ISSN No: 2454-9614

IP Spoofing Identification

M. Mohammed Imran, B. Sivaranjini*, L.Govindhasamy, P. Gunasekaran

Department Of Computer Science and Engineering, Nandha Engineering College,
Perindurai-638052, Tamilnadu, India.

*Corresponding Author: B. Sivaranjini

E-mail: theerkapalaniswamyse@gmail.com

Received: 03/11/2015, Revised: 12/12/2015 and Accepted: 04/03/2016

Abstract:

It is long known attackers may use forged source IP address to conceal their real locations. To capture the spoofers, a number of IP trace back mechanisms have been proposed. However, due to the challenges of deployment, there has been not a widely adopted IP trace back solution, at least at the Internet level. As a result, the mist on the locations of spoofers has never been dissIPated till now. This paper proposes passive IP trace back (PIT) that bypasses the deployment difficulties of IP trace back techniques. PIT investigates Internet Control Message Protocol error messages (named path backscatter) triggered by spoofing traffic, and tracks the spoofers based on public available information (e.g., topology). In this way, PIT can find the spoofers without any deployment requirement. This paper illustrates the causes, collection, and the statistical results on path backscatter, demonstrates the processes and effectiveness of PIT, and shows the captured locations of spoofers through applying PIT on the path backscatter data set. These results can help further reveal IP spoofing, which has been studied for long but never well understood. Though PIT cannot work in all the spoofing attacks, it may be the most useful mechanism to trace spoofers before an Internet-level trace back system has been deployed in real.

Reviewed by **ICETSET'16** *organizing committee*

1. Introduction

1.1 Background IP Spoofing

Which means attackers launching attacks with forged source IP addresses, has been recognized as a serious security problem on the internet for long. By using addresses that are assigned to others or not assigned at all, attackers can avoid exposing their real locations, or enhance the effect of attacking, or launch reflection based attacks. A number of notorious attacks rely on IP spoofing, including SYN flooding, SMURF, DNS amplification etc.

1.2 Motivation

However, to capture the origins of IP spoofing traffic on the internet is thorny. The research of identifying the origin of spoofing traffic is categorized in IP trace back. To build an IP trace back system on the internet faces at least two critical challenges. The first one is the cost to adopt a trace back mechanism in the routing system. Existing trace back mechanisms are either not widely supported by current commodity routers (packet marking), or will introduce considerable overhead to the routers (internet control message protocol

(ICMP) generation, packet logging), especially in high-performance networks. The second one is the difficulty to make internet service providers (ISPS) collaborate. Since the spoofers could spread over every corner of the world, a single ISP to deploy its own trace back system is almost meaningless. However, ISPS, which are commercial entities with competitive relationshIPs, are generally lack of explicit economic incentive to help clients of the others to trace attacker in their managed ases.

1.3 Our Work

Instead of proposing another IP trace back mechanism with improved tracking capability, we propose a novel solution, named passive IP trace back (pit), to bypass the challenges in deployment. Routers may fail to forward an IP spoofing packet due to various reasons, e.g., TTL exceeding. In such cases, the routers may generate an ICMP error message (named path backscatter) and send the message to the spoofed source address. Because the routers can be close to the spoofers, the path backscatter messages may potentially disclose the locations of the spoofers. Pit exploits these path backscatter messages to find the location of the spoofers. With the locations of the spoofers known, the victim can seek help from the corresponding ISP to filter out the attacking packets, or take other counterattacks. Pit is especially useful for the victims in reflection based spoofing attacks, e.g., DNS amplification attacks. The victims can find the locations of the spoofers directly from the attacking traffic.

2. Helpful hints

2.1 IP traceback IP trace back techniques

Are designed to disclose the real origin of IP traffic or track the path. Existing IP trace back approaches can be classified into 5 main categories: packet marking, ICMP trace back, logging on the router, link testing, overlay, and hybrid tracing. Packet marking methods require routers modify the header of the packet to contain the information of the router and forwarding decision. Hence the receiver of the packet can then reconstruct the path of a packet (or an attacking flow) from the received packets. There are two classes of packet marking schemes: probabilistic packet marking and deterministic packet marking.

2.2 IP spoofing observation network telescope

It is a fundamental technique for passive observation of spoofing activities on the internet. Network telescope captures non-solicited messages, which are mainly generated by victim attacked by traffic with source prefix set in the scope owned by the telescope. Then, it can be determined a part of nodes which are attacked by spoofing traffic. Currently, the largest scale telescope is the CAIDA UCSD telescope, which owns 1/256 of all the IP addresses and is mainly used to observe DDOS activities and worms. Moore el at. **[8]** Presented a technique named “backscatter analysis” which infers characteristics of dos attacks based on traces collected by the network telescope. Though ICMP error messages are mentioned in the paper, it does not further investigate these messages to trace spoofers. CAIDA provides publicly accessible data. The main analysis and experimental work of this article are performed on the data supplied by CAIDA. The MIT spoofer project tries to disclose which networks are able to launch spoofing based attacks. Volunteer participants install a client that tests the spoofing ability of their hosts and networks. The statistic result shows 6700 ass out of 30205 do not filter spoofing

2.3 Classes and causes of path backscatter path backscatter messages

It can be triggered for various reasons. Based on rfc792, there can be totally 5 types of path backscatter messages, as listed in the following sections. There are a number of codes associated with each type. The combination of type and code specifies the cause that the router decides to send the ICMP message. We name the combination of type and code by class. We use the names defined to denote the classes of path backscatter messages. In the path backscatter dataset from CAIDA , totally 23 classes of path backscatter messages are found, 11 of them are listed in table i. Messages belonging to the other 12 types are very rare. We do not find all the possible classes.

3. Time exceeded

Time exceed_intrans messages are triggered by packets with zero TTL value. Such messages are the most common path backscatter messages. Though the attackers can set the initial TTL value to be large enough to avoid triggering such messages, they may intentionally send packets with small initial TTL values, which trigger routers on the path to generate TTL exceeding messages to consume the processor resource of the router. In general such attacks target the routers rather than hosts.

4. Destination unreachable

Unreach-filter-prohib, unreach-net-prohib and unreach- host-prohib messages are mainly triggered by filtering mechanisms deployed between the spoofing origin and the victim, e.g., access control list (ACL). A result of the MIT spoofer project shows 80% filters are deployed one IP hop from the source, and over 95% of blocked packets are filtered at the source as. Thus, such messages can be from the gateways near the spoofers. It should be noted that at least part of the spoofing traffic from the spoofers has been filtered. Considering the filtering granularity may be coarse, the remaining spoofing messages can still reach the victims. Thus, trace back in such a scenario is still valuable. Unreach_host and unreach_net messages are generated if there is no route to the destination. Such messages are mostly triggered by attacking traffic launched against a private or unallocated address prefix. Whenever a spoofer sends packets to a private address, if the spoofer is attached to a public network or the victim address is not in the same private network of the spoofer, such ICMP messages will be generated when the spoofing packets arrive at the DFZ (default-free zone). We find a large number of such messages whose original destination is a private address. Such messages may be triggered by attacks against hosts behind nat or in vpn. Unreach_needfrag messages are generated if the size of the attacking packets are larger than the mtu of a hop on the path, but they don’t fragment flag is set. Such messages may be generated due to attacks against the routers. Besides, we think such messages can be triggered occasionally. Attackers use large packet to consume the bandwidth of the target. Due to forged addresses are used, the attackers cannot get the ICMP message and are unaware of that the attacking packets are dropped on path.

4.1 Source Quench:

Source quench messages are generated when the router has no buffer to queue the original packet. It can be resulted from the aggregated attacking traffic is too large to be forwarded by the router. In general such messages are generated near the victim. However, if there are a large number of attackers in the same network/as, it is possible to trigger such messages on the gateways near the attackers.

4.2 Redirect:

Redirect host and redirect net messages are generated if the spoofing origin has two or more gateways and a gateway, g1, finds the spoofing packet should be sent to another gateway, g2, as this is the shortest path. As multi-homed networks become common, such messages may be generated with higher probability. Because this message is generated by gateways near the spoofing origin, it is particularly helpful to find the location of the origin.

5. Parameter problem:

Paramaribo messages are generated if the router finds a problem with the header parameters in the original packet. Such messages are rare in the dataset. Possibly they are triggered by malformed attacking packets or just some type of attack.

5.1 Collection of path backscatter messages though path backscatter

It can happen in any spoofing based attacks, it is not always possible to collect the path backscatter messages, as they are sent to the spoofed addresses. We classify spoofing based attacks into four categories, and discuss whether path backscatter messages can be collected in each category of attacks.

It should be noted that if the routing has not constraint, packets from any node v $\in$ v to od can bypass any intermediate node r. Then the tracking is meaningless. Fortunately, it is not the case in real networks.we make use of two assumptions on the routing respectively: 1) loop-free assumption: this assumption states there is no loop in the paths. This assumption always holds unless misconfiguration or the routing has not converged. 2) valley-free assumption: this assumption states there should be no valley in the as level paths. Though the increased complexity of as relationshIP has reduced the universality of this assumption, it is still the most common model of as level routing. In the following subsections, we discuss how to perform pit based on each of the assumption respectively.

5.2 tracking on loop-free assumption

Based on the loop free assumption, a vertexv is in the $\varphi(r,od)$ if and only if there is at least one loop-free path from v to OD passing R. Denote a loop-free path from v to u by lfpath(v,u), which is a sequence of verticals along the path. Then the suspect set is $\varphi(r,od)=\{\, v|\exists lfpath(v,od),r \in lfpath(v,od)\}$. To find all the satisfying verticals through enumerating is almost impossible for large-scale networks. We designed an algorithm.. This algorithm first finds a shortest path from R to OD. From the second vertex along the path, it checks if the removal of the vertex can break R and OD. Whenever such a vertex c is found, removing the vertex from g, and the set containing all the verticals which are still connected with r is just the suspect set. The algorithm to determine the suspect set based on loop-free assumption.

The following theorem can be proofed to illustrate the correctness of the algorithm. The proof of this theorem is placed at the appendix a. Theorem 1: from the second vertex along path(r,OD), remove the first articulation point c whose removal will break r and OD. Denote the sub graph containing r by sg(R). If and only if v is in sg(r), there exists a loop-free path from v to OD containing r. Apparently, to determine a suspect set whose size is no larger than n requires the vertex number connected with r is no more than n in g –cut edge(R,OD). Especially, if the size of suspect set is 1, the degree of r must be one, and OD must not be r. 2) tracking on valley-free assumption: based on the valley-free assumption, a vertex v is in the $\varphi(R,OD)$ if and only if there is at least one valley-free path from v to OD passing R. Denote a valley-free path from v to u by vfpath(v,u), which is a sequence of verticals along the path. Then the suspect set is $\varphi(R, OD)=\{ v|\exists vfpath(v,OD),R \in vf\ path(v,OD)\}$. The valley-free assumption can be only used in as-level topology. Considering the scale of as-level internet topology, for a path backscatter message (R, OD), it is very costly to find all the ases that has a valley-free path to od through r. At first we introduce the concept of customer cone [36], which means “as a, plus as customers, plus its customers’ customers, and so on”. The customer cone of as v is denoted by cone(v). Then we can proof the following theorem: theorem 2: when $od / \in cone(r)$, if and only if $v \in cone\ (r)$, there is a valley-free path from v to od passing r.

5.3 Where are the spoofers?

In this section, we present the locations of the spoofers captured. This result is achieved through combining the tracking mechanisms proposed in section IV. The procedure is as follows. For each path backscatter message, at first we check whether it belongs to the special classes listed in section iv-c. If yes, the reflector should be near the attacker. We simply use the source as of the message as the location of the spoofer. If the message does not belong to the types, it is mapped into an as tuple. Determine whether the as tuple can accurately locate the source as of the attacker based on the mechanisms proposed in section iv-b. Because we perform tracking at the as level, we only use the valley free assumption which results in better tracking capability than the loop-free assumption. Then if the as tuple can accurately locate the source as of the message, the source as of the spoofer is just this as. Then we also use the source as the location of the spoofer. We do not further investigate the location of the spoofers inside the ash because we do not know the inner structure and address allocation in the ases. However, at least the messages of the special classes listed in section iv-c can help locate the network of the spoofer. We got 2788 ases in which there are spoofers. 914 of them are located by the mechanisms in section iv-b, and 2148 are located based on the special classes of path backscatter messages. There are 274 ases located by both mechanisms. The full list of the ases can be fetched from http://tinyurl.com/lp959y4.

The captured ases are only a small portion of all the ases. We believe this result underestimated the total number of ases with spoofers reside in. Considering the limitation of the backscatter collection capability of the CAIDA network telescope, the uncertainness of path backscatter generation and the available datasets we can access, we are not able to provide a complete list of all the ases in which there are spoofers. Here we just present this partial result to illustrate the effectiveness of this proposed tracking mechanism. It can be the basis for further potential works. Besides, it should be noted that the ases with spoofers in are not the ases which indulge spoofing. Actually, there are a number of path backscatter messages are generated because of the filtering performed by the ases.

5.4 An Aggregated Attack

We performed tracking based on all the path backscatter messages whose original destination is 194.97.x.y, which is assigned in as5430. There are totally 13511 such path backscatter messages, and 30 ases are located (15 of them are located by the special classes of path backscatter messages, and 19 of them are located based on the valley-free assumption. 4 of them can be located by both mechanisms). We plot them in the as-level topology as fig. 21. We make use of shortest valley-free path to connect the ases with spoofers located and the victim as. This proposed mechanism certainly does not work in all the attacks and cannot capture all the spoofers, but it does tell something about the spoofing attacks. At least, the luckiest victims are able to locate some ofthe spoofers.this is valuable until an as-level trace back system is established.

6. Conclusion

We try to dissipate the mist on the locations of spoofers based on investigating the path backscatter messages. In this article, we proposed passive IP trace back (pit) which tracks spoofers based on path backscatter messages and public available information. We illustrate causes, collection, and statistical results on path backscatter. We specified how to apply pit when the topology and routing are both known, or the routing is unknown, or neither of them are known. We presented two effective algorithms to apply pit in large scale networks and proofed their correctness. We demonstrated the effectiveness of pit based on deduction and simulation. We showed the captured locations of spoofers through applying pit on the path backscatter dataset. These results can help further reveal IP spoofing, which has been studied for long but never well und

References

[1] Ektachauhan* , Sonia Vatta., International Journal Of Advanced Research In Computer Science And Software Engineering, Cyber Security In Data Mining Using Homomorphic Encryption Volume 3, Issue 6, June 2013 Issn: 2277 128x

[2] Parsikalpana,Ravindran Et Al., Data Storage Security Using Partially Homomorphic Encryption In A Cloud, International Journal Of Advanced Research In Computer Science And Software Engineering 3(4),April - 2013, Pp. 603-606.

[3] Rajan.S.Jamgekar, Geetashantanu Joshi , File Encryption And Decryption Using Secure Rsa, International Journal Of Emerging Science And Engineering (Ijese) Issn: 2319–6378, Volume-1, Issue-4, February 2013.

[4] Shilpa M Pund, Chitra G Desai, Implementation Of Rsa Algorithm Using Mersenne Prime, International Journal Of Networking & Parallel Computing Www.Cirworld.Com (Issn: 2319-4529) Volume 1, Issue 3, Dec 2012-Jan 2013.

[5] B.Persis Urbana Ivy, Purshotammandiwa.Mukeshkumar , A Modified Rsa Cryptosystem Based On 'N' Prime Numbers , International Journal Of Engineering And Computer Science Issn:2319-7242, Volume1 Issue 2 Nov 2012 Page No. 63-66.

[6] Samoud Ali, Cherifadnen, Rsa Algorithm Implementation For CIPhering Medical Imaging, Samoud Ali, Et Al International Journal Of Computer And Electronics Research [Volume 1, Issue 2, August 2012] Issn: 2278-5795.

[7] Taher Elgamal. A Public Key Cryptosystem And A Signature Scheme Based On Discrete Logarithms. In G. R. Blakley And David Chaum, Editors, Crypto 1984, Volume 196 Of Lecture Notes In Computer Science, Pages 10–18.Springer, 1984.

[8] Jasleen Kour, Deepankar Verma, International Journal Of Engineering Research In Management & Technology Issn : 2278-9359 (Volume-3, Issue-5).