach to traceback. Section 3 makes a survey of existing traceback techniques and identifies a model for network forensics. We propose a model for traceback in Section 4. Conclusion and future work are given in section 5.

I. Baggili (Ed.): ICDF2C 2010, LNICST 53, pp. 129–136, 2011.

2 Assumptions and Requirements for Forensic Traceback

Network forensics is the science that deals with capture, recording, and analysis of network traffic [2]. Network forensic systems are classified [3] into different types, based on various characteristics. We extend two more classes:

- Time of Analysis: *Real time* forensics involves live network security surveillance and monitoring. *Post mortem* investigation of packet captures is done offline.
- Data Source: *Flow based* systems collect statistical information as network traffic flows. *Packet based* systems involve deep packet inspection.

We identify a set of requirements and make necessary assumptions for traceback in the context of network forensics. We limit our work to the post mortem, packet based network forensics.

Assumptions: The number of packets generated for DDoS attacks are huge and many techniques have been designed to exploit this situation. Network forensics may handle attacks which may involve very few packets. The assumptions made for designing traceback mechanisms for DDoS are modified to suit the investigation of cyber crimes. They are given below:

- attackers are able to generate and send any packet
- multiple attackers may act in a coordinated fashion
- attackers are aware of the traceback ability
- routers possess limited processing and storage capabilities
- routers are rarely compromised and all routers may not participate in traceback
- suspicious packet stream may consist of just a few packets

Requirements: Goals for effective and efficient traceback can be designed for evaluating existing traceback solutions and build new mechanisms in the future. Some of the key requirements, specifically to suit network forensics traceback, include the following:

- compatibility with existing network protocols, routers and infrastructure
- simple and minimal number of functions to be implemented on transit routers
- support for incremental implementation, partial deployment and scalability
- minimal overhead of time and resources (processing, bandwidth, memory)
- fast convergence of the traceback process involving a few packets
- minimal involvement of the internet service provider (ISP) in the process

3 IP Traceback

IP Traceback [4, 5] problem is defined as “identifying the actual source of any packet sent across the Internet”. The traceback mechanism is shown in Fig. 1. We consider proactive measures applicable only for network forensics like logging, packet marking, hybrid approaches and AS-level traceback techniques.

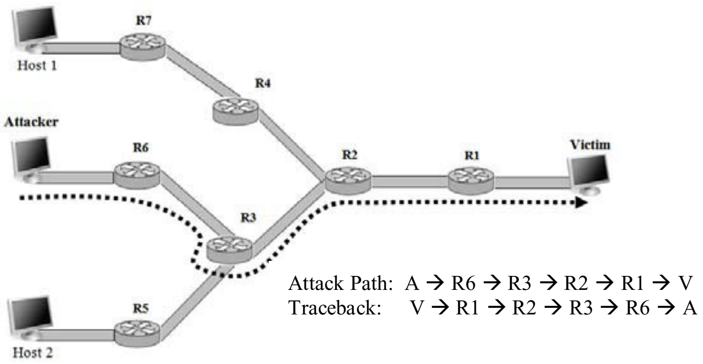
R7
Host 1
R4
Attacker
R6
R2
R1
Victim
R3
R5
Host 2
Attack Path: A → R6 → R3 → R2 → R1 → V
Traceback: V → R1 → R2 → R3 → R6 → A

Fig. 1. IP Traceback

3.1 Background

The true origin of an attacker can be identified by *logging* packets at key routers and later mining them for attack path reconstruction. Snoeren et al. [6] proposed *source path isolation engine* (SPIE) capable of tracing a single IP packet. A hash of multiple fields in the IP packet header is computed and logged in the digest tables. Baba and Matsuda [7] propose an *autonomous management network* (AMN), where monitoring manager receives requests from sensors which detect attacks. It queries the tracers maintaining log information about incoming packets to traceback.

Packet-marking involves placing the routers' part or complete address into the IP packet randomly with a fixed probability or only once deterministically. Savage et al. [8] proposed *probabilistic packet marking* (PPM) where each router probabilistically marks the Identification field in the IP packets (one in 25) with partial address information. Song and Perrig [9] proposed *advanced and authenticated packet marking* (AAPM) to reduce the storage space requirements by encoding the IP address into an 8 bit hash value used message authentication codes (MAC) to prevent packet content tampering. Dean et al. [10] proposed *algebraic packet marking* (APM) that employs algebraic techniques from the field of coding theory to calculate the values of 15-bit marks as points on polynomials. Yaar et al. [11] proposed *fast internet traceback* (FIT) that has three elements, a fragment of the hash of the marking router's IP address, the number of the hash fragment marked in the packet, and a distance field. Belenky and Ansari [12] proposed *deterministic packet marking* (DPM) where only the ingress edge routers mark the packets and all other routers are exempt from marking. Rayanchu and Barua [13] propose a *deterministic edge router marking* (DERM) where the 16-bit hash of the IP address is used to mark each packet.

Hybrid traceback approaches integrate packet marking and packet logging in a novel fashion to achieve the advantages of both the techniques. Duwairi and Govindarasu [14] propose *distributed link list traceback* (DLLT) where a router decides to mark the packet, stores the current IP address along with the packet ID in the marking table maintained at the router, and then marks the packet by its own IP address, and forwards the packet. Jing et al. [15] propose *hierarchical IP traceback system* (HITS) with three components for marking, evidence collection and traceback processing.

Gong and Sarac [16] develop *hybrid single packet IP traceback* (HIT) based on marking and logging. Traceback enabled routers audit traffic and a traceback server having the network topology information constructs attack graph by querying routers. Jing and Lin [17] propose *logging and deterministic packet marking* (LDPM) which traces the special edge connecting ingress and border routers.

Autonomous System can be a group of networks regulated by one or more entity, which enforces a clearly defined routing policy. An AS number (ASN) is a 16-bit integer, assigned and managed by IANA. Paruchuri et al. [18] propose *authenticated autonomous system traceback* (AAST) which probabilistically mark packets with AS number at AS Border Routers using 19 bits for ASN and the distance field. Gao and Ansari [19] propose *autonomous system based edge marking* (ASEM) in which only the ingress edge routers of each AS, mark packets with ASN according to certain probability. Packets are not remarked by all other routers. Korkmaz et al. [20] propose *AS-level single packet traceback* (AS-SPT) which logs packet digests at the border routers of participating ASes and traces toward packet origin at the AS-level. Castelucio et al. [21] propose an AS-level overlay network that operates on the border routers of an AS and builds an overlay network after exchanging BGP information.

3.2 Related Work

Focus of the IP Traceback approaches were in mitigating DDoS attacks by identifying the attack traffic and restrict it from reaching the victim. Relation between the major traceback mechanisms is shown in Fig. 2 to identify the suitable model for forensics.

Few researchers have identified the need to perform network traceback for other attacks. Carrier and Shields [22] propose the Session TOken Protocol (STOP) based on the Identification protocol (IDENT) and is aimed to automatically trace attackers logging through a series of stepping stones. Demir et al. [23] propose two approaches, session based packet logging (SBL) and SYN based packet marking (SYNPM), for traceback by providing simple and effective logging. Cohen [24] explores the problem of determining the real source behind the NAT gateway.

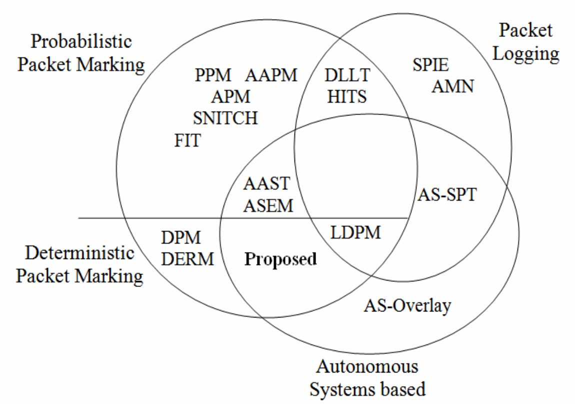

Fig. 2. Relation between various traceback techniques

4 Proposed Model

We propose an IP traceback model for network forensics based on the assumptions and requirements as listed earlier. The architecture is shown in Fig. 3. Our technique is based on deterministic packet marking (DPM) using an Autonomous System (AS) approach. We use a two level traceback mechanism, where the first level involves deterministic marking of each packet by the first ingress edge router within the AS and the second level involves marking each packet by the AS edge router (ASER). In both the levels, once the packet is marked, it cannot be marked by any other router. Every outbound packet is marked and inbound packets are not marked. A single packet is sufficient to detect the source as each contains the information about the AS and the edge router which first marked the packet.

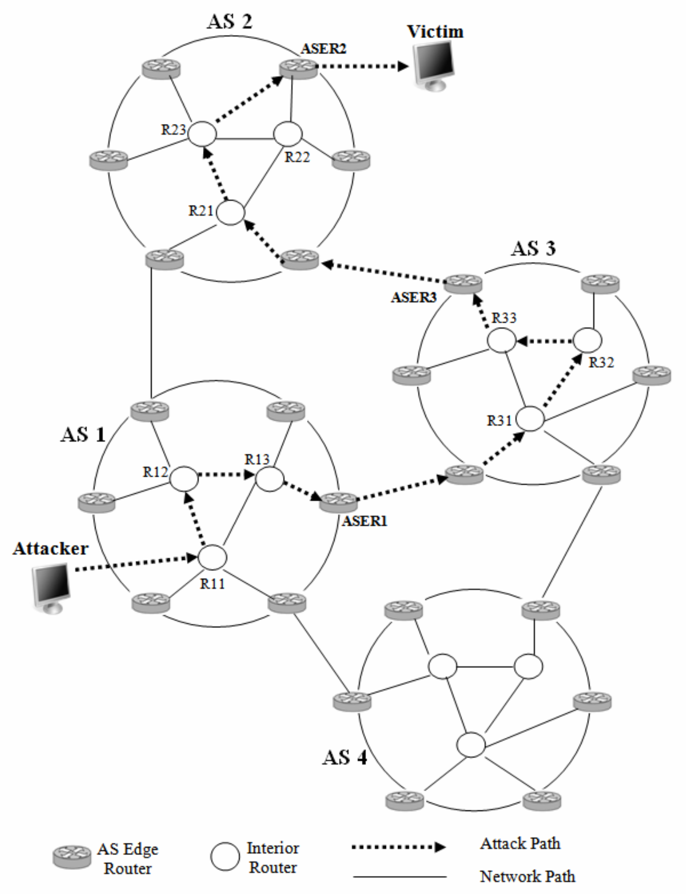

Fig. 3. Proposed Architecture for IP Traceback

4.1 Mark Information Encoding

We propose to use the 16-bit ID field, 3-bit fragment flag field and 13-bit fragment offset field in the IP header to store the marking information. These 32 bits were designed to hold information about fragmentation. The fragmentation traffic is very rare in Internet these days (about 0.25% of all traffic) [16]. The mapping between fields in the IP header and the marking fields is shown in Figure 4.

16 bits of ID field are used to hold the AS number (ASN), a globally unique number used to identify an AS. We use the ASN for marking rather than IP addresses as it is easy to mark it in the available 16 bits and will result in less number of false positives. This type of encoding was also done in [27]. After the marking is done, the reserved flag bit (first bit of the flag field immediately after the ID field) is set to 1.

The next 16 bits following the ID field are made up of 3-bit flag and 13-bit offset field. 12 most significant bits of the 13-bit offset is used to store the hashed IP address of the first ingress router traversed by the packet [18]. The remaining least significant bit is used as a flag to indicate that the marking has taken place. We may also use all the 16 bits also to store a 16-bit hash value of the ingress router address [13].

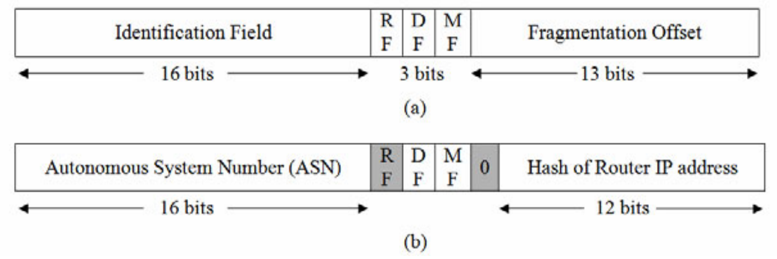

Fig. 4. Marking encoding (a) fields in the IP header (b) fields overloaded for marking

4.2 Marking Operation

We have proposed a two-level marking mechanism. The algorithms are as follows:

```
program MarkIngressEdgeRouter ($R_i$)
        for each outbound packet P
            if P.offset[0] == '1' then
                forward (P)
            else
                set P.offset[0] == '1'
                write HashIP($R_i$) into P.offset [1..12]
            end if
        end for

program MarkASEdgeRouter ($ASER_i$)
    for each outbound packet P
            if P.flag [0] == '1' then
                forward (P)
            else
                set P.flag [0] == '1'
                write ASN ($ASER_i$) into P.Identification
                only if P is forwarded to another ASER
            end if
        end for
```

4.3 Traceback Operation

Traceback operation is simple as each packet holds the information required to identify the AS and the first ingress router. The 16-bit identification field in the IP header gives the ASN, identifying the source AS of the packet. The 12-bit hash value in the offset field is used to extract the ingress router IP address using the hash function**.** This information can be extracted even from a single attack packet.

4.4 Analysis

A *single packet* can give the information till the ingress router to which the attacker was connected. There is *no additional storage* required, neither at the router nor at the victim. None of the packets are logged and nor information about the packets is stored. The *processing overhead is nominal* as the marking operations are simple functions, which can be easily performed. The hashing of router IP addresses can be calculated in advance, stored and be made available when required. The processing overhead may increase when full bandwidth traffic has to be marked. There is very *less infrastructure change* which needs to be made, as marking is done only twice. The technique is *scalable* and can handle many attackers as each attackers information can be given by a single packet. Number of *false positives is less* as the entire information is coded into a single packet. The hashing functions may have collisions yielding a few false positives. There is a *considerable amount of ISP involvement* needed as an AS can be thought of as an ISP. It must make the ASN available for routers to be enabled for traceback. *Incremental deployment is limited* as the marking is done only twice and if some of the routers are not enabled, the technique may yield more false positives.

5 Conclusion

The most challenging problem for network forensics, IP traceback, was examined in this paper. A traceback model based on Autonomous System and Deterministic Packet Marking was proposed. The proposed technique can trace the attacker till the ingress edge router even with a single packet which meets the basic requirement of network forensics. It requires nominal processing and there is no storage overhead. The only drawback is the higher involvement of ISP operating the AS. Future work involves performance analysis using simulations to validate our technique in comparison with the existing traceback techniques. Accommodating fragmentation of packets, while using the 32 bits used for fragmentation to mark packets, is also a challenge.

References

[1] Lee, S.C., Shields, C.: Tracing the Source of Network Attack: A Technical, Legal and Societal Problem. In: IEEE Workshop IAS, New York, pp. 239–246 (2001)

[2] Palmer, G.: A Road Map for Digital Forensic Research. In: Proc. 1st Digital Forensic Research Workshop (DFRWS), pp. 27–30 (2001)

[3] Pilli, E.S., Joshi, R.C., Niyogi, R.: Network forensic frameworks: Survey and research challenges. Digit. Investig, available online March (2010) (in press)

[4] Gao, Z., Ansari, N.: Tracing Cyber Attacks from the Practical Perspective. IEEE Communications Magazine 43(5), 123–131 (2005)

[5] Santhanam, L., Kumar, A., Agrawal, D.P.: Taxonomy of IP Traceback. J. Info. Assurance and Security 1, 79–94 (2006)

[6] Snoeren, A.C., Partridge, C., Sanchez, L.A., Jones, C.E., Tchakoutio, F., Kent, S.T., Strayer, S.T.: Hash-Based IP Traceback. In: Proceedings of ACM SIGCOMM (2001)

[7] Baba, T., Matsuda, S.: Tracing Network Attacks to Their Sources. IEEE Internet Computing, 20–26 (March/April 2002)

[8] Savage, S., Wetherall, D., Karlin, A., Anderson, T.: Network Support for IP Traceback. IEEE/ACM Transactions on Networking 9(3), 226–237 (2001)

[9] Song, D., Perrig, A.: Advanced and Authenticated Marking Schemes for IP Traceback. In: Proceedings of the IEEE INFOCOM 2001, Arkansas, USA (2001)

[10] Dean, D., Franklin, M., Stubblefield, A.: An Algebraic Approach to IP Traceback. ACM Transactions on Information and System Security 5, 119–137 (2002)

[11] Yaar, A., Perrig, A., Song, D.: FIT: Fast Internet Traceback. In: Proc. IEEE 24th Ann. Joint Conf. Computer and Comm. Societies (INFOCOMM 2005), pp. 1395–1407 (2005)

[12] Belenky, A., Ansari, N.: On Deterministic Packet Marking. Computer Networks 51, 732–750 (2006)

[13] Rayanchu, S.K., Barua, G.: Tracing Attackers with Deterministic Edge Router Marking (DERM). In: Ghosh, R.K., Mohanty, H. (eds.) ICDCIT 2004. LNCS, vol. 3347, pp. 400–409. Springer, Heidelberg (2004)

[14] Duwairi, A., Manimaran, G.: Novel Hybrid Schemes Employing Packet Marking and Logging for IP Traceback. IEEE Trans. Parallel and Dist. Sys. 17(5), 403–418 (2006)

[15] Jing, Y.N., Tu, P., Wang, X.P., Zhang, G.D.: Distributed log based scheme. In: Proc of 5th Int'l. Conf. on Computer and Information Technology (2005)

[16] Gong, C., Sarac, K.: A More Practical Approach for Single-Packet IP Traceback using Packet Marking and Logging. IEEE Trans. Parallel and Dist. Sys. 19(10), 1310–1324 (2008)

[17] Jing, W.X., Lin, X.Y.: IP Traceback based on Deterministic Packet Marking and Logging. In: Proc. IEEE Int'l. Conf. on Scalable Computing and Comm., pp. 178–182 (2009)

[18] Paruchuri, V., Durresi, A., Kannan, R., Iyengar, S.S.: Authentic Autonomous Traceback. In: Proc. 18th Int'l Conf. Adv. Info. Networking and Appln., pp. 406–413 (2004)

[19] Gao, Z., Ansari, N.: A practical and robust inter-domain marking scheme for IP traceback. Computer Networks 51(3), 732–750 (2007)

[20] Korkmaz, T., et al.: Single packet IP traceback in AS-level partial deployment scenario. Int. J. Security and Networks 2(1/2), 95–108 (2007)

[21] Castelucio, A., Ziviani, A., Salles, R.M.: An AS-level Overlay Network for IP Traceback. IEEE Network, 36–41 (2009)

[22] Carrier, B., Shields, C.: The Session Token Protocol for Forensics and Traceback. ACM Trans. on Info. System Security 7(3), 333–362 (2004)

[23] Demir, O., Ping, J., Kim, J.: Session Based Packet Marking and Auditing for Network Forensics. Int'l. Journal of Digital Evidence 6(1), 1–15 (2007)

[24] Cohen, M.I.: Source attribution for network address translated forensic captures. Digit. Investig. 5(3-4), 138–145 (2009)