Abstract

Facing the daunting threats from ever-increasing cyber attacks, especially Denial of Service (DoS) attacks, a promising approach to deterring attackers is to detect, record and analyze the attacks, and finally prosecute the attackers using the evidence collected through the above process. IP traceback, the key to this process, is the identification of the source of packets sent across the network, performed in an attempt to hold attackers accountable. In this paper, we first identify and state the traceback problem, and analyze the current weaknesses in the Internet environment and other issues that contribute to the difficulties in tracking and tracing attacks. We then survey the current start-of-the-art traceback techniques. Issues of IPv6 relating to traceback and IP traceback in Wireless Ad-hoc networks are also discussed. Finally, we summarize and propose a list of features that are desirable for an efficient and effective traceback technique.

Keywords: security, protection.

1 Introduction

Network computing has become an integral part of our daily life with the emergence of a global digital society. The advances in network technology have dramatically boosted our productivity and even the way we live. On the other hand, this widespread reliance on networked computing makes attacks launched over the Internet more devastating than ever before. Meanwhile, the Internet is becoming readily accessible to the general public as well as malicious attackers, while numerous government departments, business corporations and educational institutions etc. are all rushing to migrate their data and services to the Internet. Many of these data and services contain a wealth of sensitive information which lures cyber attackers from all over the world. A direct result of this is a rapidly increasing amount of attacks occurring over the Internet every year. For instance, the number of computer security incidents handled by the CERT Coordination Center (CERT/CC) [1] has grown from 6 in 1988 to 137,529 in 2003.

Among the countless cyber attacks, Denial-of-Service (DoS) attacks, and their variation – Distributed DoS attacks – are some of the hardest attacks to defend against. A DoS attack is designed to flood a victim’s connection with an overwhelming amount of attacking packets in order to render the victim’s services inaccessible to its legitimate users. The DDoS attack is a distributed variation of a DoS attack, and is even more difficult to deal with. In DDoS attacks, packets are launched from multiple, distributed sources coordinated by a single hijacking entity. DoS attacks are among the hardest security problems to address because they are simple to implement, difficult to prevent, and very difficult to trace. Therefore, DoS attacks pose an enormous challenge to the safety of the Internet.

Unfortunately, mechanisms for dealing with DoS attacks, and especially DDoS variants, have been far from satisfactory. Most work in this field is aimed toward solving the problem by tolerating attacks [2, 3, 4, 5, 6]. However, this approach neither eliminates the problem, nor thwarts future attack intentions.

Another promising approach in the battle against cyber threats, particularly DoS attacks, is to detect, record and analyze attacks, and finally to prosecute attackers using the evidence collected through the above process. Computer and network forensics is such an approach, and an emerging discipline that is becoming more important as society recognizes the seriousness of cyber attacks. This method involves capturing, recording, and analysis of network events in order to discover the source of security attacks or other problem incidents.
It attempts to deter hackers from attacking a system, stop on-going attacks, and it searches for evidence after an attack has occurred, finally prosecuting attackers if laws have been violated.

IP traceback is an indispensable part of network forensics. It is the process used to identify the source of packets sent across the network in an attempt to make attackers accountable. The success of network forensics is largely dependent upon whether we have efficient and effective traceback techniques.

The remainder of this paper is organized as follows: In Section 2, we analyze the difficulties involved with traceback and the shortfalls in today’s Internet environment that contribute to these obstacles. In Section 3, we survey current state-of-the-art traceback techniques and analyze the advantages and disadvantages of each of these approaches. In section 4, the issues of IPv6 relating to traceback are discussed. IP traceback in Wireless Ad-hoc networks is discussed in section 5. Finally, in section 6, we summarize and propose a list of desirable features that an efficient and effective traceback technique should have.

2 IP Traceback: a challenging task:

Although IP traceback is a theoretically perfect solution for stopping DoS attacks, and deterring malicious attackers by holding them accountable, determining the true source from which attack traffic is generated is surprisingly difficult. This impediment is in part due to flaws in the design of TCP/IP protocol and growing complexity of today’s Internet as well as the requirement of global efforts and commitment in tracing and tracking attack packets.

2.1 The Internet was never designed as a secure environment.

Development of the Internet began in 1969 as a research project sponsored by the Advanced Research Projects Agency (ARPA) of the Department of Defense (DoD). Originally, the Internet was only accessible to university researchers and was designed to explore the potential ways of using of networking to facilitate collaboration in research. The Internet environment was assumed to be benign, trustworthy and secure. There was no concern with regard to the possibility of cyber attacks by the Internet’s end users because the end users were mainly trustworthy researchers. Therefore, the original Internet protocols are presenting numerous shortcomings while striving to meet today’s networking needs.

In his 1985 report on TCP/IP weaknesses, Morris writes:

“The weakness in this scheme [the Internet Protocol] is that the source host itself fills in the IP source host id, and there is no provision in ... TCP/IP to discover the true origin of a packet.” [7]

The foundation of today’s Internet communications continues to rely upon the TCP/IP protocol suite designed in 1980s. The IP protocol allows the source host itself to fill in the source address, and it has no way to verify that the IP address in the source field is the sender’s true address – no entity in an IP network is officially responsible for ensuring the source address is correct. The network routing infrastructure is stateless and based largely on destination addresses. So, one machine is capable of masquerading as another machine or, even worse, as a router for a period of time. Robert Morris used this weakness, along with the predictability of the sequence number, in 1985 to show that any host can act as a trusted host, thus carrying out secure tasks [7]. Therefore, the weaknesses in the TCP/IP protocol and the stateless, anonymous nature of the Internet allow source IP addresses to be easily spoofed. Consequently, source addresses in IP packets cannot be considered reliable.

Most attackers take advantage of additional techniques to ensure that their identity cannot be discovered through the packet source address. One such technique for masking the source host is by “laundering” attack packets through intermediate hosts. Consequently, exposed IP source addresses could only be the ones of laundering hosts or compromised network entities.

2.2 Internet Complexity

In the early days of the Internet, only a small number of systems were attached to the network, and these machines were typically administered by university researchers with a reasonable amount of skill in configuring and maintaining at least a basic level of system security. In this booming Internet era, millions of networks are interconnected, and thousands of new devices are joining this noisy electronic community every second. These networked entities are connected to the Internet through various mechanisms and bring with them numerous vulnerabilities, as well as increasing complexity. Many of these end-systems are networked via residential broadband connections, and are operated by general users with little or no knowledge regarding network security. These vulnerable systems can be easily exploited by intelligent hackers to hide their real locations.
Furthermore, many advances in network technologies have even deteriorated the tracing and tracking environment. For instance, spoofed source addresses are legitimately used by network address translators (NATs), Mobile IP, and various unidirectional link technologies such as hybrid satellite architectures.

### 2.3 IP Traceback: global efforts and commitments.

The emergence of the Internet brings the world into a digital society which is beyond the governance of any single country. Today, spoofed attacking packets can come from anywhere in this cyber world and be routed through multiple countries before reaching their final victims. Therefore, tracing and tracking these packets often require the efforts and commitments from multiple involved countries. However, unlike well-organized world trade treaties among counties, there are no global commitments regarding IP traceback. Furthermore, different countries may have different legal issues regarding IP traceback. A legal traceback technique deployed in one country may violate another country’s protection of privacy. Consequently, in reality, IP traceback is often limited in the victim’s own country or a few cooperative countries.

Besides the above difficulties, there are a number of additional and significant challenges and issues pertaining to the construction of a successful tracing system, including determining which packets to trace, maintaining privacy (a tracing system should not adversely impact the privacy of legitimate users), and minimizing overhead (both in router time spent in tracking other than forwarding packets, and in storage used to keep information).

Even though IP traceback is a daunting challenge, courageous researchers continue to seek out solutions. During recent years, numerous traceback mechanisms have been proposed or developed. In the next section, we will survey these state-of-the-art traceback techniques.

### 3 State-of-the-art IP Traceback techniques

Several techniques and approaches are currently being proposed or developed to trace IP packets to their origins and reduce the anonymity afforded by IP spoofing. These techniques can be mainly classified into the following six categories: ingress filtering, link testing, logging, ICMP traceback, IP packet marking and IP Sec approach. In the following sections, we will discuss the strengths and weaknesses of each of the above six techniques.

#### 3.1 Ingress Filtering

In essence, ingress filtering [8] is not an approach to trace back IP addresses, but rather it addresses this problem indirectly by eliminating the ability to forge source addresses. Ingress filtering is a method by which routers are configured to block packets that arrive with illegitimate source IP addresses. Consequently, this approach requires routers with adequate processing power to examine the source address of every packet and sufficient knowledge to distinguish between legitimate and illegitimate addresses. Therefore ingress filtering is most feasible in customer networks or at the border of Internet Service Providers where address ownership is relatively unambiguous and traffic load is low [9,10]. As traffic is aggregated from multiple ISPs into transit networks, there is no longer enough information to unambiguously determine if a packet arriving on a particular interface has a “legal” source address. Moreover, examining packet source address becomes prohibitive on high-speed links.

The main disadvantage of ingress filtering is that its effectiveness depends on widespread, if not universal deployment. A secondary problem is that, even if ingress filtering were universally deployed at the customer-to-ISP level, attackers could still forge addresses from the hundreds or thousands of hosts within a valid customer network [11]. The third problem is the enormous overhead for examining IP headers, especially on high-speed links.

#### 3.2 Link Testing

A link testing approach typically starts from the router closest to the victim and interactively tests its upstream links until a determination can be made as to which one was used to carry the attacker’s traffic. Ideally, this tracing procedure is repeated recursively on the upstream router until the source is identified. Unfortunately, this technique assumes that an attack remains active until the completion of a trace, and is therefore incapable of
tracing completed or aborted attacks. There are two different schemes of this technique: input debugging and controlled flooding.

### 3.2.1 Input Debugging

Many routers provide a useful function referred to as input debugging. Input debugging allows an operator to filter particular packets on an egress port and determine the ingress port through which they arrived. A trace using input debugging [12] is implemented as follows: first, the victim must develop an attack signature that describes a common feature contained in all the attack packets. The victim then needs to report this signature to a network operator, who subsequently installs a corresponding input debugging filter on the victim’s upstream egress port. Upon successful debugging, this filter has the capacity to determine the associated input port, and hence which upstream router originated the traffic. The process is then repeated recursively on the upstream routers, until the originating source is reached or the trace leaves the ISP’s border. In the latter case, the upstream ISP must be contacted and the procedure repeated. As this demonstrates, the manual input debugging process is tedious and requires a large administration overhead. To overcome this obstacle, several automatic debugging tools have been developed. One such system, named CenterTrack [12], provides an improvement over hop-by-hop backtracking by dynamically rerouting all of the victim’s traffic to flow through a centralized tracking router [12]. Once this reroute is complete, a network operator can then use input debugging at the tracking router to investigate where the attack enters the ISP network.

The most obvious disadvantage of an input debugging approach, even with automated tools, is its considerable management overhead. Communicating and coordinating with network operators at multiple ISPs requires the time, efforts and commitment of both the victim and the remote personnel. A secondary problem is that this approach only works for ongoing attacks.

### 3.2.2 Controlled Flooding

Controlled flooding [13] tests links by flooding them with large bursts of traffic and observing how this perturbs traffic from the attacker. Using a pre-generated “map” of Internet topology, the victim forces selected hosts along the upstream route into iteratively flooding each incoming link on the router closest to the victim. Since router buffers are shared, packets traveling across the loaded link – including any sent by the attacker – have an increased probability of being dropped. By observing changes in the rate of packets received from the attacker, the victim can therefore infer which link they arrived from. The flooding procedure is then recursively applied on successive upstream routers until the source is reached.

This approach is quite controversial. The principal concern with this approach is that it is itself a DoS attack – exploiting vulnerabilities in unsuspecting routers to achieve its ends. The secondary limitation is that the victim has to have a good topological map of large sections of the Internet in addition to an associated list of “willing” flooding routers. Controlled flooding is also incapable of tracing DDoS attacks because the link-testing mechanism is inherently noisy and it can be difficult to discern the set of paths being exploited when multiple upstream links are contributing to the attack. Finally, like all link-testing schemes, controlled flooding is only effective at tracing on-going attacks and cannot be used to trace completed attacks.

### 3.3 Logging at Routers

#### 3.3.1 Naïve logging

Using the approach of naïve logging [9, 12], every packet that passes through key routers is logged at the input and the output interfaces. Data mining techniques can then be used to determine the path that the packets traversed. Logs can be maintained for some length of time. In case of an attack, the logs can be studied and the path to the attacker can be found easily. This method does not require the attack to be active, but it requires an enormous and impractical amount of storage. Logging every packet also imposes a considerable process overhead on routers. Therefore a sampling mechanism is needed to reduce both process overhead and storage requirements.
3.3.2 Hash-based IP traceback (SPIE)

While logging an entire packet is problematic and impractical, Snoeren and his colleagues [14,15] propose an improved logging technique. This scheme works as follows: for each packet passing through the router, a 32-bit hash digest of the first 28 invariant bytes (masked IP header plus 8 bytes of payload) of IP packet is generated. It was found that 28 bytes were enough to distinguish most packets. This digest is then recorded in a special filter called a Bloom filter, which is an efficient memory structure. Later on, when a packet needs to be checked against those that have passed the router, a digest can be generated to be fed into the bloom filter, and the result will reveal any matches.

To reconstruct the path, the victim sends a traceback request with the DoS packet to all routers that it is directly connected to. The routers verify the packet using their Bloom filters, and upon successful verification they check their tables. If any of the routers succeed in producing a match, that router queries its upstream routers with either the DoS packet or, in the case of a transformation, the reconstructed packet.

A main desirable feature of this approach is the capability of tracing a single packet. However, even with only a hash digest of packets stored at routers, there remains a major space constraint considering the enormous number of packets passing through loaded routers.

3.4 ICMP Traceback

This approach [16] uses explicit router-generated ICMP traceback (iTrace) messages. Every router samples, with low probability, one of the packets it is forwarding, copying the contents into a special ICMP traceback message that includes information about the adjacent routers along the path to the destination. During a flooding-style attack, the victim host can then use these messages to reconstruct a path back to the attacker.

This approach has many advantages when compared to the link testing and logging methods. For instance, it does not have huge management overhead, nor is there an enormous storage requirement, etc. However, there are several disadvantages in the scheme complicating its use. For instance, ICMP traffic is increasingly differentiated and may be filtered or blocked from normal traffic; The ICMP Traceback message relies on an input debugging capability (i.e. the ability to associate a packet with the input port and/or MAC address on which it arrived) that is not available in some router architectures; If only some of the routers participate, traceback will be broken on its way to the origins; Finally, it must deal with the problem of attackers sending false ICMP Traceback messages.

An intention-driven iTrace [17] is also introduced to reduce the unnecessary iTrace messages and thus improve the performance of iTrace systems.

3.5 IP Packet Marking

This approach is executed as follows: as a packet passes through routers, the routers insert traceback data into the packet to be traced at a later time. So, in an attack, the marking contained in traceback packets will reveal the path along which the attacker lies. But with this simple marking method, if every packet is marked, both the overhead and performance impact will be unacceptable. The idea of randomly encoding traceback data in IP packets was first presented by Savage, et al [10]. After that, this probabilistic packet marking (PPM) has attracted a great deal of attention, and has become one of the most promising approaches to address the traceback problem. In the above paper, a scheme was proposed in which routers would insert adjacent edge information into the ID field of packets with a small probability to avoid excessive overhead on the router’s packet marking. Upon an attack, the victim can collect enough packets to reconstruct each edge of the attack path, and consequently the full attack route.

The biggest disadvantage of this scheme is the combinatorial explosion during the edge identification step and the few feasible parameterizations. Also, this scheme requires a large number of packets for full-path traceback, while there are some attacks that require just a single packet – ping of death, land attack, etc where this scheme fails. The work of Song and Perrig provides a more in depth analysis of the faults of this scheme [18].

Song and Perrig proposed a modification [18] to Savage’s edge-identification-based PPM approach. Their approach can reduce storage requirements by storing a hash of each IP address instead of the address itself. This approach also enhanced Savage’s work by adding authentication to the embedded encodings of tracking
information. There have been two other notable proposals for IP traceback since the original proposal. Unfortunately, Song and Perrig’s approach requires that all victims have a current map of all upstream routers to all attackers (although a description is made of how such maps can be maintained).

Dean et al. proposed an algebraic approach [19] to IP marking traceback. The basis for this approach lies in mathematical techniques that were first developed for problems related to error correcting codes and machine learning. More specifically, their scheme encodes path information as points on polynomials. Algebraic methods are subsequently used to reconstruct these polynomials at the victim. This scheme demonstrates improved robustness over previous combinatorial approaches, both for noise elimination and multiple-path reconstruction. But, there remain many similar problems to those of other marking schemes including a large number of necessary packets, failure to detect multiple attackers at different distances, etc.

There are many other variations of the PPM methods including [20] and [21] etc. But the main PPM principles underpinning the technique have not been significantly changed.

3.6 IP Traceback with IPSec

In [22], Chang et al. proposed an IDS framework called DECIDUOUS. In this framework, they proposed a mechanism of locating attack sources using IPSec. This mechanism is based on the assumption that SAs are operating in tunnel modes and the complete network topology is known to the tracing system. The underlying principle is as follows: if attack packets are authenticated by a security association between the victim and a router. Then the attack must be originated on some host further than this router. Otherwise, the attack is originated on some host between the victim and this router. Therefore, by establishing security associations with suspicious routers from which the attack may come through, it is possible to locate the attack source. The strength of this approach is that IPSec is widely deployed and IPsec is proven to be very reliable. No new functionality needs to be developed by ISPs to enable this scheme. However, this mechanism requires the complete network topology is exposed to every system so that authenticated tunnels could be built between the system and any router in the network to locate attack sources. This assumption makes the mechanisms infeasible for tracing attacks across network domains and not scalable for widely deployment.

4 IP Traceback and IPv6

IPv6 is the next generation IETF standard protocol, designed to replace the current IPv4. The most salient feature of IPv6 is an enormously expanded address space compared to IPv4. IPv6’s address space is 128 bits whereas IPv4 contained only 32 bits. The direct consequence of having a much longer address field is that we can provide a distinct and static IPv6 address to every network device in the predictable future. This is a significant advantage from the perspective of tracking and tracing packets using a particular static IP address instead of ephemeral dynamic IP addresses. In addition, the IPv6 header design is particularly flexible and efficient, providing for a sequence of extension headers to carry optional information. This feature is especially desirable for IP marking since marks would no longer need to be broken across multiple packets.

But IPv6 definitely is not the ultimate solution for IP traceback. In a recent paper [23], researchers discussed the issue of IP traceback for IPv6. It is found that packet transformations introduced by IPv6, namely tunneling and address manipulation, constitute additional problems for tracing packets in IPv6 context. They presented a new ICMPv6 protocol message for the traceback coordination.

5 IP Traceback for Wireless Ad-hoc Networks

The above discussed traceback techniques are mainly addressing tracing and tracking IP packets in ad-hoc wired networks. Recently, wireless networks have been widely deployed and wireless Internet access has become readily accessible to public. Consequently, intelligent attackers are starting to exploit attacks launched over wireless network or to target wireless services. Therefore, researchers are challenged with a new task: tracing and tracking IP packets on wireless ad-hoc networks.

In wireless ad-hoc networks, devices typically only have limited bandwidth, computational resources, and battery power. Furthermore, packet routing is more unpredictable due to the mobility feature of wireless nodes. Therefore, these barriers constitute additional difficulties in locating attack sources.
In [24], Thing and Lee investigated the feasibility of applying the existing IP traceback techniques on wireless ad-hoc networks. Due to its large storage requirement, SPIE is not suitable for wireless traceback. Therefore, their analysis focused on PPM and ICMP traceback techniques. Their study shows that the performance of the above two traceback in wireless ad-hoc networks depended on the network size, the routing protocols and the traceback mechanisms used. While static routing makes traceback much easier, with dynamic routing, the traceback techniques would have to be applied with higher probability settings to make multiple attack paths reconstruction possible. Their qualitative analysis was based on network simulations and has not been tested in real world.

6 Conclusion

Tracing and tracking IP packets back to their origins is one of the most difficult dilemmas in network security. Although researchers have suggested numerous approaches addressing this problem, none of them has been fully satisfactory. While seeking out myriad solutions, they often introduce new problems in the process. In the following, we summarize and propose a list of desirable features that an efficient and effective IP traceback technique should possess:

- Low network traffic overhead.
- Minimal deployment overhead.
- Compatibility with existing network protocols.
- Minimal overhead in terms of time, resources and computation.
- Effectiveness and robustness against DDoS attacks.
- Capability of tracing and tracking small amount of packets, and even a single packet.
- Resistance to IP mark-spoofing and other manipulations of detection marks and messages.
- Capability of tracing and tracking after attacks are completed.

In addition to technique difficulties, IP traceback also involves many security policy and legal issues. For instance, different AS may have different conflicting policies. One ISP may be reluctant to participate in another ISP’s traceback process because the process may expose its internal routing infrastructure. Logging and analysis of IP packets passing through routers may also violate the owner’s privacy and involves many legal issues. In [25], Lipson provided an in-depth analysis of policy issues involved in IP traceback.

As the Internet extends its fingers to cover the entire globe, cyber attacks can be launched from any corner of the world. Thus, in many circumstances, IP traceback requires international or global involvement. Consequently, IP traceback goes beyond a technique problem and may involve complicated policy and legal issues, potentially requiring worldwide cooperation.

References