Survey of Traceback Methods

Bhoomi Patel, M.E., CSE Department, Parul Institute of Engineering & Technology, Vadodara, India
Sumitra Menaria, Assistant Professor, CSE Department, Parul Institute of Engineering & Technology, Vadodara, Gujarat, India

Abstract
Now days, internet is expanding day by day. Everyday some new threats are detected and some new attacks like Dos and DDoS are launched. So the problem of identifying such kinds of threats is very difficult in network security. It is very difficult to distinguish between normal traffic and DDoS traffic because DoS hide their actual identity. In this situation, it is necessary for the victim to identify the exact source or the origin of packet by using traceback method. It means that it is important to find the router, from which, the packet was entered into network so that attack rate can also be reduced. This paper presents different traceback methods to trace the packets or to identify the origin of a packets on network

Keywords: Dos, DDoS, Traceback, Threat, Packet marking

Introduction
Classification Of Traceback Methods
There are two major categories of traceback methods[1]: preventive and reactive. The main aim of preventive method is to prevent DoS attack. To prevent DoS attack, preventive method takes some precautionary steps. A wide range of solutions has been proposed, however, this problem still remains as open. The main aim of reactive method is to identifying the source of attacks. This is very important because attackers spoof their addresses. So, techniques are needed to trace back to the source of attack.

Different Traceback Methods
A. Preventive Methods
1. Ingress Filtering
B. Reactive Methods
1. Link Testing
   a) Input Debugging
   b) Controlled Flooding
2. Logging
3. ICMP Traceback
4. Packet Marking algorithm
   a. PPM
   b. DPM

Survey Of Different Traceback Methods
Preventive Methods
1. Ingress Filtering
This technique was first introduced by savage[2]. One way to address the problem of anonymous attacks is to eliminate the ability to forge source addresses. For this purpose, ingress filtering method is used, to configure routers to block packets that arrive with illegitimate source addresses [3]. It must require that all routers have sufficient power to determine source address of every packet and must have sufficient amount of knowledge to differentiate between legitimate and illegitimate addresses. This technique is most feasible where the load of traffic is relatively low in network. But when the traffic is aggregated from multiple ISPs into transits network, it can not exactly determine if a packet arriving on a particular interface has a legal source address. If this technique can be used widely then it can improve the Internet’s robustness to denial-of-service attacks. But it is prudent to assume that such a system will never be full proof and so that reactive traceback technologies have more importance.

1. REACTIVE METHODS
   1. Link Testing
   Input Debugging
   In this technique, particular packets can be filtered at some egress port and can determine that the packets arrived from which ingress port[3]. First, the victim must recognized that it is being attacked and generate an attack signature that describe a common feature contained in all attack packets. This signature is sent by victim to network operator, frequently via telephone, which can then install input debugging filter on egress port of victim. This process is repeated recursively until the original ingress port is identified. Hence in this way which upstream router originated the traffic can be identified.

   Disadvantage: Management overhead, Communicating and co-ordinating with network operator require more time and attention, If network operator does not have appropriate technical skills then traceback cannot be complete.

   Controlled Flooding
   Only DoS attack can be handled by this technique. It completely relies on the fact that, during DoS attack, the links of the attack path should be heavily loaded. In this technique, the drops in the rates of attack packets should be observed by carefully measuring incoming traffic to the attacked system and loading the links of suspected path even more. Until the source of the attack is identi-
fied, this process can be repeated. So if increasing the load of a particular link results in the drop in the rate of the attack traffic, then this link is on the attack path. Once the DoS attack is identified by victim, equipment to measure the load of link is used to generate traffic on network and then only the traceback will begins.

The concept is illustrated in figure 1. First of all, router closest to the victim, are determined and in this case it is router R12. Then we have to consider all the routers, which are connected with R12. So, consider routers R9, R10, R11. Now consider above all three routers link with R12 and observe packet drop rate. So first, short burst of traffic is generated from R11 hopping that the rate of attacking packet, to the victim will drop but here in this case, it did not. So, the link between R11 and R12 and all the paths which may use it, are excluded from the set of possible path. Now load the link between R10 and R12, but here also no drop in the rate of packets was observed and so this link is also excluded. But when we load the link between R9 and R12, the desired drop in the rate of packet is observed. So we conclude that the link between R9 and R12 belongs to the attack path. This process is repeated until source of attack is identified.

The way the links are suggested to be loaded is by using the chargen service on each router [5].

Advantage: Does not rely on ISP co-operation.
Disadvantage: Most routers have chargen disabled. In fact it comes disabled by default now on most of routers. It can handle only DoS attack so traceback is limited to one attacking stream only. It cannot trace the attack when it is over [4].

2. Logging or SPIE (Single Packet IP Traceback)
It is a technique to log packets at routers in the network and then determine the network paths, which packets traversed, using data extraction technique [6]. There are major three units of it: DGA, SCRA, STM. The functionality of DGA is to record packet digest. Every router captures partial packet information of every packet which will be used in future to determine if packet passed through that router or not. Here network is logically divided into regions. SCRA (SPIE Collection and Reduction Agent) is responsible for a particular region. In every region, SCAR connect to all DGA’s and can collect necessary information from DGA [5]. STM (SPIE Traceback Manager) is a central management unit [5] and whole SPIE system is controlled by it. The request for tracing an IP packet is first received by STM then it dispatch the request to SCAR and then collect information from all related SCAR to make a complete attack path. When the packets traverse through network, the digest of packets will be stored in DGA. To produce several digest, constant field from IP header and 8 bytes of payload of each packet, are hased by some hash function. There is bloom filter, in which, digest are stored. When 70% of given bloom filter is stored then another one is used.

The concept of SPIE is illustrated in figure 2. When victim send request to STM (step-1) then STM send request to appropriate SCAR (step-2). So all DGA send copies of digest to it’s SCAR as a reply for their request. Then SCAR perform analysis of tables, so that SCAR can know that, which routers in the region, if any, forwarded the packet. Now SCAR can reconstruct the path along which the packet traversed through the region and reports it to STM (step-4). Based on these information, STM can reconstruct the whole path through network.

Disadvantage: ISP involvement is very high because ISP has to purchase equipment for SCAR and at least one STM. More processing overhead occurred because packet digest is stored in bloom filter for every packet. If SCAR may unlink with rest of the network then the whole segment of the path can be incorrect and not generate correct traceback.

3. ICMP (Internet Control Message Protocol) Traceback
In this method, the packets can be marked by each routers on network statistically (1 in every 20,000 packets recommended) and produce iTrace or ICMP message directed to the same destination as shown in figure 3 [5]. Next hop information, previous hop information and information about timestamp is shown by iTrace message. To identify the actual path of attack, initially
TTL field’s value is set to 255. TTL field is used to sort the addresses and to reconstruct the attack path. In this method, the traceback information is not completely in-band because routers on the network generate new packet with iTrace message. If the victim is under DDoS attack and so the volume of packets going to victim, will be large then victim will eventually get all the address of the routers on the attack path, which implement iTrace. But the major problem with this method is that, if victim undergoes a major DDoS attack then the chance of receiving useful iTrace message is very less [7]. To resolve the above problem, value or weight has to be attached with every iTrace generated. This value depends on frequency of iTrace being sent to victim, time since the attack has begun and distance from the victim. Icaddie ICMP method is used based upon the number of packets, after which to generate iTrace message [8]. For this purpose, timer is placed with each router to indicate, how long it has not received a traceback message. So, ICMP messages are sent to find the source of forged packets [9].

Fig.3 : ICMP traceback

**Advantage:**
- Good scalability.

**Disadvantage:**
- Generate additional traffic, ICMP messages could be filtered out by some routers, ISP involvement, Number of packets for path reconstruction is very large, Require large amount of memory, No extra mechanism to prevent attackers to generate fake iTrace, Ability to handle major DDoS attack is poor.

4. Packet Marking Algorithm

**Probabilistic Packet Marking (PPM)**

In this technique, all routers in the network select randomly some of the packets with constant probability and then mark those selected packets by their own IP address or portion of IP address [10]. Whenever victim receive large number of marked packets then after it can start to reconstruct the attack path. The packets, which are already marked by previous routers, can also be marked by its subsequent routers. So, those routers which are far away from the victim, has more chance of overridden its mark by its subsequent routers. Similarly, routers which are nearest to victim have more chance to deliver its mark to victim.

There are major two fields of this method: address and distance. Both of these fields are stored in Identification field of IP header. Which packets have to be marked-this decision is made by each routers independently. Once a router decides to mark a packet, it write its own IP address into address field and write 0 in distance field. So, if some routers receive a packet with distance field is 0, means that packet is already marked by previous router. In such case, routers would x-or its own IP address with previous IP address. If routers do not want to mark the packet then it just increment the value of distance field. For path reconstruction, victim locates the closest router to itself by looking for the packet, where distance=0,1,2,…… respectively. The victim continues this process until locating the router that is the most far away.

**Advantage:**
- Zero bandwidth overhead because all marking information is stored in packet IP header and it is completely inband.

**Disadvantage:**
- Computational overhead, Memory overhead, Number of required packets for traceback, ISP involvement.

**Deterministic Packet Marking (DPM)**

This approach was first introduced by Belenky and Ansari [11]. It is a packet marking algorithm. In this technique, each packet will be marked whenever it enters into network and this marks cannot be changed through whole network. The interface, which is closest to the source of the packet on edge ingress router, can only mark the packets. The interface makes a distinction between incoming and outgoing packets. In this method, only incoming packets will be marked and outgoing packets can not be marked so that there is no chance to overwrite the mark by edge egress router on victim side. To mark the packets, identification fields of IP header is used, which is of 17 bits. But IP address is of total 32 bits and we have available only 17 bits in one packet. So, we need at least two packets to send complete IP address. For this purpose, IP address is divided into two parts: one part contains 0-15 bits of IP address and another part contains 16-31 bits of IP address. From total of 17 bits, 16 bits are used for portion of IP address and remaining one bit is Reserved for Flag. So, if a packet consist first part of IP address (0-15 bits) then the value of flag bit is set to 0 and if packet consist another part of IP address (16-31 bits) then the value of flag bit is set to...
1. So all the routers on the network can identify that, the packet consist which portion of IP address, depending upon the value of RF. Here the marking is performed in deterministic way. So if an attacker attempt to spoof the mark, to deceive the victim, then also this spoof marked will be overwritten with a correct mark by very first router that the packet traverse.

**Table 1. Comparison table**

<table>
<thead>
<tr>
<th>Method</th>
<th>Management overhead</th>
<th>Network overhead</th>
<th>Router overhead</th>
<th>Post-mortem capability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ingress filtering</td>
<td>Moderate</td>
<td>Low</td>
<td>Moderate</td>
<td>No</td>
</tr>
<tr>
<td>Input debugging</td>
<td>High</td>
<td>Low</td>
<td>High</td>
<td>No</td>
</tr>
<tr>
<td>Controlled flooding</td>
<td>Low</td>
<td>High</td>
<td>Low</td>
<td>No</td>
</tr>
<tr>
<td>Logging</td>
<td>High</td>
<td>Low</td>
<td>High</td>
<td>Excellent</td>
</tr>
<tr>
<td>ICMP traceback</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Excellent</td>
</tr>
<tr>
<td>Packet Marking</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Excellent</td>
</tr>
</tbody>
</table>

**Advantage**: No bandwidth overhead, Scalable, Not reveal on internal topologies of ISP.

**Disadvantage**: May increase false positive rate

**Conclusion**

This paper describe classification of major two kinds of traceback methods. It also describe detailed survey of different traceback methods. The comparison of all above methods based on management overhead, network overhead, router overhead and post-mortem capability shown in table 1.

**References**


