A Brief Survey of IP Traceback Methodologies

Vijayalakshmi Murugesan, Mercy Shalinie, Nithya Neethimani

Department of Computer Science and Engineering,Thigarajar College of Engineering, Thiruparankundram, Madurai – 625015, Tamilnadu, India
mviji@tce.edu,shalinie@tce.edu, nithyaarun@tce.edu

Abstract: The problem of identifying DDoS (Distributed Denial of Service)Attack is one of the prevalent threats in the field of Internet security. The difficulty lies in distinguishing the attack traffic from the normal traffic, as their attack origin is often hidden. Several techniques are used to detect and identify the source of DDoS attack. One of the most popular techniques in identifying the attack source is the IP traceback mechanism. Different kinds of traceback approaches are proposed with each having its own advantages and disadvantages. This paper presents and evaluates some of the existing and recently evolving IP traceback techniques with respect to their pros and cons.

Keywords: Survey; Review; IP Traceback; Spoofed IP address; DDoS Attack; Packet Marking; Logging; Hybrid

1 Introduction

Denial of Service (DoS) attack attempts to generate a huge amount of traffic to the victim and thereby disrupting the service or degrading the quality of service, by depleting the resources. Distributed Denial of Service (DDoS) attack is a distributed, co-operative and large-scale attack. Attackers can launch the attack traffic from various locations of Internet, exhausting bandwidth. The processing capacity or memory of the target machine or network is drained, taking advantage of the vulnerabilities and anonymous nature of Internet. Both these attacks have been posing a major threat to the Internet for over a decade. Now-a-days these attacks are turning to be more sophisticated. DDoS attack takes place from multiple attack path from numerous zombies controlled by an attacker. According to the recent survey of Arbor networks the impact of DDoS attack is increasing every year. Even the key players such as Microsoft,Yahoo,e-bay are counted in the list of DDoS victims. The packets sent will have spoofed IP addresses [1, 2, 3] which makes it practically difficult to identify the real location of attackers. Defending an attacker with spoofed IP address is more complex and this motivates the research on IP traceback,which is a methodology to trace the true origin of spoofed IP packets.

DDoS attacks can be launched in two forms,namely, direct attacks and reflector attacks [4]. In the Direct attack, the attacker floods the spoofed packets to the victim via zombie machines. Direct attack is further classified into Network-layer DDoS attack (e.g. Ping flood, TCP layer attacks, Routing attacks, ICMP flood etc.) and Application-layer attacks (e.g. HTTP flood, HTTPS flood, FTP flood, etc.). The Reflector attack involves sending spoofed request packets to a large number of machines (known as reflectors) that will send reply packets to the requested source. The spoofed request packet will hold the source address of the targeted victim and so the replies from all the reflector machines will flood the source, targetting the victim. ICMP Echo Request attacks commonly known as Smurf attack is a well-known reflector attack.

The counter measures against these DDoS attacks are broadly classified into proactive mechanism, reactive mechanism, and survival mechanism [4]. IP Traceback is one such reactive technique. IP Traceback is used to find the origins and attacking paths of malicious traffic. In general, IP traceback is not limited only to DDoS attack. IP Traceback is defined in [5], as identifying a source of any packet on the Internet. The task of identifying the original source of a packet is complex as the source IP address can be fake or spoofed. The source of these packets may be the actual attacker but in most cases, it might be a reflector, a zombie as stated above or a device compromised by the attacker in some other way. IP traceback techniques neither prevent nor stop the attack, they are used only to identify the source of the packets. Different IP traceback techniques are proposed only to mitigate DoS/DDoS attacks. A survey on existing IP Traceback schemes is already done and evaluated in [6], which has not included the recent developments. This paper focuses on a detailed discussion on various traceback schemes ranging from the traditional Link testing to the newly emerged Hybrid schemes and analyze them with additional evaluation metrics.

The paper is organized as follows – the classification of various schemes is dicussed in Section 2, followed by a brief description of the metrics used for evaluating the different methodologies in Section 3, and a precise comparison of schemes based on the evaluation metrics is presented in Section 4 and Section 5 forms the conclusion of the paper.

2 Classification of IP Traceback Schemes

The intent of IP Traceback mechanism is to locate the source of the packet. As the source IP address of the packet is often forged or spoofed, IP traceback mechanism is inevitable. Traditional traceback mechanisms like Link Testing which includes Input Debugging and Control Flooding [1], have emerged a decade ago and recent techniques that are either combination of or completely different from the traditional ones are discussed here. IP Traceback schemes can be applied in two ways [7] – Intra AS and Inter AS. Intra AS Technique involves

traceback within the network and Inter AS technique involves traceback across various networks. The different types of IP Traceback Schemes are shown in Figure 1 and the description of each scheme is given below.

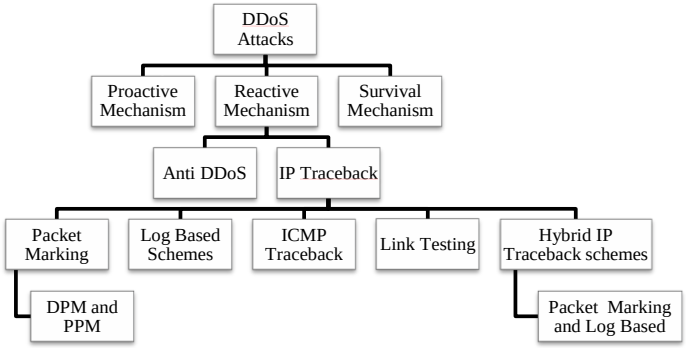

Figure 1
Classification of IP Traceback Scheme

2.1 Link Testing

The overview of link testing as shown in Figure 2 starts from the victim and traces till the attack source via upstream links with the assumption that the attack remains active until the completion of the trace. This scheme,therefore, will not be suitable to identify the attack that occurs intermittently or when the attacker is aware of the traceback scheme used. Input debugging and controlled flooding are the two varieties of Link testing. In Input Debugging [8] technique, the victim has to recognize that it is being attacked and has to develop an attack pattern (called attack signature) and check that with each of the incoming packets in the upstream routers and identify the corresponding upstream router and proceed further till the attacker. The most significant problem of this method is the management overhead, the co-ordination from the network admin. If the admin is unavailable or if he lacks the skill to assist the traceback, then the traceback may be slow or its completion could be impossible. Another variation of link testing is Control Flooding [9] which does not require any support from network operators. This technique tests incoming links of the victim by iteratively flooding each link with large bursts of traffic to see its effect on the incoming traffic. By observing the change in the rate of packets received, the victim can infer from which link the attack packets have arrived. This procedure is applied to the next upstream router until the origin of attack is reached. This kind of traceback by itself floods the network.

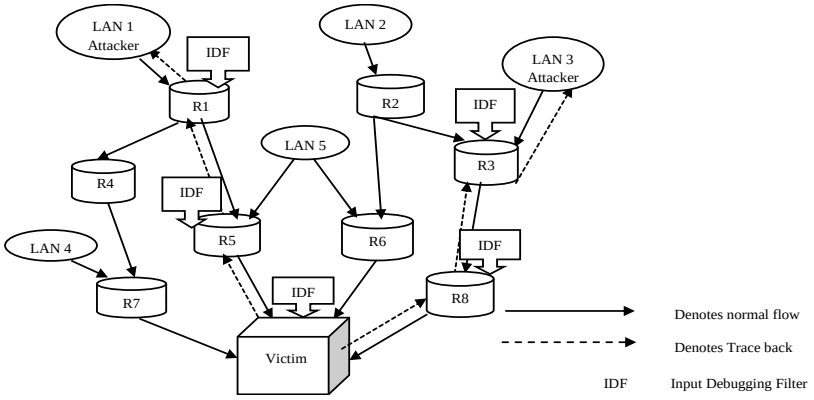

Figure 2
Link Testing

Link Testing which is also known as Hop by Hop Tracing uses an automated Pushback mechanism in [10] and it is currently supported by many router manufacturers. This uses statistical and pattern based analysis at the router closer to the victim to identify the upstream router from which the traffic has been forwarded and is repeated until the origin is reached. The statistics suggests the presence of attack and the pattern is used to distinguish the normal packets from the illegitimate attack packets.

2.2 Packet Marking

One of the common and significant techniques of IP Traceback is packet marking. The marking utilizes the rarely used fields of IP header, to store the audit trail where the field size used for marking varies from scheme to scheme. The dawn of packet marking era began with Node append, Node sampling, Edge sampling [1] marking methods etc. Each method emerged with the purpose to overcome the difficulties faced by the other. Packet marking mechanism is broadly classified into Probabilistic Packet Marking and Deterministic Packet Marking.

2.2.1 Probabilistic Packet Marking

Probabilistic Packet Marking method [1, 12] is shown in Figure 3. In this method, each router marks the packet with some probability say p for example $p = 1/100$ which implies marking one packet for every 100 packets received. The marking field uses 16 bits identification field in the header, of which 5 bits are used for marking hop count, which would be a useful information during reconstruction of attack path, and the remaining bits are used by the router to send its information.

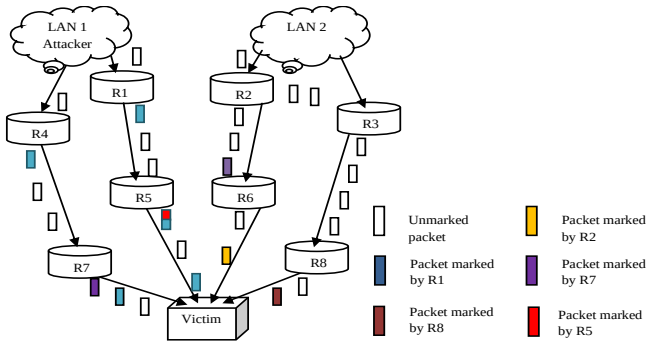

Figure 3
Probablistic Packet Marking

If the information is too large, then it is broken into fragments and marked in multiple packets. The marked packets will therefore contain only partial information of the path. This reduces the storage overhead in the packets. The victim has to receive enough number of packets to re-construct the path. This scheme does not require prior knowledge of the topology. The disadvantage of this scheme is that it produces many false positives and the mark field value written by routers far away from victim might be overwritten by the routers closer to the victim and if the attacker is aware of the scheme, then the traceback fails. This scheme is improved by [13] which uses hashed message authentication codes (HMACs) at each router, which reduces the number of packets for reconstruction when compared to previous scheme but has scalability issues and requires topology information. A traceback scheme using PPM proposed in [14] uses traffic rates of packets to identify the source of attack. Statistical analysis on the traffic rate combined with PPM is used for traceback with an assumption that the traffic follows Poisson distribution and depends on queuing model. The marking field is used to hold start, end and distance. The start and end fields store the IP addresses of the routers residing at the two end points making an edge and the distance field registers the number of hops between this edge marked and the victim. The problem of attacker taking advantage of the scheme has been overcome in [11] that recommends traceback using randomize and link approach. The main idea of this approach is that each router fragments its message into several words (pieces) and calculates checksum for the whole message named as ‘cord’. The mark value consists of checksum cord and message fragment and an index of the message fragment. The index and checksum are used to identify the message fragment during reconstruction. The total number of bits used for packet marking in this paper is 25. Reconstructing large messages requires more packets. Increasing checksum size increases security, but when the checksum bits are increased, message bits are decreased. Hence reconstruction will be time consuming. The drawback of requirement of large number of packets to traceback an attacker using PPM is addressed in [15] with minimum number of packets.

Apart from the above PPM schemes, algebraic approach [16] and Chinese remainder theorem based approach are also based on Probabilistic Packet Marking.

2.2.2 Deterministic Packet Marking

Deterministic Packet Marking scheme (DPM) shown in Figure 4 was first proposed in [17] to overcome the disadvantages of PPM. Every packet passing through the first ingress edge router is only marked with the IP address of the router. The IP address is divided into two fragments (16 bits each) and each fragment is randomly recorded into each inflowing packet. The entire IP address is recovered by the victim when the victim obtains both the fragments of the same ingress router. This scheme fails when the source address is spoofed and is also false positive. The enhanced schemes [18, 19] are proposed where the IP address is split into more fragments, and a hash function is used to contain the identity of the ingress router to decrease the false positive. Deterministic packet marking based on redundant decomposition is proposed in [20]. The knowledge of topology plays a significant role in DPM scheme's traceback. Consider the DPM scheme suggested in [21] where, it is assumed that the topology of the network is known in advance. The packet marking method involves hash of ingress router's IP address. The hash value is split into chunks and each chunk is marked into the packet randomly. With the topology known, the victim performs traceback of the marked routers. Large numbers of packets are not required for traceback in this scheme but it consumes a longer search time to identify the origin. The traceback scheme is challenged, if the topology is modified.

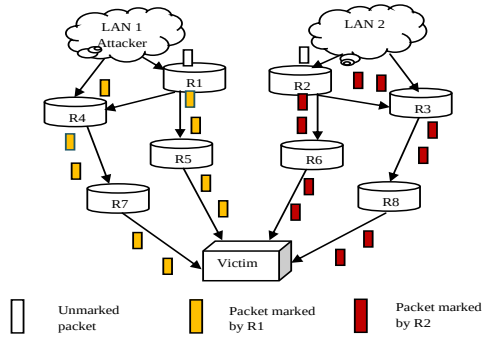

Figure 4
Deterministic Packet Marking

When an intermediate router goes off, the traceback can be carried out with the topology but might turn to be false positive. If the attacker modifies the mark field, this scheme will fail to traceback. Instead of IP address respective bit fields were marked in [22].In Flexible Deterministic Packet Marking (FDPM) scheme [23, 24], the marking field length is varied according to the requirement. The

length of the marking field is flexible and can be adjusted. Further, the marking rate can also be adaptively changed according to the incoming traffic load on the participating router. FDPM is capable of tracing a large number of real sources with low false positive rate and low resource requirement on routers.

2.3 ICMP Traceback

A traceback scheme utilizing the explicitly generated ICMP Traceback message was proposed in [25]. Each router samples the forwarding packets with a low probability (e.g. 1/20000) and sends a special ICMP message including the information like neighboring routers (forward and backward links) on the path to the destination and source along with the original (triggering) packet. Traceback packet also includes an authentication field which guards against spoofed traceback packets sent from attackers. This field can be null authentication, random strings or even HMACs. TTL is set to 255 for computing distance at the receiving end. During DDoS flooding attack, these ICMP traceback messages are used by the victim to reconstruct the path taken by the attacker. The schematic representation of the scheme is shown in Figure 5. The updated version of the previous iTrace (ICMP Traceback) scheme was proposed in [26]. iTrace scheme is considered as an industry standard by IETF. The time taken for path reconstruction by iTrace is minimized in ICMP Traceback with cumulative path (iTrace CP) [27]. This scheme is independent of the attack length. This scheme encodes the entire attack path information (i.e. contains the addresses of all the routers on the attack path) into minimal number of packets, thus minimizing the attack path construction time. This is achieved at the expense of minimal additional overhead in computation, storage and bandwidth. An enhancement to this scheme is suggested in Enhanced ICMP Traceback with Cumulative Path [28], which suggests the exponential increase in the probability of message generation with the distance in hops from the victim. The effectiveness of the scheme relies on selecting the appropriate value for the probability exponent which influences the traceback time for attack paths of different length. The iTrace scheme suffers a serious problem on the resource spent on generating the number of traceback packets which turns out to be neither useful nor informative during traceback and this issue is addressed in Intention-driven ICMP traceback [29] which enhances the probability of the router to generate useful trace messages. This is achieved by adding an additional intention bit to the iTrace message. A modification to Intention driven traceback is provided in [30] to create more effective iTrace packets to detect the origin of attack more accurately.

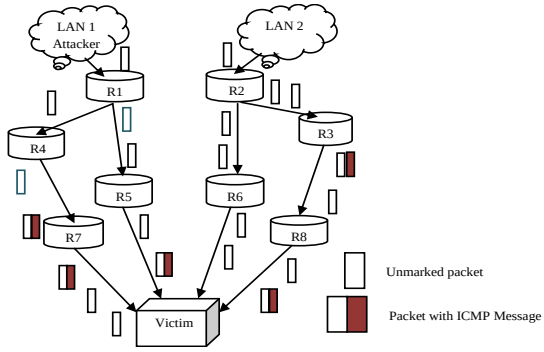

Figure 5
ICMP based Traceback

2.4 IP Logging

Logging scheme for IP traceback stores the information like packet's digest, signature, and fields of IP header on all or few routers which forward packets within the domain. It is shown in Figure 6. When an attack is detected, the victim requests the upstream router to gather information about attack packet. If the information is found, then the router is counted as a hop in the attack path and the process is repeated. The major challenges faced by this scheme is the overhead on the network and the storage requirement at core routers etc. Hash based IP traceback [31, 32] can trace even a single IP packet provided, the copy of the packet, its destination and approximate time of the packet's reception at the victim are available. A Source Path Isolation Engine (SPIE) is developed for this purpose, which stores the packet's digest. The memory requirement is minimized using Bloom filter. Bloom filter is a space efficient probabilistic data structure to test whether the given element (entity) is a member of a set. When the Bloom Filter returns a wrong value,there is a possiblity of false positive. The storage overhead at the routers is very high and the processing and the storage cost of it has been addressed in [33]. The storage requirement analysis reveals that with a packet size of 1000 bits, a duplex OC-192 link requires a computation of 60 million hash operations every second, and 44GB of storage space every hour, using the parameters suggested in [32] and the scheme [33] needs a computation of 8 million hash functions every second and requires a storage of 5.2GB to store the traffic for one hour. Another scheme for IP traceback with single packet is suggested in [34]. The disadvantage of false positive errors in traceback due to Bloom Filter is reduced in [35]. ID based Bloom Filter (IDBF) is used which requires ID table at every traceback enabled node. During Logging phase, ID table stores the node information (Node ID, Forwarder Address) in positions obtained on applying k hash functions to the payload. During Query phase, the most occurring value of Node ID is retrieved and reverted for traceback. Multiple

IDBFs are used on nodes nearer to the sink with high traffic load to avoid false positive errors closer to sink. This, in turn, consumes a lot of memory. The idea of packet logging is combined with the overlay network to improve traceback results by reducing the number of routers involved in traceback. Overlay network helps to minimize the number of hops during traceback and minimizes the router involvement. Intra-domain IP traceback using Overlay network is detailed in [7]. An OSPF-based Traceback System (OsTraS) supports partial and progressive deployment of the traceback system. This system suits even large network domains. In this approach, certain routers in the network are used as Traceback Agents (TBAs), which take part in the traceback process. The exchange of information is done using Link State Advertisement (LSAs). The traceback is done from the victim with the number of hops required across the overlay tree. Similarly, inter AS IP traceback (traceback across different ASes) using information obtained from BGP is proposed in [36]. It reduces false positive rate in traceback, at the cost of storage. To reduce storage, enhanced and specialized Bloom Filters are used. Though this minimizes the false positive rate, zero false positive rate has not yet been achieved. False negative results are also produced when the logged information is refreshed.

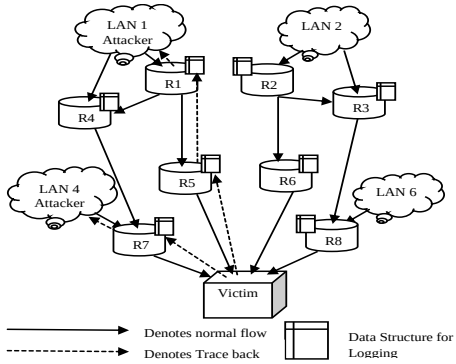

Figure 6
Logging Scheme

2.5 Hybrid Schemes

The idea of hybrid scheme combining marking and logging has been conceived to overcome the disadvantage of individual marking and logging schemes as stated above and a drastic improvement in traceback has been achieved. In [37], two hybrid schemes of IP traceback are proposed – Distributed Linked List Traceback (DLLT) and Probabilistic Pipeline Packet Marking (PPPM). The first scheme preserves the marking information at the core routers in a precise way such that it can be collected using a linked-list based approach. The second scheme aims at passing the IP addresses of the routers that were involved in marking particular

packets by stuffing them into the packets going to the same destination. This mechanism avoids the need for long-term storage at the core routers. This scheme can fail if IP marking field value is spoofed by the adversary but can be identified with the help of restrictions imposed on TTL field. When compared to IP logging schemes, processing and storage overhead at the routers are significantly minimized using this Hybrid Scheme. Single packet IP traceback [38] using hybrid scheme (referred as HIT) employs logging packet's digest at alternate routers in the attack path. Compared to SPIE,this scheme reduces the storage overhead and access time of digest tables. Over time newly emerging ideas and interface numbers of routers came into the picture. Instead of IP addresses or link information specified partially router interface numbers are marked. However, as the mark field size is limited, it still requires storage at the routers. Several techniques like Huffman codes [39], Modulo /Reverse modulo techniques like MRT [40] and MORE [41] have used router interface number rather than IP address. RIHT [42] recommends a traceback scheme that marks router's interface numbers and logs the interface numbers in the hash table when the mark field exceeds and produces zero false positive and false negative rate.

3 Evaluation of IP Traceback Techniques

This section evaluates a representative method in each of the category of IP Traceback techniques based on the following evalution metrics.

- Deployability
- Scalability
- Memory Requirement
- Router Processing Overhead
- Protection
- Paraeters needed for traceback
- Applicability on different types of attacks
- Prior knowledge of topology
- Accuracy
- Post Attack Analysis
- Attacker's Challenge Vs Scheme survival
- Router Involovement during traceback
- Number of bits overriden in IP header
- Number of Packets Required to Traceback.

Controlled flooding [9] is choosen as the representative method of Link Testing, PPM[11] is chosen as a representative method of Probablisitic Packet Marking and FDPM[24] is chosen under Deterministic Packet Marking, ITrace[25] represents ICMP based traceback technique, SPIE[32] is choosen as the representative method of Packet Logging and RIHT[42] represents Hybrid Traceback scheme.

3.1 Deployability

Deployability stands for the requirement of hardware or software installation on ISPs either partially or completely. An ideal scheme must have ease of installation on ISPs, without making much change to the existing network infratructure. For e.g., additional hardware to all ISP's for implementation of a methodology will be overhead with respect to this metric. Except ITrace all other traceback schemes require a change in the existing infrastructure to enable IP traceback because packet marking and logging is not presently supported by any of the routers.

3.2 Scalability

Scalability relates to the amount of additional configuration required on other devices needed to add a single device to the scheme. It also measures the ability of the scheme to adapt to increasing network size. The features that depend on configuration on other devices deteriorate scalability. An ideal scheme should be scalable and configuration of the devices should be totally independent of each other. As mentioned earlier marking and logging schemes require additional configuration at the routers,FDPM requires comparitively lesser configuration when compared to that of PPM,RIHT and SPIE because it requires additional configuration only at the border routers whereas the other schemes require additional configuration in the routers in the attack path.

3.3 Memory Requirement (Network/Victim)

An important metric of a traceback scheme is the amount of additional storage required either at the routers or at the dedicated traceback servers in the network, or at the victim. An ideal scheme should demand negligible or no additional storage on the network devices. ITrace and marking schemes does not require any storage at the routers whereas logging and hybrid scheme needs logging at the intermediate routers in the attack path. Using SPIE, a core router with 32 OC-192 links requires 23.4 GB[32] and RIHT requires a fixed storage of 320 KB[42] according to CAIDA dataset [43].

3.4 Router processing Overhead

Almost every traceback scheme requires processing at the routers. Processing overhead on routers is undesirable as it may result in degrading the performance of routers. Though processing occurs during traceback, it is expected to be relatively infrequent. An ideal scheme should have minimal or less processing overhead incurred on the network. Since Link testing involves every router in the

traceback process,it requires high computation at the routers in the attack path,FDPM and PPM require processing at the routers but it is relatively lesser compared to the logging based SPIE which involves every router and its neighbours in the computation. RIHT involves only the routers in the attack path with minimal arithematic computation.

3.5 Reliability

A high level protection is preferred in any traceback scheme. Protection refers to the ability of a traceback scheme to produce reliable traces with a limited number of network elements that have been challenged. An ideal scheme should act as if a device is not part of the scheme when the device becomes subverted. Schemes that are dependant on every router in the attack path fail to produce reliable result when any one of the device fails. SPIE and RIHT requires computation in every router in the attack path for tracing back to the attacker,hence they provide less reliable results if any of the router is inactive. Controlled Flooding,FDPM and PPM are more reliable compared to log based schemes and ITrace also produce more reliable results even if the intermediate routers are challenged.

3.6 Parameters Needed for Traceback

With recent advanced techniques on IP traceback, it is an important criterion to evaluate techniques based on the required paramenters to inititate the traceback process. Attack consists of flooding of attack packets along with normal packets. So traceback schemes were analysing the traffic pattern and they were in need of multiple packets. Attackers have become so clever that they have started to attack with a single packet these days. So tracing the attacker with a single packet is a desirable feature of IP traceback. RIHT and SPIE has single packet traceback capability,remaining schemes require multiple packets and link testing based scheme requires to analyse attack pattern also.

3.7 Applicability on Different Types of Attacks

This metric classifies the traceback technique based on the types of attack which it can handle. Attack could be classified into flooding based attacks and software exploit attack. In case of flooding based attack,the attacker generally pumps out a huge number of packets to the victim whereas in case of a software exploit attack,attacker generally sends only very few packets,for that matter even a single packet would be sufficient to bring down the entire server. Software exploit attack generally takes advantage of the vulnerabilities in the operating system in the victim machine and makes it incapicitated. Controlled Flooding, FDPM and PPM are dependant on multiple packets to identify the attack path hence they are not

suitable for identifyign the origin of Single packet attack,whereas SPIE and RIHT has single packet traceback capablity and hence they can be used to trace back software exploit attack as well as flooding based attack.

3.8 Prior Knowledge of Network Topology

A few schemes assume that they are aware of the topology in advance. In this changing environment one cannot always rely on a topology map. So this metric is used to analyse if the scheme requires prior knowledge about the topology. Except link testing based schemes,remaining schemes do not require prior knowledge of the topology.

3.9 Accuracy

Accuracy is the important metric which measures the precision of the scheme. False positive and False negative have to be less in an ideal traceback scheme. False positive is tracing a legitimate node as an attacker node. False negative is missing to identify the attacker node. So the traceback scheme must be able to trace most of the attackers. Since link testing based schemes are depending on the flow of the packets,they cannot diferentiate flash crowd from an attack. Since a single IP packet cannnot accommodate the entire IP address of the router, FDPM and PPM rely on multiple packets to store the IP address,which may also result in false positive. SPIE uses Bloom filter to log the hash digest. Overwriting the log may lead to false positive. RIHT claims zero false positive and false negative.

3.10 Post Attack Analysis

A few traceback schemes are capable of tracing the attacker even after the attack is stopped whereas some schemes require the attack to be alive till the traceback is completed. A traceback scheme should be able to detect the attack whether it is alive or not because the attack duration cannot be predicted. This metric evaluates whether the traceback scheme supports post attack analysis or not. Except link testing based schemes remaining schemes are capable of supporting post mortem analysis because they store the audit trail either in the IP or ICMP packets or at the routers and they are not dependant on active packet flow from the attackers.

3.11 Attacker's Challenge to the Scheme

This metric evaluates how well the proposed scheme sustains the attacker, if the attacker is well aware of the scheme. If the attacker is aware of the controlled flooding scheme,attacker can very well generate the attack with the signature which matches the normal traffic flow and mislead the traceback scheme.

Similarly if the attack is aware of DPM or PPM scheme also he can generate packets with false marking and mislead the scheme. RIHT and SPIE is also vulnerable whereas ITrace provides authentication and hence the attacker cannot easily generate false ICMP packets.

3.12 Router Involvement during Traceback

Most of the traceback schemes rely on the router to send the trace information when the packet is moving towards the victim. This is an overhead which would affect the router's performance. So its expected that atleast during the second phase, that is, during the reconstruction of attack path or identifying the attacker, the router should not be bothered. This metric evaluates whether the scheme requires router involvement during traceback or not. Link testing based schemes require the router support during traceback. FDPM,PPM and ITrace can traceback the attacker from the victim itself. They do not require router support in traceback. The number of routers (N_R) involved in traceback in SPIE is given by (1)

$$N_R = (n - 1)h \tag{1}$$

Where 'n' is the number of routers connected to the router in the attack path. 'h' is the total number of hops in the attack path. SPIE queries all the nieghbouring routers of the routers in the attack path except the downstream router. RIHT involves only the routers in the attack path. The number of routers involved in RIHT traceback process is given by (2)

$$N_R = h \tag{2}$$

According to CAIDA topology dataset[43] the average degree of a router is 3.5 and the average path length is 16. So assuming that the total number of neighbour routers is 4 the router involvement in SPIE and RIHT is given in Figure 7.

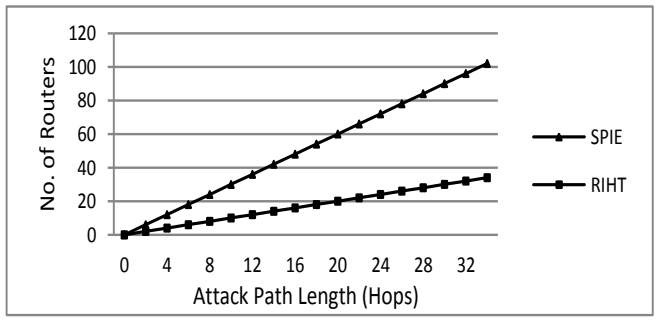

Figure 7
Router Involvement in Traceback Process

3.13 Number of Bits Overriden in IP Header

IP header as such do not have a provision to store the audit trails. So researchers started using the rarely used fields of an IP header in storing the traceback information. Most of the traceback schemes commonly override the Identification field in the IP header. Overriding the Identification field would affect the fragmented traffic. Likewise few schemes override the fragment offset and flag fields along with the Identification field to accomodate the trace information. Few schemes override TOS field. The lesser the number of bits is overriden, the better the scheme is. Figure 8 depicts the number of bits overriden by IP header by each of the marking based schemes. Link testing based schemes, log based schemes and ITrace avoid overriding the IP header,whereas marking based schemes unanimously override the IP header fields.

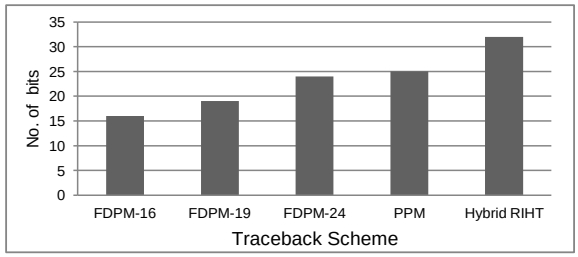

Figure 8
Number of Bits Overriden in different Schemes

3.14 Number of Packets Required to Traceback

Few schemes are capable of tracing back the attacker with the single packet. Few schemes rely on multiple packets because the entire audit information cannot be stored in a single packet. Schemes that are capable of initiating the traceback process with minimal number of packets have lesser false positives and can traceback faster compared to schemes that rely on multiple packets. SPIE,RIHT can initiate the traceback with a single packet. The expected number of packets to reconstruct the attack path in an ICMP based trace back is given by (3)

$$mH_m / q \tag{3}$$

where 'm' is the number of attackers, 'H_m' is the m^{th} harmonic number and 'q' is the probability at which the ICMP packet was generated. The expected number of packets using DPM is given by (4).

$$P = 1 - 0.5^r \tag{4}$$

where '*r*' is the number of packets needed to identify one attacker, '*P*' is the probability of identifying one IP address. Figure 9 depicts the expected number of packets to traceback one attacker in each of the schemes.

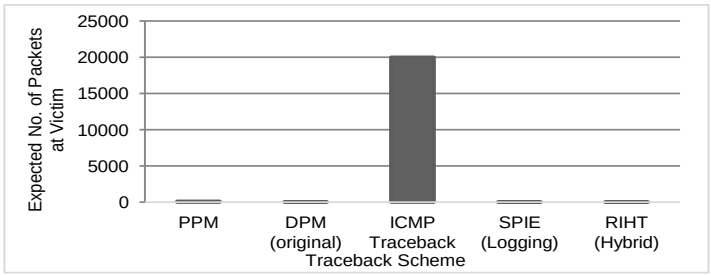

Figure 9
Expected Number of Packets at the Victim

Table 1 shows a high level qualitative comparison on various IP Traceback schemes based on the evaluation metrics.

Table 1
Comparison of various trace back schemes against the evaluation metrics

Evaluation Metrics	Link Testing	Packet Marking		Packet Logging	ICMP Traceback	Hybrid Scheme
		PPM	DPM			
Deployability	Fair	Fair	Fair	Poor.Huge memory requiremet	Good	Fair
Scalability	Poor	Poor	Fair	Fair	Good	Fair
Memory Requirement (Network)	Not Required	Not Required	Not Required	Very High	Not Required	Low
Memory Requirement (Victim)	Not Required	Very High	Medium	Not Required	Medium	Not Required
Router Processing overhead	High	Medium	Medium	High	Low	Low
Reliability	Good	Good	Good	Poor	Good and Practically feasible	Poor
Parameter needed for traceback	Attack pattern and large number of packets	Large No. of packet	Minimum number of packets compared to PPM	One Packet	No.of ICMP messages and huge number of attack packets.	One Packet

Applicability on different types of attacks	DoS	DoS/ DDoS flooding attacks	DoS/DDoS flooding attacks	DoS/DDoS flooding attacks	DoS/DDoS network layer attacks	DoS/ DDoS flooding attacks
Prior knowledge of different topology	Needed	Not needed. Faster traceback and low false positive if known	Not needed. Faster traceback and low false positive if known	Not needed	Not Needed	Not Needed
Accuracy	Medium	Medium, huge false positive rate in case of DDoS attack	Good	Medium with high false positive and false negative	Good for less number of attackers	High. less false positive and false negative rate.
Post Attack Analysis	Not Possible	Possible	Possible	Possible	Possible	Possible
Attacker challenge vs. Scheme's survival	Poor	Poor	Poor	Poor	High	Poor
Router Involvement during Traceback	High	Nil	Nil	High	Nil	High

Conclusion and Future Work

This survey paper thus provides an overview of the evolution of existing IP traceback schemes. The study shows that the focus on traceback scheme has moved from the quick traceback from the victim to the quick detection of attack before the victim is affected as most of the DDoS attacks take place from the stepping stones (compromised intermediate hosts). Traceback schemes using Watermarking technique, Information metrics like entropy, divergence and distance metric are gaining momentum and a brief study of these techniques will be provided in near future.

Acknowledgement

This work was practically supported by the Smart and Secure Environment project. The project is funded by the National Technical Research Organization (NTRO). This project is funded by the Government of India.

References

[1] S. Savage, D. Wetherall, A. R. Karlin, T. E. Anderson: Network Support for IP Traceback, IEEE/ACM Transactions on Networking, Vol. 9, No. 3, 2001, pp. 226-237

[2] J. Mirkovic, J. Martin, P. Reiher: A Taxonomy of DDoS Attacks and DDoS Defense Mechanisms, Computer Communication Review, Vol. 34, No. 2, 2004, pp. 39-53

[3] D. Moore, C. Shannon, D. Brown, G. Voelker, S. Savage: Inferring Internet Denial-of-Service Activity, ACM Transactions on Computer Systems, Vol. 42, No. 2, 2006, pp.115-139

[4] Hakem Beitollahi, Geert Deconinck: Analyzing Well-known Countermeasures against Distributed Denial of Service Attacks, Computer Comm., Vol. 35, 2012, pp. 1312-1332

[5] R. K. C. Chang: Defending Against Flooding-based Distributed Denial-of-Service Attacks: A Tutorial, IEEE Commun. Mag., Vol. 40, No. 10, 2002, pp. 42-51

[6] Belenky, N. Ansari: On IP Traceback, IEEE Commun. Mag. Vol. 41, No. 7, 2003, pp. 142-153

[7] A. Castelucio, A. T. A.Gomes, A. Ziviani and R. M. Salles: Intra-Domain IP Traceback using OSPF, Computer Comm., Vol. 35, 2012, pp. 554-564

[8] R. Stone: Centertrack: An IP Overlay Network for Tracking DoS Floods, Proceedings of the 9th conference on USENIX Security Symposium, Berkeley, USA, 2000, pp. 199-212

[9] H. Burch, B. Cheswick: Tracing Anonymous Packets to Their Approximate Source, Proceedings of the 14th USENIX Conference on System Administration, New Orleans, LA, USA, 2000, pp. 319-328

[10] J. Ioannidis and S. M. Bellovin: Implementing Pushback: Router-based defense against DDoS attacks, in Proc. Network and Distributed System Security Symp., 2002

[11] M. T. Goodrich: Probablistic Packet Marking for Large Scale IP Traceback, IEEE/ACM Trans. Networking, Vol. 16, No. 1, Feb 2008, pp. 15-24

[12] S. Savage, D. Wetherall, A. R. Karlin, and T. Anderson: Practical Network Support for IP Traceback, in Proc. ACM SIGCOMM, 2000, pp. 295-306

[13] D. Song and A. Perrig: Advanced and Authenticated Marking Schemes for IP Traceback, in Proc. IEEE INFOCOM, 2001, pp. 878-886

[14] T. K. T. Law, D. K. Y. Yau, and J. C. S. Lui: You Can Run, But You Can't Hide: An Effective Statistical Methodology to Trace Back DDOS Attackers, IEEE Trans. Parallel Distrib. Syst., Vol. 16, No. 9, Sep. 2005, pp. 799-813

[15] Yaar, A. Perrig, and D. Song: FIT: Fast Internet Traceback, Proc. IEEE INFOCOM, 2005, pp. 1395-1406

[16] D. Dean, M. Franklin, and A. Stubblefield: An Algebraic Approach to IP Traceback, in Proc. Network and Distributed System Security Symp. (NDSS), 2001, pp. 3-12

[17] Belenky and N. Ansari: IP Traceback with Deterministic Packet Marking, IEEE Comm. Letters, Vol. 7, No. 4, 2003, pp. 162-164

[18] Belenky and N. Ansari: Tracing Multiple Attackers with Deterministic Packet Marking (DPM), in Proc. IEEE PACRIM'03, Victoria, BC, Canada, Aug. 2003, pp. 49-52

[19] Belenky and N. Ansari: On Deterministic Packet Marking, Computer Networks, Vol. 51, No. 10, 2007, pp. 2677-2700

[20] G. Jin and J. Yang: Deterministic Packet Marking Based on Redundant Decomposition for IP Traceback, IEEE Comm. Letters, Vol. 10, No. 3, 2006, pp. 204-206

[21] Wang Xiao-jing, Wang Xiao-yin: Topology Assisted Deterministic Packet Marking for IP Traceback, The Journal of China Universities of Posts and Telecommunications, Vol. 17, No. 2, April 2010, pp. 116-121

[22] Y. Kim, J. Y. Jo, and F. L. Merat: Defeating Distributed Denial-of-Service Attack with Deterministic Bit Marking, Proc. IEEE Global Telecomm. Conf. (GLOBECOM '03), 2003, pp. 1363-1367

[23] Y. Xiang, W. Zhou, and J. Rough: Trace IP Packets by Flexible Deterministic Packet Marking (FDPM), Proc. IEEE Int'l Workshop IP Operations and Management (IPOM '04), 2004, pp. 246-252

[24] Y.Xiang, W.Zhou and M.Gu: Flexible Deterministic Packet Marking: An IP Traceback System to Find the Real Source of Attacks, IEEE Trans. Parallel and Distributed System, Vol. 20, No. 4, April 2009, pp. 567- 580

[25] S. M. Bellovin: ICMP Traceback Messages, Internet Draft: draft-bellovin-itrace-00.txt, 2000

[26] S. M. Bellovin, M. D. Leech, and T. Taylor: ICMP Traceback Messages, Internet Draft: Draft-Ietf-Itrace-04.Txt, Feb. 2003

[27] Henry C. J. Lee, Vrizlynn L. L. Thing, Yi Xu, Miao Ma: ICMP Traceback with Cumulative Path, an Efficient Solution for IP Traceback, International Conference on Information and Communications Security, Oct. 2003

[28] V. L. L. Thing , H. C. J. Lee , M. Sloman and J. Zhou: Enhanced ICMP Traceback with Cumulative Path, 61st IEEE Veh. Technol. Conf., 2005

[29] Felix, W. (2001): On Design and Evaluation of Intention-Driven ICMP traceback, Proc. IEEE Int'l Conf. Computer Comm. and Networks, IEEE CS Press, 2001, pp. 159-165

[30] Izaddoost, M. Othman, and M. F. A. Rasid: Accurate ICMP Traceback Model under DoS/DDoS Attack, in Proceedings of the 15th International Conference on Advanced Computing and Communications, 2007

[31] A.C. Snoeren et al.: Hash-based IP Traceback, Proc. ACM SIGCOMM, 2001

[32] A. C. Snoeren et al.: Single-Packet IP Traceback, IEEE/ACM Trans. Networking, Vol. 10, No. 6, Dec. 2002, pp. 721-734

[33] J. Li et al.: Large-Scale IP Traceback in High-Speed Internet: Practical Techniques and Theoretical Foundation, Proc. IEEE Symp. Security and Privacy (S&P '04), 2004, pp. 115-129

[34] T. Baba and S. Matsuda: Tracing Network Attacks to Their Sources, IEEE Internet Computing, Vol. 6, No. 3, 2002, pp. 20-26

[35] M. S. Siddiqui and S. O. Amin, C. S. Hong: Hop by Hop Traceback in Wireless Sensor Networks, IEEE Comm. Letters, Vol. 16, No. 2, Feb 2012, pp. 242-245

[36] A.Castelucio, A.Ziviani and R. M.Salles: An AS Overlay Network for IP Traceback, IEEE Network, Jan/Feb 2009, pp. 36-41

[37] B.Al-Duwari and M. Govindarasu: Novel Hybrid Schemes Employing Packet Marking and Logging for IP Traceback, IEEE Trans. Parallel Distributed Syst., Vol. 17, No. 5, May 2006, pp. 403-418

[38] Gong and K. Sarac:A More Practical Approach for Single-Packet IP Traceback Using Packet Logging and Marking, IEEE Trans. Parallel and Distributed Systems, Vol. 19, No. 10, Oct. 2008, pp. 1310-1324

[39] K. H. Choi and H. K. Dai: A Marking Scheme using Huffman Codes for IP Traceback, in Proc. 7th Int. Symp. Parallel Architectures, Algorithms Networks (SPAN'04), Hong Kong, China, May 2004, pp. 421-428

[40] S. Malliga and A. Tamilarasi: A Proposal for New Marking Scheme with its Performance Evaluation for IP Traceback, WSEAS Trans. Computer Res., Vol. 3, No. 4, Apr. 2008, pp. 259-272

[41] S. Malliga and A. Tamilarasi: A Hybrid Scheme using Packet Marking and Logging for IP traceback, Int. J. Internet Protocol Technol., Vol. 5, No. 1/2, Apr. 2010, pp. 81-91

[42] M. H. yang, M. C. Yang: RIHT: A Novel Hybrid IP Traceback Scheme, IEEE Trans. Information Forensics and Security, Vol. 7, No. 2, April 2012, pp. 789-797

[43] CAIDA's Skitter Project CAIDA, 2010 [Online] Available:http://www.caida.org/tools/skitter/