Chapter 5

Packet Marking Based Anti Spoofing Techniques

In the previous chapter, I discussed how bidirectional nature of Internet communications can be used to detect and mitigate a DDoS attack. However, spoofing allows an attacker to degrade the performance of the defense system. In this chapter, I discuss about anti spoofing defense systems which can discriminate spoofed attack packets from benign packets based on an identification mark carried by the packets. The performance of such defense systems depend on an underlying packet marking scheme which inscribes the identification mark into the forwarding packets. In this chapter, I propose and demonstrate two packet marking schemes namely XORID and SEM. In XORID one or more intermediate routers participate in the packet marking process to encode the path traveled by the forwarding packets as their identification mark. XORID is proposed to overcome shortcomings associated with existing similar techniques. On the other hand, SEM explores the benefits of marking the packets deterministically at the source end to help defending against spoofing attacks.

5.1 Introduction

In IP spoofing, an attacker sends attack packet to the victim by replacing the source IP (SIP) of the attack packet with any IP address other than its actual one. Internet does not provide any mechanism which enables the victim to verify the genuineness of the SIPs of the received packets. This limitation allows an attacker to send attack traffic to the victim by replacing the SIP addresses of the attack packets dynamically. Thus, at the victim side it becomes difficult to discriminate an attack packet from a normal packet based on its SIP address. Even if the victim maintains a list of valid source IP addresses representing most frequently connected users as proposed in defense mechanism such as [30], such a list often becomes void if the attacker chooses the spoofed IP addresses from a set of potential users of the victim. It is not always the victim which only receives spoofed packets. In a DRDoS (Distributed Reflector Denial of Service) attack [7, 145], the attacker sends spoofed request packets to a set of public servers, where the spoofed SIP is the IP address of the victim. Thus, the public servers send the responses, which are usually many times larger than the request messages, to the victim. As a consequence the victim encounters a huge surge of response packets coming from all these servers. The servers which are commonly used for this purpose usually communicate over UDP where no handshaking is required before sending the responses, such as NTP servers, DNS servers and SMTP servers. Thus, IP spoofing plays an important role in executing a DRDoS attack. An anti-spoofing mechanism is capable of identifying the incoming spoofed request packets to such servers, which is helpful to reduce the intensity of DRDoS attacks. Other than DRDoS attack, IP spoofing is also used to send direct attack packets from a set of zombies to the victim by randomly setting the SIP field of the attack packets. Under such a situation, the victim cannot block the attack traffic based on their SIP addresses.

An anti-spoofing mechanism should be capable of discriminating the attack traffic from legitimate traffic irrespective of their SIP addresses. In this paper we discuss about source identification mark based anti-spoofing schemes which discriminate spoofed packets from legitimate packets based on an identification mark carried by the packets. Such techniques allow the victim to discriminate a spoofed packet from a legitimate packet even if both of them carry the same SIP. For example in Figure 5.1, the victim V receives two packets sent from A and B. However, B spoofs the SIP of the sent packet as the IP address of A. An identification mark based anti-spoofing technique allows V to discriminate the legitimate packets sent from A, from the spoofed packets sent from B, even if all the packets carry the same SIP.

The performance of such techniques depend on the underlining packet marking technique. Here, we propose a path encoding scheme, referred to as XORID, which enables a packet to carry a source identification mark in its 16 bit ID field. The source identification mark is derived from the path traversed by a packet from its source to destination. XORID is proposed with enhanced performance in comparison with several other similar techniques such as PI, StackPI and ANTID. The basic assumption in XORID and existing related techniques such as PI, StackPI and ANTID is that, all the packets reach a specific destination from a specific source by traversing the same sequences of routers, i.e., by following the same path. However, such an assumption is not always true. To address this issue, in this paper we also propose a source end packet marking system, referred to as Source End Marking (SEM). In SEM, the source identification marks are written into the outgoing packets by the first SEM enabled router from the source of the packet in the source network.

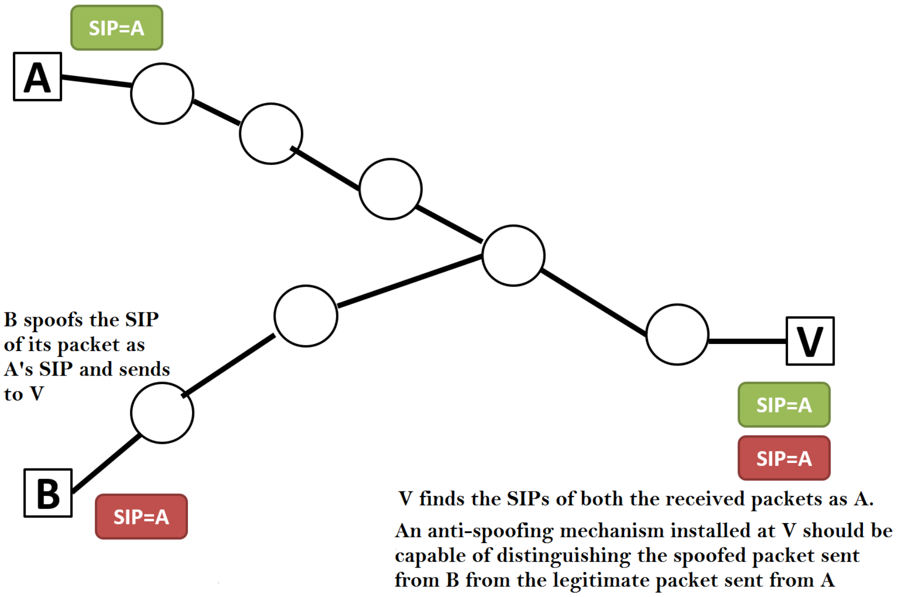

SIP=A
A
B spoofs the SIP of its packet as A's SIP and sends to V
B
SIP=A
V
SIP=A
SIP=A
V finds the SIPs of both the received packets as A.
An anti-spoofing mechanism installed at V should be capable of distinguishing the spoofed packet sent from B from the legitimate packet sent from A

Figure 5.1: Identification mark based anti-spoofing technique

The rest of the chapter is organized as follows: Section 5.2 discusses the motivation behind the proposed schemes. Section 5.3 discusses the proposed path encoding scheme, XORID. Section 5.4 discusses a general architecture of spoofing attack defense system based on source identification marks of the received packets facilitated by XORID (or any other similar techniques). In section 5.5 we discuss SEM, a source end packet marking scheme. Conclusions are drawn in section 5.6.

5.2 Motivation

The motivation behind the proposed techniques is to present a strong SIP verification technique to defend spoofing based DDoS attacks, which can overcome the following performance related issues of other competing techniques such as StackPI, PI, and ANTID.

1. Our proposed techniques are to improve the performance of existing similar techniques. For example, StackPI is vulnerable to attackers which are closer to the victim network. In StackPI all the sources which are i hops away from the victim network are distributed over k^i different codes, where k is the number of bits inserted by a single marking router. Table 5.1 shows the distribution of the codes among the sources located at different hops from the destination for both 1-bit and 2-bit StackPI schemes with respect to mwest monitoring point topology. The *Hop Count* column represents different path lengths, *Paths* column represents number of distinct paths of the corresponding length, *SP1 Codes* column represents different distinct codes assigned by StackPI 1-bit scheme to different paths of the corresponding length, *SP2 Codes* column represents different distinct codes assigned by StackPI 2-bit scheme to different paths of the corresponding length and *XORID Codes* column represents different distinct codes assigned by XORID scheme to different paths of the corresponding length. It has been observed that sources which are closer to the victim network are distributed over fewer numbers of different codes. It allows the attacker to place its zombies closer to the victim and mimic the sources which are at the same distance from the victim network with a very high probability of success. For example, by placing the zombies at a distance of 2 hops from the victim network an attacker can mimic all the 2 hops away genuine sources with a success probability of 1/16 (assuming k=2). StackPI is an enhanced marking scheme of PI, thus, PI also suffers from the above mentioned shortcomings.

2. Although ANTID does not suffer from these shortcomings however, a determined attacker can set the appropriate distance value against a spoof SIP in a transmitted attack packet[107]. Using such a technique the attacker can increase the probability of an attack packet reaching the destination with a valid mark from $\frac{1}{2^{16}}$ to $\frac{1}{2^{11}}$.

In this work we first propose a marking scheme referred to as XORID to address the above mentioned limitations of the existing marking schemes. XORID has the following benefits over the other methods

- XORID distributes all the 2^{16} possible marks with equal probability to all the potential paths, irrespective of the distance of the source to the victim. Hence, unlike StackPI, an attacker's attempt to mimic sources which are closer to the victim network will not succeed.

- XORID does not leave any scope for the attacker to specify any part of the mark with correct values. Hence, unlike ANTID, knowing the correct distance from a specific source to the victim does not do any good to the attacker.

All the above mentioned schemes PI, StackPI, and XORID encode the path traveled by the packets from a specific source to a destination as the identification mark of the source. The two main limitations of such an approach are given below

- All packets from a specific source to the destination might not always travel the same path due to legitimate network operations such as load balancing.
- To be effective such marking schemes need the participation of most of the intermediate routers.

To address these limitations, we further investigate the possibility of a source end marking scheme referred to as SEM (Source End Marking).

Like many other existing works such as [26, 27, 29, 113, 141, 143, 144], the proposed schemes assume the availability of K bits in the packet header which can be used by an intermediate router to encode additional mark into the forwarding packets. In [28, 113], the authors identify the ID field (16 bit), fragmentation tag bit (1 bit) and type of service field (8 bit) of an IPV4 packet header which can be used for marking purpose for different forensic activities. Similarly, the 24 bit Flow Label field of an IPV6 packet header can be used for marking purpose[142].

Table 5.1: Code distribution of different schemes for mwest monitoring point

	Mwest			
Hop Count	Paths	SPI1 Codes	SP2 Codes	XORID Codes
1	374	2	4	374
2	819	4	16	813
3	2287	8	64	2238
4	5960	16	250	5690
5	12181	32	900	11157
6	21985	64	2696	18719
7	31765	128	6509	25170
8	38095	256	12194	28891

5.3 XORID

In XORID each router maintains a 16 bit random key. We assume that the attacker does not know the keys maintained by the intermediate routers which are traveled by a packet from a given source S to the victim D. The XORID scheme works as follows. Whenever an XORID enabled router receives a packet, it first checks if the packet is received from its local LAN or not. If so, the router sets the 16 bit ID field of the packet with its own 16 bit key. On the other hand if the packet is received from one of its upstream routers, the receiving router first extracts the 16 bit ID of the received packet. Then the extracted 16 bits are XORed with the 16 bit secret key maintained by the router. The 16 bit result of the XOR operation is then used to replace the ID value of the forwarded packet. A demonstration of XORID marking scheme is shown in Figure 5.2. In the figure A is the source, V is the destination and $Router_i, i = 1, 2, ..n$ are the intermediate routers through which the packets from A traverse to reach V. When $Router_1$ receives a packet from A it simply copies its 16 bit secret key K_1 to the ID field of the packet and forwards the packet to the next hop $Router_2$. When $Router_2$ receives the packet it extracts the 16 bit ID field of the packet and performs XOR operation on it with its 16 bit secret key K_2. The 16 bit result of the XOR operation is then set as the new value of the ID field of the forwarded packet. The same process is then repeated up to the last hop $Router_n$.

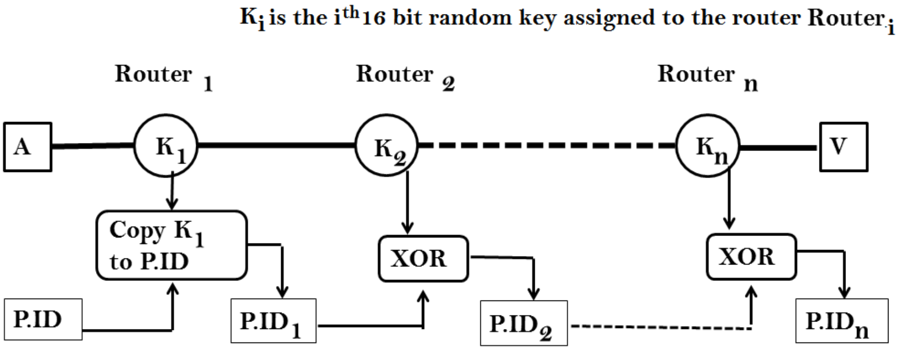


Figure 5.2: Demonstration of XORID marking scheme

Based on this simple marking scheme we state the following proposition

Proposition 1. 16-bit IDs are uniformly distributed among different paths irrespective of the number of hops between the source and destination

Explanation: For any two 16-bit numbers, say A_1 and A_2, if $a_i \epsilon A_1$ and $b_i \epsilon A_2$ are two arbitrary bits, then $P(a_i = 1) = P(a_i = 0) = P(b_i = 1) = P(b_i = 0) = \frac{1}{2} = 0.5$,

where i=1,2,..,16 and P represents the probability of occurrences.
Similarly, $P((a_i \oplus b_i) = 0) = P((a_i \oplus b_i) = 1) = \frac{1}{2} = 0.5$
Thus, if A_1 and A_2 are completely random, then $P(A_1) = P(A_2) = \frac{1}{2^{16}}$ and $P(A_1 \oplus A_2) = \frac{1}{2^{16}}$. It implies $P((..((A_1 \oplus A_2) \oplus A_3)..) \oplus A_k) = \frac{1}{2^{16}}$

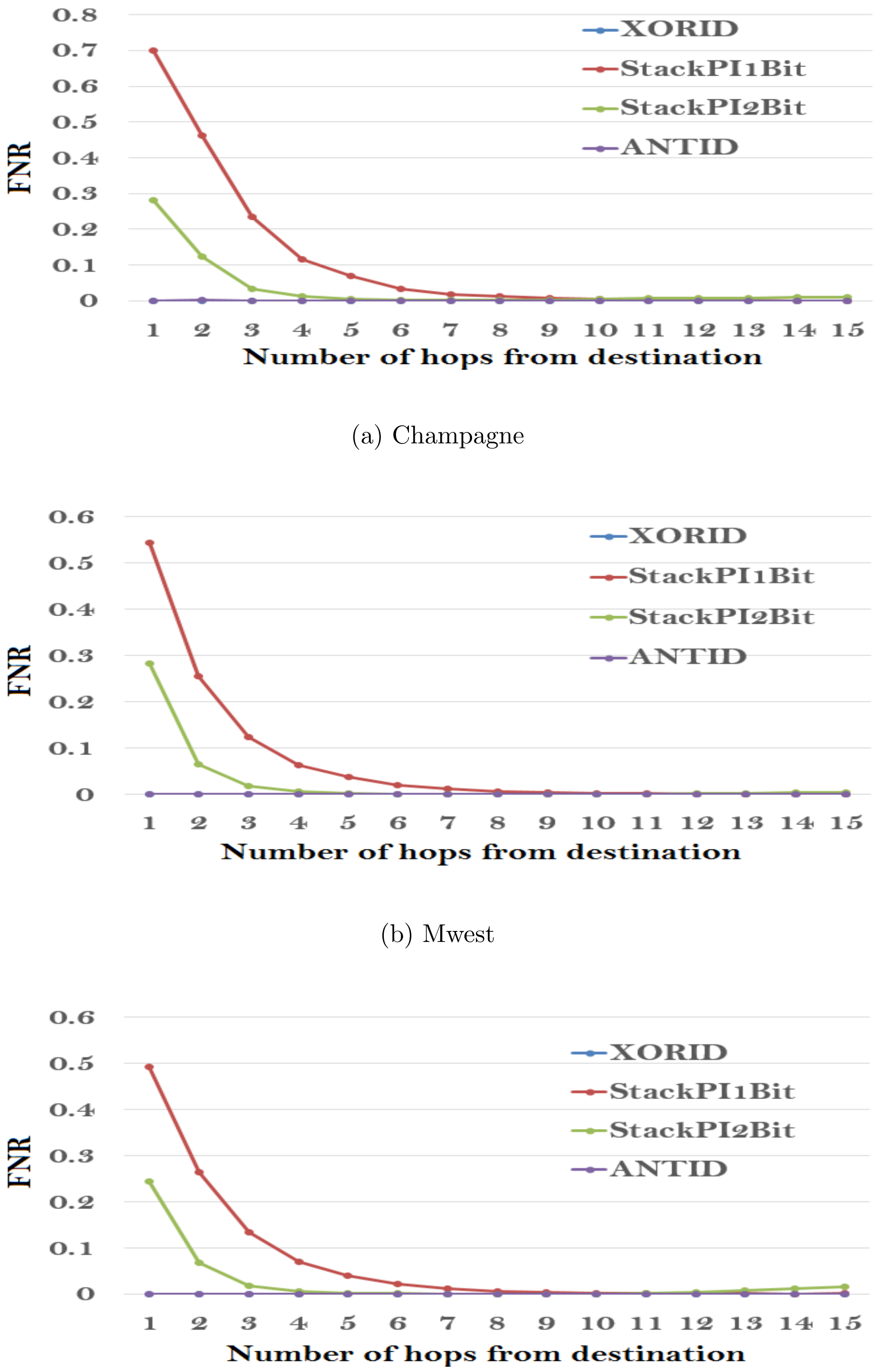
The distribution of the final marks received by the victim for different sources do not depend on the number of intermediate routers, and hence our scheme assigns 16 bit ID to paths with different lengths with equal probability of $\frac{1}{2^{16}}$.
In XORID scheme, a marking router alters the 16 bit ID value of the received packets based on its own secret key and the mark carried by the packets. Hence, XORID does not suffer from the problem of overflow, as it is the case with StackPI which allows XORID to encode a path up to the last hop towards the victim.

5.3.1 Analysis of XORID

To evaluate XORID marking scheme we use CAIDA's ITDK0304[64] Internet topology dataset. Skitter maintains trace route informations from selected monitoring points to different destinations. The datasets contain both complete as well as incomplete paths. Also for a single destination there are multiple paths in some of the datasets. For our experiments we consider only the complete paths. In case of multiple paths to the same destination from the monitoring point, we chose the longest path sequence. Thus, after preprocessing we get a tree, where the root of the tree is the destination and the leafs are the sources. Out of skitters 23 monitoring points, we chose the topology generated from champagne, mwest and sjc monitoring points in our experiments. However, topology related to other monitoring points also exhibit almost same characteristics.

5.4.1.1 False Negative Rate Analysis

To compare the false negative rate (FNR) of XORID with that of ANTID and StackPI schemes, we randomly selected 10000 different sources to construct the mapping table. Here we represent these set of sources as B (B for benign) and the rest of the sources as A (A for attack). For each distance $d = 1, 2, 3, ...31$ we select a subset of attack sources $A_d = \{x | distance(x, D) = d \ AND \ x \epsilon A\}$ and a subset of benign sources $B_d = \{x | distance(x, D) = d \ AND \ x \epsilon B\}$, where $distance(x, D)$ is the number of hops between x and D. For each attack source in A_d attack packets are sent to D by spoofing a randomly selected source IP address from corresponding B_d. At the destination end a received packet is accepted only when

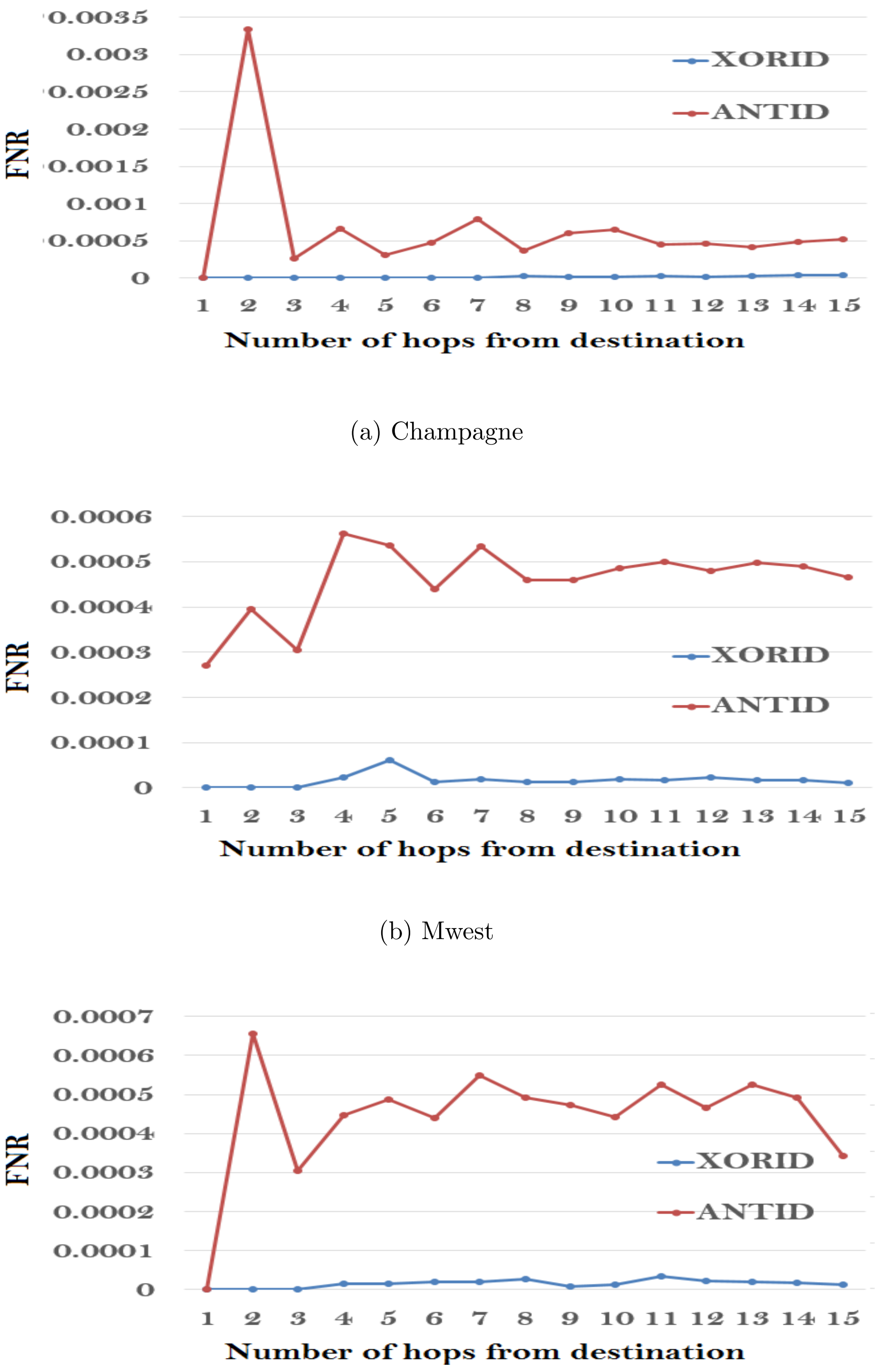


(a) Champagne

(b) Mwest

(c) Sjc

Figure 5.3: FNR comparison of XORID with StackPI and ANTID



(a) Champagne

(b) Mwest

(c) Sjc

Figure 5.4: FNR comparison of XORID with ANTID

its mark is same as the mark stored in the mapping table against its SIP. Thus all the accepted packets sent from the attack sources are considered as false negative. For each distance $d_i$we calculated the $false\ negative\ rate\ (FNR)$ as

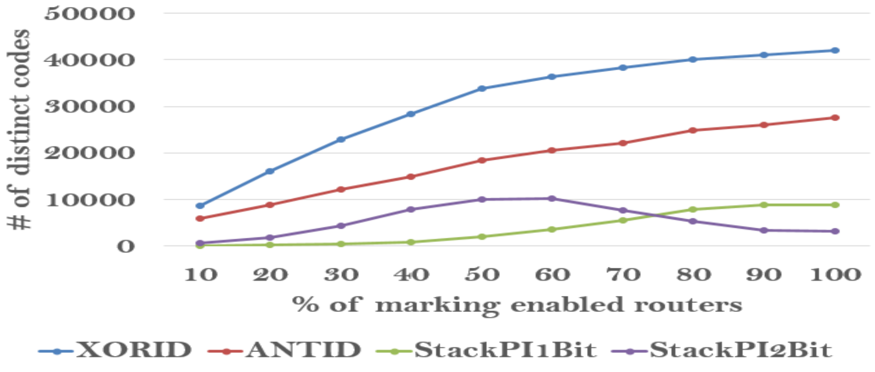
$$FNR = \frac{\#\ of\ accepted\ attack\ pakets}{Total\ number\ of\ attack\ packets\ received} \tag{5.1}$$

Figure 5.3 shows the FNR comparison of XORID with that of StackPI and ANTID for the above mentioned topologies . We see that the FNR of StackPI is much higher when the attacker mimics sources closer to the victim. However, the FNR of XORID and ANTID remains stable at all distances. Although the FNR of ANTID is comparatively lower than that of StackPI, our proposed scheme XORID achieves a better FNR over ANTID. We plot the FNR of ANTID and XORID separately at Figure 5.4 to show the fact.

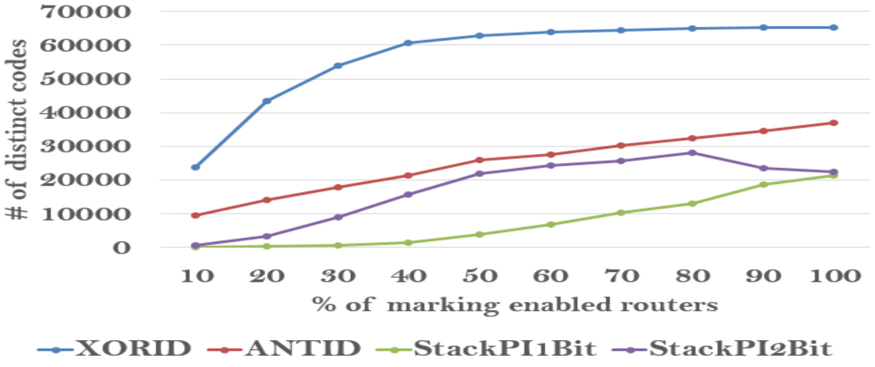
5.4.1.2 Incremental Deployment Analysis

All the above mentioned marking schemes assume the participation of intermediate Internet routers in the marking process. However, it will be quite unreasonable to expect the up-gradation of all the Internet routers in a single step. In this section, we compare the effectiveness of these schemes when a certain percentage of intermediate routers are marking enabled.

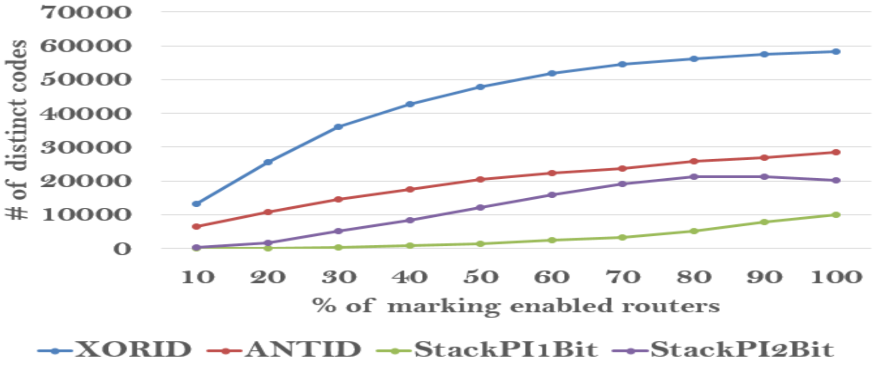
Such a scheme assigns a 16-bit code to each of the sources communicating to a destination. The assigned code depends on the path traveled by the packets from their sources to the destination. Since, the number of possible paths to a specific destination could be many times larger than the total number of 2^{16} different codes, such schemes might assign two or more different paths the same 16-bit code. Thus, one of the desired characteristics of such a scheme is to assign as many distinct codes as possible among all different paths. To observe the distribution of distinct codes among the paths under partial deployment scenario, we conducted the following experiment. For each of the topologies we randomly chose a fixed percentage of routers as marking enabled routers and counted the distinct number of codes assigned by different schemes to all different paths. The result is shown in Figure 5.5, where the horizontal axis represents the percentage of participating routers and the vertical axis represents the number of distinct codes assigned among all different paths. From Figure 5.5, we see that the distinct number of codes assigned by XORID scheme is many times larger than that of the other schemes PI, StackPI and ANTID. Even under partially deployed scenario where only 20-30% of the routers participate in the marking process, XORID assigns more number of codes to the paths than the distinct codes assigned by other schemes when all routers are considered as participating routers. Thus, XORID is more suitable under partial deployment scenario than the other discussed schemes.



(a) Champagne



(b) Mwest



(c) Sjc

Figure 5.5: Incremental deployment Analysis of XORID, StackPi and ANTID

5.4 A General Architecture of Source Identification Mark Based Spoofing Attack Defense System

In this section, I discuss a general architecture of source identification mark based spoofing attack defense system to demonstrate how the source identification mark carried by the incoming packets can be used to detect and defense spoofing based DDoS attacks, such as those described in section 1. Figure 5.6 shows the block diagram of such a defense system.

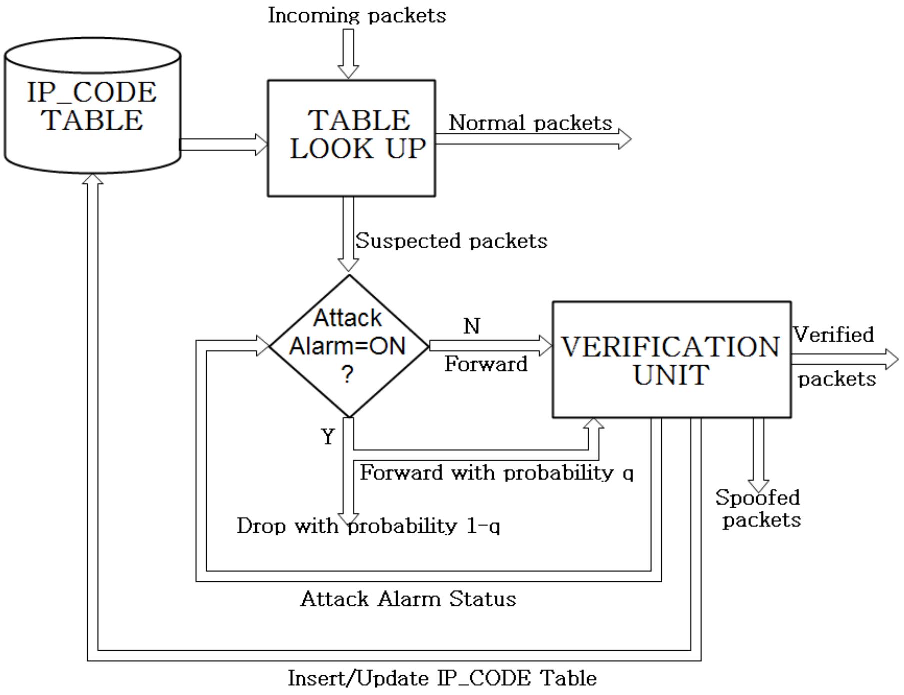


Figure 5.6: General architecture of source identification mark based spoofing attck defense system

To make use of the mark made available by an underlining marking scheme such as StackPI or XORID, the defense system maintains a table, referred to as *IP_Code* table, to keep track of the source IP addresses of the peers communicated with the protected site along with their corresponding marks. At any instant of time, an arrived packet's source IP and corresponding mark is verified against the *IP_Code* table. There are three different outcomes

- The SIP is not recorded in the table, indicating a new source IP.
- The SIP address is recorded but with a different mark. In ideal situation such

an outcome should be considered as a spoof incident. However, based on the marking technique used, there could be legitimate causes of such a mismatch. For example in case of StackPI, ANTID and XORID a packet reaching the destination through a different path other than the usual one might cause such a mismatch.

- The SIP address is recorded with the same mark as that of the received packet.

The defense system considers the first two outcomes as suspected and forwards such traffic to another unit, referred to as the verification unit. The task of the verification unit is to explicitly verify the genuineness of the claimed SIP of a received packet. The explicit verification can be done by sending a request probe message to the claimed source IP address and then the mark carried by the response probe packet can be used to decide the genuineness of the source. Typical request/response probe messages could be

- ICMP echo request and its corresponding response message
- A UDP packet sent to an arbitrary port and the corresponding ICMP port unreachable message, assuming no valid application is running on that port.
- A TCP packet sent to an arbitrary port and corresponding RST message.

The verification unit maintains a state table to keep track of the request/response probe messages to/from different sources. Also, each entry in the state table is associated with buffer space to accommodate further packets received from a source which is waiting for a response probe message. When the verification unit receives a packet, it first checks its state table for an entry corresponding to the source IP address of the packet. If an entry is found it indicates a probe request is already sent to the SIP address. In this case the packet is quid in the corresponding buffer. Otherwise a new entry is allocated in the state table corresponding to the source IP address of the received packet. The packet itself is then buffered and a probe request packet is sent to the source. When a probe response packet is received by the verification unit, the mark of the response packet is compared with that of the mark of the packet which initiated the corresponding probe. If both the mark matches the verification unit forwards all the buffered packets to the destination. The verification unit then either enters a new entry or updates an existing entry in the mapping table. On the other hand, if the received mark is different than the original mark, the verification unit considers the source as spoofed and drops all buffered packets against the SIP address. The verification unit maintains a counter, referred to as $spoof\_count$ to indicate the spoofed incidents encountered within a certain time interval. Every time a spoof event is detected the counter is incremented by one. A large value of this counter clearly indicates a ongoing spoofing attack. The defense system generates an attack alarm if the $spoof\_count$ value is greater than a threshold T. Under normal condition suspected sources received by verification unit is expected to be low. However, under a spoofing attack the number of distinct SIPs could be very large as the attacker can potentially use any IP as the SIP of an attack packet. Hence, under an attack the verification unit might end up sending a large number of probe packets to all these spoofed sources. To avoid such a situation when an attack is detected the defense system considers all the new and suspected traffic as attack traffic and drops them with high probability q. Only a small percentage of suspected sources are considered for verification under an attack condition. The attack is considered to be going on as long as the $spoof\_count$ value is above the threshold. Under an attack two types of benign packets will suffer

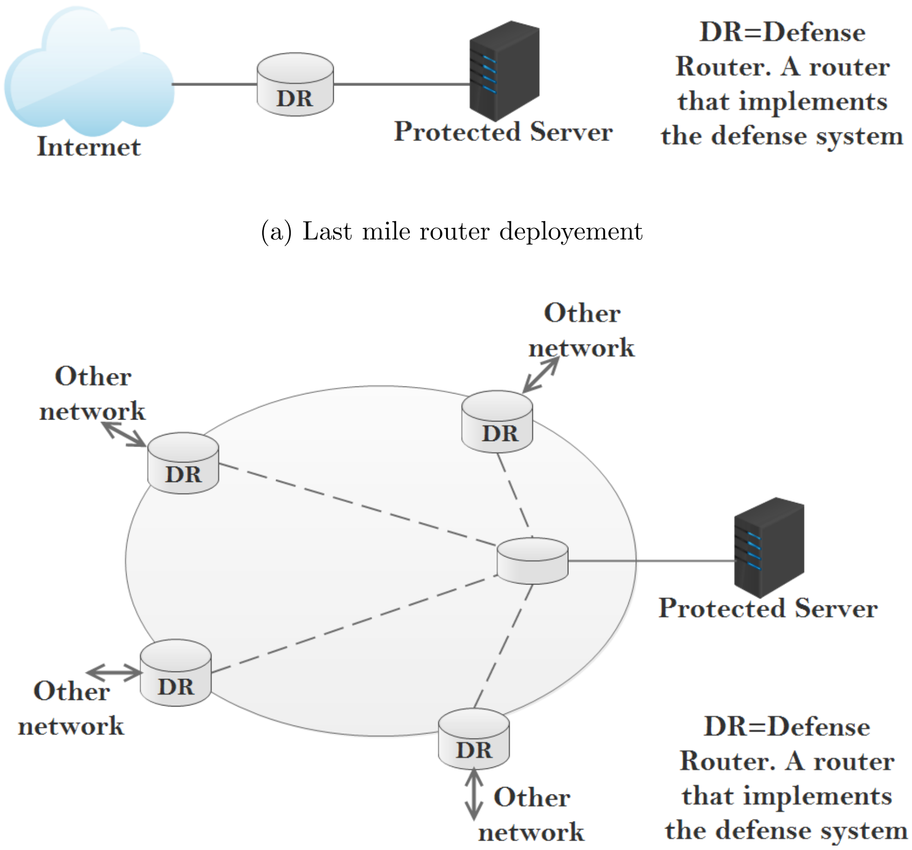
- Whose source IPs are not listed in the IP_Code table
- Whose SIPs are listed but mark is different

The first error could be minimized by keeping the size of the IP_Code table large enough to accommodate most of the previously seen SIPs along with their marks. However, the defense has less scope to improve the collateral damage caused by the second type of error. Although the fraction of such benign traffic is relatively small, the dropped traffic might affect some of the ongoing communications. For example, one or more dropped packets in a TCP communication might activate TCP's congestion control mechanism and thus might reduce the benign user's throughput.

5.4.1 Deployment Location of The Defense System

The defense is a victim end defense system and typically placed just in-front of the protected server, as shown in Figure 5.7(a).

However, it should be noted that such a defense mechanism assumes that even under an attack situation all the traffic destined to the victim are available for inspection by the defense system, i.e. the link carrying traffic to the defense point is not saturated and no legitimate packets are dropped up to this point. Hence, if the attack traffic bandwidth is greater than the link capacity the defense might


Internet
DR
Protected Server
DR=Defense Router. A router that implements the defense system
Other network
Other network
DR
DR
DR
DR
Protected Server
Other network
Other network
DR=Defense Router. A router that implements the defense system

(a) Last mile router deployement

(b) Edge router deployement

Figure 5.7: Different deployment locations of the anti spoofing defense system

miss legitimate packets even before they are examined by it. To avoid such a situation the defense could be placed at the edge of the victim network as shown in Figure 5.7(b). In such an arrangement each edge router of the destination network implements the defense and inspects the incoming packets entering through them individually. The capacity of the links carrying traffic to the edge routers from the Internet is much higher than the link capacity between a server and its first router. Hence, the probability of legitimate packet drop at these points are comparatively less. Another advantage of such deployment is it allows attack traffic to be handled at different points through which they enter the victim network.

5.5 Limitations of Path Encoding Schemes

In the above section, I discussed about source identification mark based anti spoofing techniques such as PI, StackPI, ANTID and XORID which allows a packet to carry a 16 bit mark which is based on the path traveled by the packet from its source to the destination. In such a scheme if P_1 and P_2 are two different packets which traveled through the sequence of routers $r_1, r_2, r_3, ..r_n$ and $q_1, q_2, q_3, ..q_m$ respectively to reach a specific destination D, then at the destination end

- $P_1.mark = P_2.mark$ if the sequence of routers $r_1, r_2, r_3, ..r_n$ is same as $q_1, q_2, q_3, ..q_m$. i.e. packets traveling the same path to reach the destination will always carry the same mark.
- Probability of $P_1.mark = P_2.mark$ is extremely low if the sequence or routers $r_1, r_2, r_3, ..r_n$ is different from $q_1, q_2, q_3, ..q_m$. Typically these marks are of 16 bits in size. Hence, theoretically probability of any two random paths to a specific destination D encoded with the same mark is $\frac{1}{2^{16}}$

There are two main disadvantages of such a path based marking schemes.

- A defense mechanism based on such techniques assumes that packets from a specific source S to a specific destination D usually travel the same route. However, the Internet is based on the *best effort service to deliver packets from source to destination* policy and hence, the paths traveled by the packets between any $< S, D >$ pair might change dynamically. The fraction of such traffic is low under normal condition and hence, explicit verification of such packets can be done under normal condition. However, under an attack condition such *out of the path* packets could be very high and the defense might not have a way to distinguish legitimate packets with different marks(caused by legitimate path change) from attack packets.
- These techniques assume the participation of a large fraction of intermediate routers in the marking process. Internet is composed of many autonomous systems managed by different administrations. Hence, such a global participation might not be possible.

5.5.1 Source End Marking(SEM)

In this section, I discuss a source end marking scheme referred to as SEM to overcome the above mentioned shortcomings of path based marking schemes. A source network implements SEM in all the routers which connects one or more sources with the network. Figure 5.8, shows a typical network showing the routers at which SEM should be implemented. In SEM each packet is marked when it enters

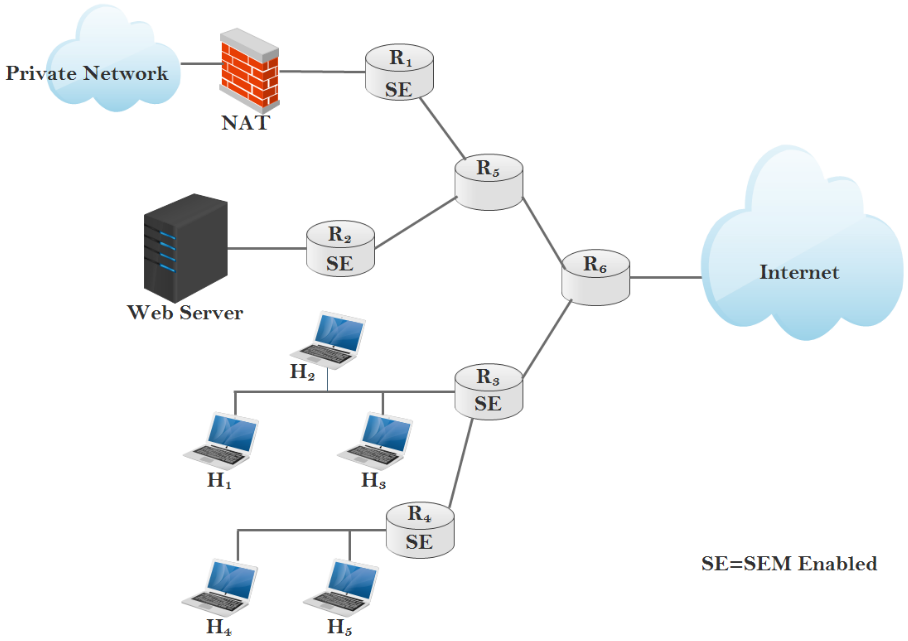


Figure 5.8: SEM enabled routers in a network

into the Internet through one of the SEM enabled routers. SEM uses a function f to generates the mark m of a packet based on the following 3 parameters

- a secret key, K, maintained by an autonomous system,
- the IP address of the ingress interface $ingressIP$, through which a packet enters the network, and
- the destination IP address of the packet, $destIP$.

5.6.1.1 The Mark Generation Function

The function f, which is used to generate the mark m, from the above mentioned parameters could be any mathematical function with the following characteristics

- it maps the input to a 16-bit bit pattern more or less evenly, i.e., for any given K, $ingressIP$ and $destIP$, the probability of getting a specific m is approximately $\frac{1}{2^{16}}$.
- it is deterministic so that the same combination of parameters always results in the same mark.
- computational cost of the function is as light as possible.

- it is infeasible to recreate the original parameters from the obtained mark except by using a brute force search of all possible combinational of parameter values, and
- a small change in the parameter values should return a mark such that the new mark and the old mark appears completely uncorrelated.

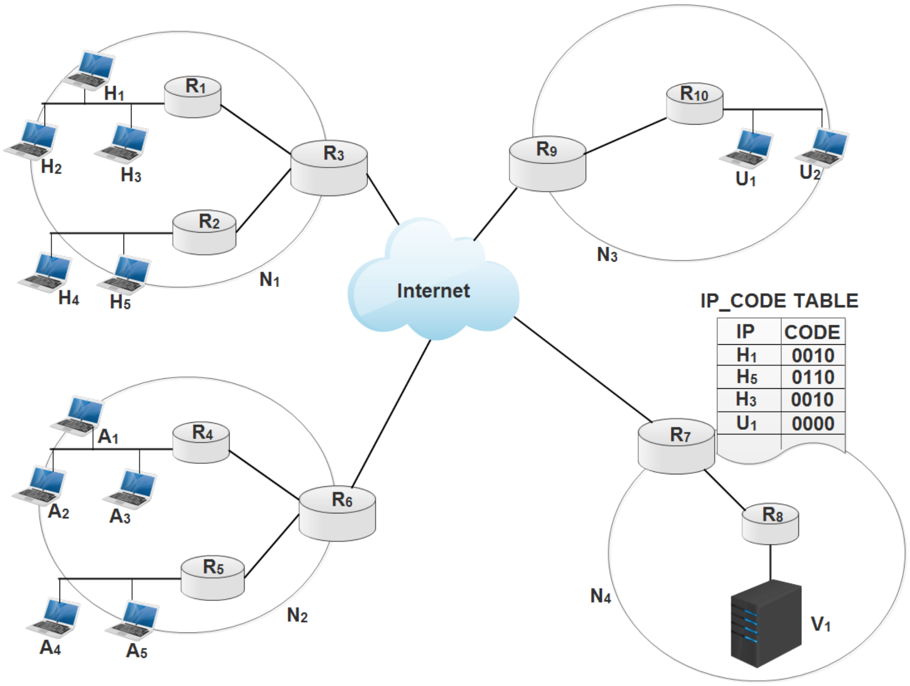
The first characteristic is desirable so that the possibility of acceptance of an attack packet by the defense system becomes minimum. The second characteristic makes sure that all packets from a specific source IP belonging to a particular AS always carry the same marks to a specific destination. The light computational cost of the mark generation function is desirable as this function is to be executed for every outgoing packets by a SEM enabled router. The fourth and the fifth characteristics are important to protect the scheme against replay attacks.

The above mentioned requirements can be met by using a cryptographic hash function such as MD5 [96]. One possible approach to implement f is to return the last 16 bits of the MD5 hash value of the concatenation of the parameters. i.e., $m = f(K, ingressIP, destIP) = last16Bits(MD5(K + ingressIP + destIP))$,where, $last16Bits$ returns the last 16 bits of the input bit string, $MD5$ returns the MD5 hash value of the input string and + represents concatenation of strings.

One important point is to be noted that the mark generation function, f, is not a secret to the attackers. I.e., an attacker is assumed to know the mathematical steps which are invoked in the mark generation process. The only component which is not known to the attackers is the value of the secret key K maintained by an AS.

5.6.1.2 The Parameters of Mark Generation Function(f)

In this section, I describe the parameters used by SEM scheme to generate the mark corresponding to an outgoing packet. Figure 5.9 is used to discuss the importance of each of these parameters. In Figure 5.9, N_1 is a SEM enabled network, i.e., all the outgoing packets from this network are marked by the first router along the path towards any destination. N_2 is an ill-protected network, i.e., one or more sources from this network can send spoofed packets to any destination over the Internet. N_3 represents a network which does not implement SEM to protect its users from being spoofed. However, N_3 might implement some other source end anti spoofing techniques such as ingress/egress filtering. N_4 represents a network which has deployed anti spoofing defense system, similar to the one described in section 5.5, and makes use of the marks carried by the incoming packets in their 16-bit ID field to distinguish the spoofed packets from the benign packets.



IP_CODE TABLE

IP	CODE
H1	0010
H5	0110
H3	0010
U1	0000

Figure 5.9: SEM enabled routers in a network

- *Secret key, K*: Each autonomous system(AS) maintains a secret key, K, which is shared by all SEM enabled routers in that AS. Although, the network administrator of an AS can chose any key, but, while choosing the key it should be noted that a small length key can easily be cracked by an attacker using brute-force approach. On the other hand, an extremely long key might slow down the key generation function.

  The secret key prevents an attack source within an unprotected network (i.e. a network from which injection of spoofed packets into the Internet is possible) to generate the correct mark against the SIP which is being spoofed. Even if, the attacker is assumed to know the ingress interface IP address associated with the actual source IP, without knowing the secret key it is impossible for an attacker to compute the correct mark against the specific source IP address.

  For example, in Figure 5.9, an attacker A_1, from N_2 wants to send attack packets to victim V_1 in network N_4 by spoofing the IP address of H_1 in network

N_1, which is a legitimate user of V_1. In this case, since the attacker has no clue about the key maintained by N_1, it can not generate the actual mark against H_1. Hence, the only way left to A_1 is to set the mark of the attack packets to some random pattern. However, since the victim is protected by an anti spoofing defense system deployed at R_7, such attack packets will be discarded immediately by the defense.

- *Ingress IP address, ingressIP*: The IP address of the ingress interface associated with a source IP address is used to protect internal spoofing within a SEM enabled network. Since the mark of a packet depends on the ingress IP address through which a packet enters the network, the possibility of an attack source mimicking another IP address withing the same network is very limited. An attack source can spoof only those legitimate sources which are connected via the same ingress interface. For example in Figure 5.9 if H_1 is an attack source it can spoof only the IP addresses of H_2 and H_3 as they share the same ingress interface. However, if H_1 attempts to spoof any other sources connected via a different ingress interface, for example H_4, the spoofed packets from H_1 will be marked with different marks than the usual mark associated with the actual source H_4.

- *Destination IP address, destIP*: The destination IP address is used in the mark generation process to make sure that different destinations receive different marks for the same source. Hence, even if an attacker collects the mark carried by the packets from a specific source S_1 to a specific destination D_1, it does not do any good to the attacker to send spoof packets with S_1 as the source to another destination D_2.

For example, A_1 might try the following sequence of actions in an attempt to invade the defense system. First, A_1 sends a probe packet (A TCP SYN, or ICMP echo request packet) to H_1. From the reply packet A_1 extracts the mark and then use the mark to send attack packets to V_1 by spoofing itself as H_1. However, since the mark in the reply packet from H_1 to A_1 is generated based on A_1's IP address, it is unlikely that the same mark will be generated for the genuine packets which H_1 sends to V_1. Thus, A_1's attempt to reply the mark will fail with a very high probability.

5.5.2 How SEM is Different From Path Based Marking Schemes

SEM is different from existing path based marking schemes such as PI, StackPI, ANTID and XORID in that

- SEM code is inserted at the first router through which a packet enters into the Internet. Hence, even if two different packets from the same source reach a specific destination through two different paths the mark of the packets will be identical.
- SEM is a source end marking scheme i.e. SEM does not need the participation of the routers of the intermediate networks, whereas performance of existing schemes highly depend on the fraction of intermediate routers participating in the marking process. The best result is possible only when all routers along a path participates in the marking process in such schemes.
- Deployment of SEM at a source network makes the sources un spoofable, i.e. a specific destination can always verify the source IP of a received packet based on the mark carried by it. Hence, deployment of SEM protects the user of a source network from
  - Being rejected by a defense system when the protected site is under a spoofed DDoS attack. For example, in Figure 5.9, say, A_1 from N_2 wants to send attack packets to V_1 in N_4 by spoofing the SIP address of the attack packets as the IP address of H_1. Since the attacker never knows the actual marks carried by the genuine packets from H_1 to V_1, the best option for the attacker is to put some random marks in the attack packets. The anti spoofing defense system deployed at R_7 can detect majority of the attack packets based on their unusual marks. Thus, an attacker can't perform malicious activity by mimicking itself as a legitimate user from a SEM protected network. On the other hand, if A_1 wants to send attack packets to V_1 by mimicking itself as U_1, A_1 can execute the following steps. First, A_1 sends a probe packet to U_1. From the response packet from U_1, A_1 extracts the 16-bit ID field. Then, A_1 sends attack packets to V_1 by setting the ID field of the attack packets as the extracted ID of the response probe packet and the SIP of the attack packets as the IP address of U_1. The anti spoofing technique at R_7 won't be able to distinguish the attack packets based on their

marks. At later stage, the malicious packets from A_1 (which are spoofed to disguise as U_1) might be detected by other DDoS defense modules and based on such deployed DDoS defense system the IP address of U_1 might be considered as an attack source. Under such a situation, the genuine packets from the U_1 to V_1 will also be discarded by such DDoS defense system.

– Being the victim of a DRDoS attack. If the packets from a source is marked with SEM then it gives the reflection servers such as NTP, DNS and SMPT a way to verify an incoming request packet before actually sending the response to the source IP carried by the packet. For example, in Figure 5.9, consider the situation where A_1 wants to launch a DRDoS against H_1 using V_1 as a reflector. To achieve this A_1 will first send request packets to V_1 by mimicking itself as H_1. However, as explained above, the anti spoofing defense system deployed at R_7 will immediately detect such spoofed packets and will discard them all. Thus, H_1 will not receive any unwanted responses from V_1. On the other hand A_1 can easily launch a DRDoS against any source in network N_3. Say A_1 wants to launch a DRDOS attack against U_1 using V_1 as a reflector. To achieve this A_1 first collects the default ID value of the packets transmitted from U_1. Then, A_1 sends request packets to V_1 mimicking itself as U_1 by setting the ID value of the attack packets appropriately. The anti spoofing defense deployed at R_7 wont be able to distinguish the bogus request packets based on their marks. Hence, V_1 will receive the bogus requests and accordingly will send the responses to U_1.

It should be noted that deployment of existing schemes such as PI, StackPI, ANTID and XORID only at the source network does not facilitate this advantages to the source network.

Table 5.2: Node distribution of SEM scheme for different topologies

Topology	No of leaf nodes	Total No of intermediate routers	No of SEM enabled nodes	% of SEM enabled nodes	No of different 16 bit prefixes of the SEM enabled routers
Champagne	101274	55822	40685	72.88	5380
Mwest	555043	109250	93264	85.36	6778
Sjc	269849	88038	71332	81.02	6476

5.5.3 Experimental Evaluation of SEM

To evaluate the performance of SEM the same preprocessed network topology datasets as described in Section 5.4.1 are used. In the topologies all the leaf nodes are considered as individual sources and the nodes which are one level up from the leafs are considered as the set of routers through which one or more sources are connected with the Internet. To perform the experiment we considered the first level routers as SEM enabled routers. The SEM enabled routers are assigned a random key of length more than 32 bits based on their 16 bit suffix, i.e. all the SEM enabled routers having the same 16 bit suffix are considered to be in the same autonomous system in the experiment. A pictorial representation of the experimental arrangement is shown in Figure 5.10. In the figure R_1, R_2, R_3 and R_4 are SEM enabled routers as they connect one or more sources to the rest of the Internet. Out of these SEM enabled routers R_1 and R_2 are considered to be in the same autonomous system (AS_1) as they have the same 16 bit suffix address (192.168). Similarly R_3 and R_4 are considered to be in another autonomous system

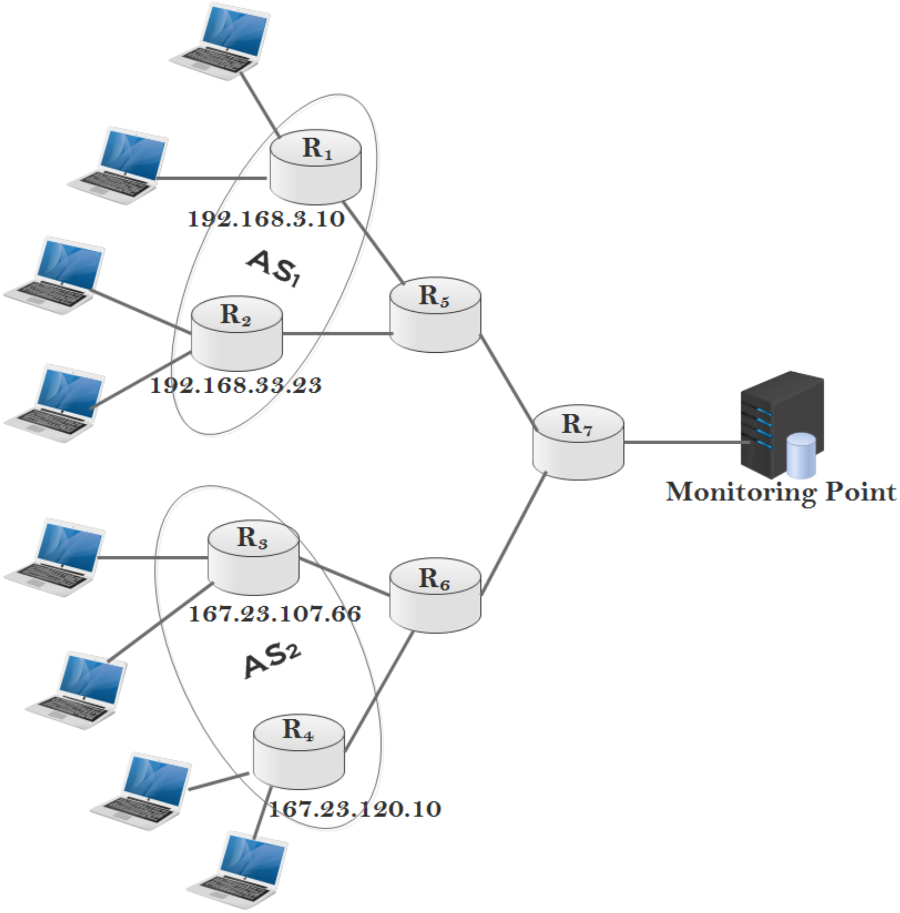


Figure 5.10: Pictorial representation of the experimental arrangement

(AS_2).

Table 5.2 lists the number of leafs, number of SEM enabled routers, number of different 16 bit suffixes of the SEM enabled routers and the total number of routers in each of the three considered topologies.

To compare the performance of SEM with StackPI, ANTID and XORID, we randomly choose 10000 different sources to construct the corresponding IP_Code tables at the destination end. The rest of the sources are then used to send attack packets to the destination by replacing the SIP address as on of those listed in the mapping table. At the destination end the FNR is calculated using Equation (5.1). Figure 5.11 shows the experimental FNR of each of these schemes.

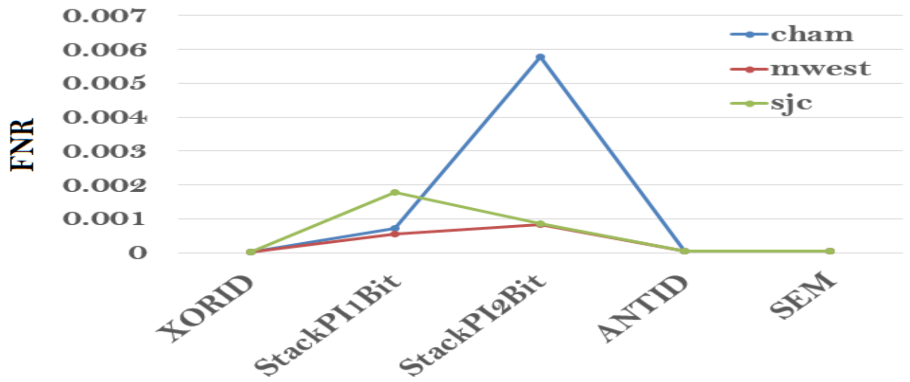


Figure 5.11: FNR comparison of SEM with StackPI, ANTID and XORID

We see that the FNR of ANTID and XORID seems to be comparable with that of SEM. However, it should be noted that SEM requires only the first level routers (i.e., the routers through which one or more sources are connected to the Internet) to participate in the marking process, whereas in case of StackPI and ANTID only the first level router implementation will mark all the packets from a specific source S to any destination $D_i, i = 1, 2, ..n$ with the same mark. Thus an attacker can easily collect the mark associated with a source first and then can send spoofed attack packets to the victim with matching pair of SIP and corresponding mark to bypass the defense completely.

5.6 Discussion

In this chapter, I propose and demonstrate two packet marking schemes namely, XORID and SEM, to inscribe a mark into the forwarding packets. I discuss how such marks can be used by a defense system to effectively detect spoofed packets. The experimental results show that XORID performs better than other similar packet marking schemes such as PI, StackPI and ANTID where, the final mark of a packet depends on the intermediate routers traveled by the packet to reach the destination. On the other hand experimental results and theoretical explanation of SEM show that a source end marking scheme is more beneficial in terms of accuracy and incremental deployment. I also demonstrate how participating in SEM can protect a network from being the victim of reflection DDoS attacks. In the next chapter, I discuss about another approach to mitigate spoofing attack, called as IP traceback.