A Novel Attack Path Reconstruction Based on Packet Logging & Marking Scheme

S.Prathyusha¹, M.V.Sruthi², S.Anjani Prasad³
M.Tech Student, Dept of CSE, Gates Institute of technology, Gooty, India¹
Assistant Professor, Dept of ECE, PVKK college of Engineering, Anantapur, India²
Assistant Professor, Dept of CSE, Gates Institute of technology, Gooty, India³

Abstract: Computer network attacks are on the increase and are more sophisticated in today’s network environment than ever before. One step in tackling the increasing spate of attacks is the availability of a system that can trace attack packets back to their original sources irrespective of invalid or manipulated source addresses. Most of these schemes require very large number of packets to conduct the traceback process, which results in lengthy and complicated procedure. IP traceback is just the technique to realize the goal, it reconstructs IP packets traversed path in the Internet to determine their origins. To reconstruct the path of a packet and identify the source of the attack, the victim requires a map of the routers. The victim matches packet markings with the routers on the map and can thus reconstruct the attack path. There are two major kinds of IP traceback techniques, which have been proposed as packet marking and packet logging. Here in this paper we presented a novel attack path reconstruction based on packet logging and marking techniques which shows improved accuracy, practicality and low storage and number of routers.

Keywords: IP traceback, packet logging, packet marking

I. INTRODUCTION

During the past decade, a lot of attention has been focused on the security of Internet infrastructure in place as being part of transmission, reception and storage of paramount importance within the increasing ecommerce applications. Yet, with high-profile Distributed Denial-of Service (DDOS) attacks, numerous ways have been elaborated to identify the source of these attacks and one methodological approach is using IP traceback. The goal of IP traceback is to trace the path of an IP packet to its origin. The most important usage of IP traceback is to deal with certain denial-of-service (DoS) attacks, where the source IP address is spoofed by attackers. Identifying the sources of attack packets is a significant step in making attackers accountable. In addition, figuring out the network path which the attack traffic follows can improve the efficacy of defense measures such as packet filtering as they can be applied further from the victim and closer to the source. Two main kinds of IP traceback techniques have been proposed in two orthogonal dimensions: packet marking [1] and packet logging [2]. In packet marking, the router marks forwarded IP packets with its identification information. Because of the limited space in packet header, routers probabilistically decide to mark packets so that each marked packet carries only partial path information. The network path can be reconstructed by combining a modest number of packets containing mark. This approach is known as probabilistic packet marking (PPM) [3]. The PPM approach incurs little overhead at routers. But it can only trace the traffic composed of a number of packets because of its probabilistic nature.

In packet logging, the IP packet is logged at each router through which it passes. Historically, packet logging was thought to be impractical because of enormous storage space for packet logs. Hash-based IP traceback approach [4] records packet digests in a space-efficient data structure, bloom filter [5], to reduce the storage overhead significantly. Routers are queried in order to reconstruct the network path. the information required to achieve traceback is either stored at different points (mostly on routers) along the path that a packet traverses or that path and usually other incidental paths are analyzed to gain information that will be used in traceback. It is this distinction that we employ to further divide network based schemes into packet logging schemes and network analysis schemes.
1.1 Packet Logging Schemes:

In packet logging schemes, traceback is accomplished by requiring nodes that a packet traverses to log information about that packet in such a format that can later be queried to discover if the packet has been seen at that node. The big drawback here is the huge amount of storage that may be needed at each node especially at nodes connected to high speed links that witness a lot of traffic. One of the most definitive works in IP Traceback is the Hash-based IP Traceback scheme [3] which computes and stores a Bloom filter digest for every packet at each node. The scheme comprises a Source Path Isolation Engine (SPIE) which instead of storing whole packets, saves space by storing only computed hashes for each packet based on the invariant parts of the IP header along with the first 8 bytes of the payload. The use of a Bloom filter was innovative in addressing the storage space problems of packet logging schemes but despite this, the requirements in terms of computation and space are still formidable. As noted in [4] “assuming a packet size of 1000 bits, a duplex OC-192 link requires 60 million hash operations to be performed every second, resulting in the use of SRAM (50ns DRAM is too slow for this) and 44GB of storage per hour, with the parameters suggested in” [3]. To resolve this, [4] went ahead to propose a packet logging based traceback scheme that is scalable to such high link speeds by sampling and logging only a small percentage of packets and then using more sophisticated techniques to achieve traceback. We note however that the approach taken by [4] loses the single packet traceback ability originally possessed by [3].

1.2 Packet Marking Schemes

Distinct from the ICMP messaging scheme, packet marking schemes encode information about the path a packet traverses in the packet itself, usually in rarely-used fields within the IP header. Apart from the well-known issue of finding enough space in the current IP header in which to place traceback information, another common problem with packet marking schemes arises from the additional computational tasks placed on routers during the marking process. There are two common methods to achieve this

1.2.1 Deterministic Packet Marking Schemes

Deterministic packet marking involves the marking of each individual packet with a node’s marking information at one or more nodes reached by the packet in transit in such a way that traceback by can realized by considering the marks on each packet. These marks are usually placed in reserved spaces in the IP header (most proposed solutions use the Identification field).

The overhead associated with this method are mostly in terms of the packet space and modification required in the IP header, as well as the router processing time required. Various schemes differ in the type of markings they use, as well as in the points where a node is required to mark. For example in [10], the marking code is a half of the IP address octets (i.e. 16
bits) plus one extra bit to indicate which half and a 17 bit storage on the packet, comprised of the 16-bit ID field and the 1-bit reserved Flag bit is used to hold this information, while in [21] an encoding scheme based on router-level topology is proposed to be stored in a futuristic IPv6 packet header that makes provision for 52-bits for traceback purposes.

1.2.2 Probabilistic Packet Marking
Probabilistic schemes counter the space constraint in deterministic packet marking by requiring that a packet is marked with only a part rather than the whole of the path it traverses. Such schemes make the reasonable but not foolproof assumption that attacks are usually made up of a large number of packets, and so by aggregating these partial path information from a number of packets, traceback can be achieved along with path reconstruction.

The two types have their own features: Probabilistic Packet Marking incurs little overhead when routers mark packets in a low marking rate, but the victim needs a large number of packets to reconstruct the path to the source. It is more suitable for flooding DoS traceback, and does not have the capability to trace a single packet. While SPIE extracts the digests of packets and stores them in a space-efficient data structures known as bloom filter [15], which decreases the storage overhead and makes the packet logging scheme practical. It can trace small packets flows, even a single packet. However, it is still a challenging task for its practicality due to its remaining high storage overhead. Therefore, it is attractive to propose an effective IP traceback mechanism with the combination of the two traceback techniques, which is called a hybrid IP Traceback scheme. At present, HIT (Hybrid Internet Traceback) proposed by Gong Chao is the most representative. HIT borrows the main idea of packet logging, and records packet digests in every other router. The marking routers do not record digests, but write their ID information into some certain fields of IP header. It is efficient to reduce the huge storage overhead of SPIE. However, there are some drawbacks of HIT. Firstly, it may return incorrect path even the false source; then it still has a great demand for storage, which would limits its practicality.

In this paper we have presented a unique approach for attack path reconstruction based on packet marking and packet logging. The contribution of our approach is (1) to reduce the storage overhead at routers to roughly one half, and (2) to reduce the access time requirement for recording packets by a factor of the number of neighbor routers. For each arriving packet, routers always commit marking operation, but commit logging operation when needed (generally alternately). The packet digest is stored in the same fashion as the hash-based approach. But the mark is stored in a space-efficient fashion so that the storage requirement for marks is negligible. Thus the storage overhead at routers is reduced to one half. Each router maintains a different digest table for each of its neighbor routers. Packets coming from different neighbor routers (with different marks) can be recorded in corresponding digest tables simultaneously. That reduces the access time requirement by a factor of the number of neighbor routers.

The rest of this paper is organized as follows. Section II puts our approach in the context of the related work. Section III describes our IP traceback approach in detail. Section IV analyzes the resource requirements and performance of our approach.

II. BACKGROUND AND RELATED WORK

2.1 Background: Denial of Service (DoS) is a threatening intrusion in the Internet now. According to the size of attack packet flow, it can be classified to two groups. One type is to consume the resource of the victim with a huge number of meaningless packets, which can be named as resource consuming DoS. It is the major type, and can make the victim resource exhaustive in a very short time. The other type is to make the victim cannot provide service with the vulnerability of the software or service running on the victim, which named software exploit DoS. The key feature of it is the smaller packet flows, even a single packet. It has become an important part of DoS in recent years. What we should do is try our best to defend the serious cyber attack. Unfortunately, IP network is designed for freedom and resource share without much consideration of security issues, and anonymous access and non-state are two key characteristics. It means no guarantee for the authenticity of the source address, and no record about the transmission path of packets. In addition, attacker may insert an arbitrary address into the source address field of a packet, which is known as IP spoof, and IP spoof makes it more difficult to defend DoS intrusion. Therefore, it is an extraordinary challenge for us to trace back to the source of DoS. IP Traceback technique is studied to resolve the problem, and it can be defined that equipments in IP network record packets state information in the network with a certain mechanism, reconstruct the complete path and find the source reliably in the end.

2.2 Related work: Packet logging scheme makes routers record the state information of packets, and it has the capability to trace a single packet. Therefore it can provide the straightforward evidence for traceback. It becomes practical when SPIE
appeared, but it still needs great storage space. Packet marking approach makes routers write ID information into packet IP header, and reconstruct the complete path in the victim node. It does not increase storage burden on routers. Several researchers combine the features of the two approaches to propose hybrid IP traceback approaches [16-18]. The approaches are outstanding for their low storage overhead, and single packet traceback capability. In hybrid IP traceback, each traceback-enabled router will conduct logging or marking. However, hybrid IP traceback approaches should not be the simple combination of the two methods, and we must pay much attention to some key issues. Firstly, in normal condition, a marking router has no more than one neighboring logging router, which logged the attack packet. However, in some certain situations, packets will follow some special routes, and a marking router may have more neighboring logging routers. Hybrid IP traceback may return false paths, which could lead to the failure of traceback. Secondly, in the marking process, marking routers insert ID into IP header without logging, which can reduce storage overhead. The marking space in IP header is limited, and full utilization of the space can hold more ID information, which can reduce the percentage of logging routers in all traceback routers. In this way, the IP traceback approach is more practical.

The IPv6 implementation of this traceback scheme needs more research because IPv6 has built-in security mechanisms such as authentication headers to provide origin authentication. Therefore, we leave it as our future work. This message can show the state information instead of logging in the router, which can reduce the storage overhead. Like packet marking approaches, this make use of some fields in the IP header when conduct marking. The utilization of IP header for marking. Three fields in IP header are used for marking: they are Identification, Reserved flag and fragment offset. Similar to other PPM, this traceback reuses the Identification field, which is for IP fragmentation. Measurement studies show that less than 0.25% of packets are fragmented in the Internet [19]. Savage et al. [8] had some discussion about the backward compatibility issues of utilization of this field and confirmed its utilization. This method reuses Reserved Flag (RF) field for marking as suggested in [20]. The fragment offset field will become meaningless if the IP Identification filed is reused for other purposes. Therefore, some traceback approaches utilize this field [21, 22], PPIT follows this way. Someone may argue that the value in the fragment offset field could cause some compatible problems. In order to avoid such issues in PPIT, the marking information will be removed before the packet arrives at the destination. In a certain area, IP address is too long to be taken in the IP header marking space and is unnecessary. In PPIT, each marking router is assigned a 14-bit ID number, which is to differentiate it to other neighboring routers. Muthuprasanna et al. [23] studied the unique ID number assignment problem for Internet routers and proposed an approach for internet coloring. They report that 14 bits are enough for a unique ID number assignment within a three-hop neighborhood and 12 bits for two-hop neighborhood, based on the analysis of several Internet topology data sets. Furthermore, the same ID number can be assigned to routers more than one in different three-hop neighborhood. If these neighborhoods do not have a common router. Therefore, we can use such short ID information to replace IP address. PPIT makes use of the Identification field to store routers ID in 14 bits, which is called router ID field. The ID is generated for each traceback-enabled router, which is set by the network administrators. For each arriving packet, the current router first examines the router ID number marked in the packet header to check whether it is valid. The router ID number carried by a packet p is valid at a router r if it equals to the ID number of some neighbor router of router r. That is, the packet p is forwarded from some neighbor router to router r. If the router ID number is valid, based on the logging flag bit in the packet, the router may choose to commit (1) only marking operation, or (2) both marking and logging operations. If the upstream router logged the packet (logging flag is 1), the current router chooses to only mark the packet; if the upstream router didn’t log the packet (logging flag is 0), the current router chooses to both mark and log the packet. If the router ID number is not valid, that means the arriving packet came directly from the sender host or an attacker which sends packets with forged mark. In this case, the router chooses to commit only marking operation.
The effectiveness of IP traceback increases greatly with widespread deployment of traceback-enabled routers in the network. However, it is likely that hybrid IP traceback approach does not require all routers to be traceback-enabled. All traceback-enabled routers form an overlay network. If the traceback server has the topology knowledge of that overlay network and each traceback-enabled router knows its neighboring traceback-enabled routers, this uniquepath reconstruction approach still works.

### III ATTACK PATH RECONSTRUCTION

To reconstruct the path of a packet and identify the source of the attack, the victim requires a map of the routers. The victim matches packet markings with the routers on the map and can thus reconstruct the attack path. Obtaining or constructing this map is not difficult. A number of tools are available that can be used to obtain a map of the the routers and the Internet. If a router commits logging operation on an attack packet, examining digest tables at that router will not only confirm that router is in the attack path, but also find out its upstream router in the attack path since each digest table is annotated with an upstream router’s ID number. Given an attack packet and victim, the traceback server could infer the last hop router and whether the last hop router committed logging operation based on the logging flag bit carried by the attack packet.

- If the traceback server infers a router logged the attack packet, examining the digest tables at that router would identify its upstream router in the attack path.
- If the traceback server infers a router didn’t log but marked the attack packet, querying the neighbor routers of that router in the RPF manner and examining the digest tables on these neighbor routers would identify the upstream router. The attack graph can be constructed using those two methods alternately. Figure 3.1 shows how to construct attack graph.

**FIG3.1:** Attack path construction. Solid arrows represent the attack path; dashed curves represent the first method; dashed arrow represent the second method. Router D and H logged the attack packet.

- If the packet undergoes transformation at the current router, commit both marking and logging operations on the packet, and record the transformation information in the transform lookup table. Given a packet, consulting the transform lookup table.
table can get to know whether the packet was transformed and the original packet can be reconstructed. The implementation of the transform lookup table is described in [4].

- If the packet is a fragmented packet, compute and store the packet digest in a particular digest table which is only for fragmented packets and is managed in the same way as the hash-based approach.
- Otherwise, follow the algorithm in packet operating procedure. Attack graph construction is also improved accordingly.

When the traceback server examines the digest tables at a router, it also consults the transform lookup table at that router and reconstruct the attack packet to its original form if possible.

1) If the attack packet provided by victim is not a fragmented packet, the traceback server employs the procedure to construct attack graph which is similar to the one presented in Section III-B. The only difference is that it is not any longer true that routers in an attack path log an attack packet alternately. It is possible that two adjacent routers, say m and n, both log a same packet p because the upstream router m commits logging operation on packet p, and p undergoes transformation at the downstream router n. During traceback process, when moving to the upstream router m from router n which logged and transformed packet p, the traceback server can not assume router m did not log packet p. The traceback server needs to examine the digest tables at m to figure out whether m marked p only or both marked and logged p, then takes proper action accordingly.

2) If the attack packet is a fragmented packet, starting from the last hop router, the traceback server queries routers in the RPF manner and examines the digest tables recording digests for fragmented packets to construct the attack path.

### 3.1 Overheads on Routers

In the hybrid approach, when a packet is traversing the network, each router on the route commits marking operation on the packet, every other router commits logging operation. Keeping different digest table for each neighbor router makes the storage overhead for router ID numbers negligible. So the overall storage overhead at routers is reduced to roughly one half. In addition, since the router keeps separate table for each neighbor, packets coming from different neighbor routers can be recorded in corresponding digest tables simultaneously as long as each digest table has its own read/write hardware support. Thereby the access time requirement for recording packet digests is reduced by a factor of the number of neighbor routers. When taking into account packet transformation and backwards compatibility, the hybrid approach can handle those two issues in return for a modest increase in storage and access time requirements. But the percentage of IP traffic undergoing transformation and the percentage of IP fragmented traffic are small (3% [10] and 0.25% [11]), hence the increases of storage and access time requirements should be trivial.

Fig3.1: detailed procedure of path reconstruction
3.2 Traceback Process:

During the traceback process, the total number of the digest tables examined is an index reflecting the overhead on the traceback server and the speed of the traceback process. Suppose time synchronization is maintained between adjacent routers, and each router has \( n \) neighbors on average. Then, during the traceback process, the ratio of the number of digest tables examined in the hybrid approach to that in the hash-based approach is between \( n/2 \) and \( 1/2 \), depending on the average link latency between routers. The mathematical deduction below is based on average values of parameters and omits small value constants. Suppose each router has \( n \) (\( n \geq 2 \)) neighbor routers on average, and the traffic load at the router is from each neighbors equally. Let the average time interval covered by one digest table in the hash-base approach be \( t_h \), and the average time interval covered by one digest table in the hybrid approach be \( t_c \). Then

\[
t_c = t_h \times n.
\]

(1)

Suppose the attack path is \( m \) hops long from the attacker to the victim. Let the average link latency between routers be \( l \). If the average link latency between routers is larger than the average time interval covered by one digest table, multiple digest tables covering continuous time periods at one router or one interface will be examined during the traceback process. Suppose the average time interval covered by one digest table is \( t \), then \( n \) tables need to be examined in order to locate the digest of attack packet. In the hash-based approach, in order to move one hop upstream along the attack path from the current router during the traceback process, the digest tables at \( n \) neighbor routers need to be examined (actually \( n-1 \), we omit that constant for simplicity). The number of digest tables examined is

\[
N_h = m \times n \times \left\lceil \frac{t}{t_h} \right\rceil = m \times n \times \left\lfloor \frac{l \times n}{t_c} \right\rfloor.
\]

(2)

In the hybrid approach, in order to move from the current router which marked attack packet to the upstream marking router which is 2 hops away, the digest tables at all interface of \( n \) neighbor routers need to be examined, there are \( n^2 \) interfaces totally (actually \( n-1 \)). The number of digest tables examined is

\[
N_c = \frac{n}{2} \times n^2 \times \left\lceil \frac{l}{t_c} \right\rceil.
\]

(3)

We omit the odd case for simplicity.) Hence the ratio of the number of digest tables examined in the hybrid approach to that in the hash-based approach during the traceback process is

\[
r = \frac{N_c}{N_h} = \frac{n}{2} \times \frac{\left\lfloor \frac{l}{t_c} \right\rfloor}{\left\lceil \frac{l}{t_h} \right\rceil}.
\]

(4)

When \( l \leq t_h \), we have \( \left\lceil \frac{l}{t_c} \right\rceil = \left\lceil \frac{l}{t_h} \right\rceil = 1 \). Then, \( r = \frac{n}{2} \).

When \( l = a \times t_c \) (\( a = 1, 2, 3 \ldots \)), we have \( \left\lceil \frac{l}{t_c} \right\rceil = a \) and \( \left\lfloor \frac{l}{t_h} \right\rfloor = a \times n \). Then, \( r = \frac{n}{2} \times \frac{1}{n} = \frac{1}{2} \).

When \( l = a \times t_c + r \) (\( a = 1, 2, 3 \ldots \), \( 0 < r < t_c \)), we have \( \left\lceil \frac{l}{t_c} \right\rceil = a + 1 \) and \( \left\lfloor \frac{l}{t_h} \right\rfloor = \left\lfloor \frac{l \times n}{t_c} \right\rfloor = a \times n + \left\lceil \frac{r \times n}{t_c} \right\rceil \).

Because \( 1 \leq \left\lceil \frac{r \times n}{t_c} \right\rceil \leq n \), so \( a \times n + 1 \leq \left\lceil \frac{l}{t_h} \right\rceil \leq (a+1) \times n \).

Then,

\[
1 = \frac{a+1}{(a+1) \times n} \leq \frac{\left\lceil \frac{l}{t_h} \right\rceil}{\left\lceil \frac{l}{t_c} \right\rceil} \leq \frac{a+1}{a \times n + 1} < \frac{1}{n} + \frac{1}{a \times n + 1} < \frac{2}{n}.
\]

So \( \frac{1}{2} \leq r < 1 \).

In summary,

\[
\frac{1}{2} \leq r \leq \frac{n}{2}.
\]

(5)
According to the deduction above, when \( l_{tc} \), \( r < 1 \). That is, when the link latency between routers is large enough, less digest tables are examined in hybrid IP traceback approach than hash-based IP traceback during the traceback process. If time synchronization is not maintained between adjacent routers, more digest tables need to be examined at each router. That is equivalent to the prior case with an increased link latency between routers. We could get a similar conclusion.

IV CONCLUSION

In this paper, we proposed a new attack path reconstruction approach which is based on both packet marking and packet logging. Our approach has the ability to track packet back to its origin. Compared to hash-based IP traceback approach, it reduces the storage overhead to roughly one half and improves accuracy on the access time by a factor of the number of neighbor routers.

V ACKNOWLEDGMENT

I thank Mrs M.V.Sruthi, S.Anjani Prasad for their comments and valuable suggestions on this manuscript.

REFERENCES


Biography

S. Prathyusha B.Tech is pursuing her M.tech in Gates Engineering college, Gooty, India.

Mrs. M. V. Sruthi, M.Tech (Ph.D) is an Research scholar and Assistant Professor in the department of Electronics and communications engineering, P.V.K.I.T. Anantapur, India.

S. Anjani Prasad is working as an Assistant Professor in the Dept of CSE, Gates Institute of technology, Gooty, India.