IJISET - International Journal of Innovative Science, Engineering & Technology, Vol. 2 Issue 11, November 2015.
www.ijiset.com
ISSN 2348 – 7968

IP Traceback Using Path Backscatter Message

Kirti Panmand[1], Shivesh Kumar[2], Akash Singh [3], Amit Tiwari [4]

[1] Computer engg, Savitribai Phule Pune University, Pune, Maharashtra, India

[2] Computer engg, Savitribai Phule Pune University, Pune, Maharashtra, India

[3] Computer engg, Savitribai Phule Pune University, Pune, Maharashtra, India

[4] Computer engg, Savitribai Phule Pune University, Pune, Maharashtra, India

***Abstract*— It is well known to us that attackers use forged source IP address to conceal their real locations. Various mechanism were proposed to find the spoofers. However due to challenge of deployment there has been not a widely adopted IP Traceback solution.**

This paper proposed Passive IP Traceback (PIT) that overcome the challenge of deployment. This mechanism investigate path backscatter message i.e ICMP error message to find the spoofers. In this way, PIT can find the spoofers without any deployment requirement. It also illustrates the causes, compilation, and the statistical results on path backscatter, demonstrates the processes and efficiency of PIT, and shows the captured locations of spoofers through applying PIT on the path backscatter information set. These outcome can help further reveal IP spoofing, which has been deliberate for long but never well understood.

***Keywords*— Twitter, Traffic event detection, tweet classification, text mining, social sensing.**

I. BACKGROUND

IP SPOOFING, which means attackers throwing attacks with bogus source IP addresses, has been acknowledged as a severe safety problem on the Internet for long. By using addresses that are allotted to others or not allotted at all, attackers can escape revealing their real locations, or enrich the effect of attacking, or launch reflection based attacks. A number of disreputable attacks rely on IP spoofing, including SYN flooding, SMURF, DNS amplification etc. A DNS amplification attack which rigorously ruined the service of a Top Level Domain (TLD) name server is reported in [1]. However there has been a widespread conventional wisdom that DoS attacks are thrown from botnets and spoofing is no longer dangerous, the report of ARBOR on NANOG 50th meeting demonstrates spoofing is still significant in witnessed DoS attacks [2]. Certainly, based on the caught backscatter messages from UCSD Network Telescopes, spoofing activities are still commonly observed. To find the origins of IP spoofing traffic is of great significance. As long as the actual locations of spoofers are not revealed, they cannot be prevented from launching more attacks. Even just oncoming the spoofers, for example, defining the ASes or networks they reside in, attackers can be found in a smaller area, and filters can be placed closer to the attacker before attacking traffic get accumulated. The last but not the least, identifying the origins of spoofing traffic can help build a reputed system for ASes, which would be supportive to push the corresponding ISPs to verify IP source address.

II. INTRODUCTION

Though, to find the origins of IP spoofing traffic on the Internet is thorny. The research of classifying the origin of spoofing traffic is categorized in IP traceback. To form an IP traceback system on the Internet faces at least two critical challenges. The first one is the rate to adopt a traceback mechanism in the routing system. Present traceback mechanisms are either not widely sustained by current commodity routers (packet marking [3]), or will introduce significant burden to the routers (Internet Control Message Protocol (ICMP) generation, packet logging), particularly in high-enactment networks. The second one is the difficulty to build Internet service providers (ISPs) cooperate. Since the spoofers could spread over all corner of the world, a single ISP to deploy its own traceback system is almost meaningless. Though, ISPs, which are commercial entities with inexpensive relationships, are generally lack of explicit economic inducement to help clients of the others to trace attacker in their accomplished ASes. Since the deployment of traceback mechanisms is not of clear improvements but apparently high overhead, to the best knowledge of authors, there has been no installed Internet-scale IP traceback system till now. As a consequence, despite that there are a lot of IP traceback mechanisms suggested and a large number of spoofing activities detected, the real locations of spoofers still remain a secret. Given the

difficulties of the IP traceback mechanisms deployment, we are considering alternative direction: tracking the spoofers without deploying any supplementary mechanism. In another word, we try to reveal the location of spoofers from the traces produced by existing widely implemented functions on commodity routers when spoofing attacks happen.

III. RELATED WORK

Though PIT is used to execute IP traceback, it is very diverse from existing IP traceback mechanisms. Thus, the related work introduces existing IP traceback mechanism, and the second introduce the IP spoofing remark activities.

IP traceback technique is intended to release the real origin of IP traffic or track the path. Existing IP traceback methods can be differentiate into five main categories: packet marking, ICMP traceback, logging on the router, connection testing, superimpose, and hybrid tracing.

Packet marking method require router modify the caption(header) of the packet to hold the information of the router and sending decision. Therefore the receiver of the packet can then rebuild the path of a packet (or an attacking flow) from the conventional packets. There are two classes of packet marking systems: probabilistic packet marking [3], [4] and deterministic packet marking [5]. Packet marking method are generally considered to be lightweight because they do not cost storage resource on router and the connection bandwidth source. However, packet marking is not a commonly supported function on router; thus, it is difficult to allow packet marking traceback in the network.

Diverse from packet marking methods, ICMP traceback, produces addition ICMP messages to a collector or the endpoint. The ICMP messages can be used to rebuild the attacking path. Such like, if iTrace is enabled, router generate ICMP samples to destination with a sure probability. The shortcoming of ICMP traceback is significant additional traffic will be produced to consume the already stressed bandwidth resources. Likewise, when the attack is alongside the bandwidth of the casualty, the increases traffic will make the attack more serious. ICMP generation can be execute by the processor, but significant overhead will be introduce to the processor. Attacking path can be rebuilt from register on the router when router makes a record on the packets dispatched [6]. Bloom filter is used to decrease the number of bits to store a packet. However, to achieve a low sufficient collision probability in current high-speed network, the storage cost is still too large for commodity router.

CenterTrack suggests offloading the suspect traffic from edge router to special tracking router through a overlay network. Though such a mechanism can decrease the requirement on edge router, the managing of the tunnel and the overlay network will be significantly growth the network management overhead. It is found if many of of ASes can join the overlay network, the spoofer can be accurately located(found). However, the challenge in practice is how to make the ASes cooperate. The intra-domain version of this work can avoid this problem, but it is needed to update router to adopt modification on OSPF. The above mechanisms can be joint to achieve better tracing capacity and decrease the cost. There are a number of hybrid mechanisms employment both packet marking and logging. Though the overhead on routers can be compact, they require the routersto support both mechanisms or strategies.

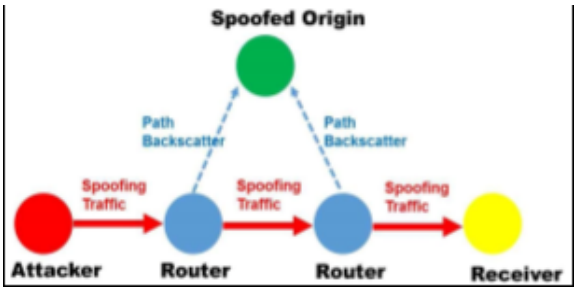

Fig. 1.The scenario of path backscatter generation and collection.

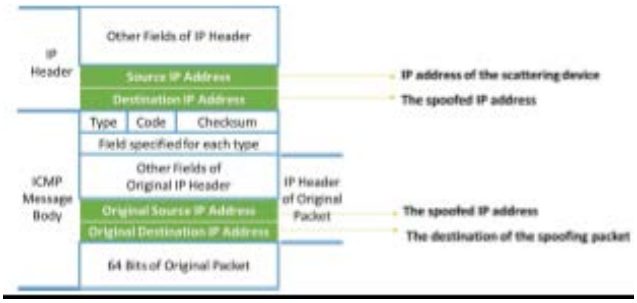

Fig. 2.The format of path backscatter messages.

IV. PATH BACKSCATTER

A) OVERVIEW

All packets not reach their destinations. A network devicecan may fail to forward a packet due to several reasons. In certain conditions, its generate an ICMP error message,this type of error messages called path backscatter messages. it will be sent to the source IP address indicated in the original packet. If the source address is fake, the message will be sent to the node who actually owns the address. This incomes the victims of reflection based attack, and the hosts whose addresses are used by spoofers, are possibly to collect such messages. This consequence is shown in Fig. 1. As specified by RFC792 , the

format of the path backscatter messages, is showed in Fig. 2. Each message contains the source address of the reflecting device, and the IP header of the original packet. Thus, from each path backscatter, we can get

1) the IP address of the reflecting device which is on the path from the attacker to the destination of the spoofing packet;

2) the IP address of the original destination of the spoofing packet. The original IP header also contains other valuable information, e.g., the remaining TTL of the spoofing packet. Note that due to some network devices may perform address rewrite (e.g., NAT), the original source address and the destination address may be different.

TABLE I

PATH BACKSCATTER CLASSES

Type	Class
Time Exceeded	TIMXCEED_INTRANS
Destination Unreachable	UNREACH_FILTER_PROHIB, UNREACH_NET_PROHIB, UNREACH_HOST_PROHIB, UNREACH_HOST, UNREACH_NET, UNREACH_NEEDFRAG
Source Quench	SOURCEQUENCH
Redirect	REDIRECT_HOST, REDIRECT_NET
Parameter Problem	PARAMPROB

B. CLASSES AND CAUSES OF BACKSCATTER

Path backscatter messages can be generated for several reasons. Based on RFC792, there can be totally 5 kinds of path backscatter messages, as listed in the following section. 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.Which is known as class. In the path backscatter dataset from CAIDA, totally 23 classes of path backscatter messages are found, 11 of them are listed in above Table I. Messages going to the other 12 types are extremely incoherent. We do not need to find all the achievable classes. We try to explain the causes of the classes of path backscatter messages given in Table I based on analyzing the dataset. Exactly, we try to understand the reasons that they are generated near the spoofers. Though, considering the huge number of spoofing messages, if only a small ratio of them trigger path backscatter messages near the spoofer, the total path backscatter dataset will be appreciated. Smooth for the path backscatter messages generated far away from the spoofers, their generation locations are at least closer to the spoofers than the victim. Thus, they can be used in the first step of traceback.

1) Time Exceeded:

TIMXCEED_INTRANS messages are caused by packets with zero TTL value. This type of messages are the most common path backscatter messages. Though the attackers can set the initial TTL value to be large enough to avoid causing such messages, they may intentionally send packets with small initial TTL values, which trigger routers on the path to produce TTL Exceeding messages to devour the processor resource of the router. In general such attacks target the routers rather than hosts. We can also find the attack against a host and the attack against the nearby routers of the target host can be combined. We think the attacker may want to degrade the forwarding performance of the routers near the target host, and then less aggregated spoofing trafficare require to avoid legitimate traffic from reaching the host. Also, to regulate the correct initial TTL value to make sure the TTL exceeding event happen on the targeted router, the attacking hosts should achieve some traces. The traces using real address can be cloaked with a number of traces using forged addresses to avoid tracking. This could be the reason that we found a number of TIMXCEED_INTRANS messages from cascading routers in the dataset.

2) Destination Unreachable:

UNREACH_FILTER_ PROHIB, UNREACH_NET_PROHIB and UNREACH_ HOST_PROHIB messages are mainly caused by filtering mechanism organized between the spoofing origin and the victim, e.g., Access Control List (ACL). A result of the MIT Spoofer project shows 80% filters are organized 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, traceback in such a scenario is still valuable.

UNREACH_NEEDFRAG messages are produced if the size of the attacking packets are larger than the MTU of a hop on the path, but the Don't Piece flag is set. Such messages may be produced due to attacks against the router. Also, we think such messages can be caused occasionally. Attacker use large packet to consume the bandwidth of the target. Due to bogus addresses are used, the attacker can't get the ICMP message and are unaware of that the attacking packets are released on path.

3) Source Quench:

SOURCEQUENCH messages are produced when the router has no buffer to queue the original packet. It can be resulted from the collected attacking traffic is too large to be promoted by the router. In general such messages are produced near the victim. Though, if there are a large number of attackers in the same network/AS, it is possible to cause such messages on the gateway near the attackers.

4) Redirect:

REDIRECT_HOST and REDIRECT_NET messages are produced if the spoofing origin has more than two gateways and a gateway, G1, finds the spoofing packet should be sent to next one gateway G2, this is the shortest path. As multi-home networks become public, such messages may be produced with higher probability. Because this message is produced by gateways near the spoofing origin, it is mainly helpful to find the location of the origin.

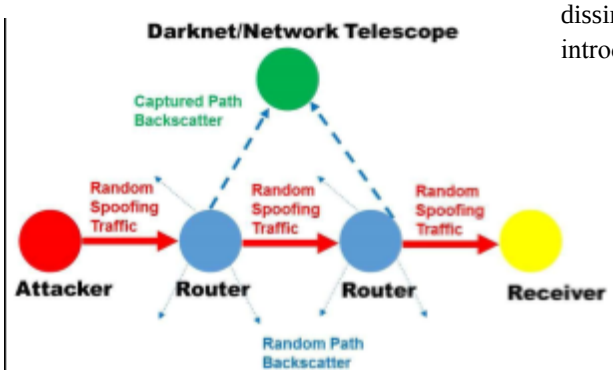

Fig. 3. Network telescope captures path backscatter in random spoofing attacks.

As mention in RFC792, G1 should checked the address of the packet and G2 are in the same network. Though, the dataset is composed by a network telescope, and seemingly any G2 and the address of a network telescope must not in the same network. It may be due to misconfiguration or operations varying with the standard.

5) Parameter Problem:

PARAMPROB messages are produced if the router finds a problem with the header parameters in the original packet. those messages are infrequent in the dataset. Possibly they are caused by malformed attacking packets or impartial some type of attack.

C. COLLECTION OF PATH BACKSCATTER MESSAGES :

However path backscatter can happen in any spoofing based attacks, it is not permanently possible to gather the path backscatter messages, as they are sent to the spoofed addresses. We categorize spoofing based attacks into four categories, and discuss whether path backscatter messages can be composed in each category of attacks.

1) Multiple Sources, Single Destination:

In such attacks, the source address of spoofing packets is selected from a set of applicant addresses. Particularly, this set comprises a`ll the addresses. Such attacks are named random spoofing. Random spoofing is naturally used to reduce the resource of the target, e.g., SYN flooding. Network Telescopes can be used to confinement path backscatter messages in random spoofing attacks. As shown in Fig. 3, in random spoofing attacks, the path backscatter messages are sent to the randomly spoofed addresses. Because the addresses owned by network telescopes can be used in random spoofing attacks, the network telescopes are possibly able to capture part of the path backscatter messages.

2) Single Source, Multiple Destinations:

In this type of attacks, all of the spoofing packets have the similar or same source IP address. The packets are sent to dissimilar destinations. Such packets are classically used to introduction reflection attacks.

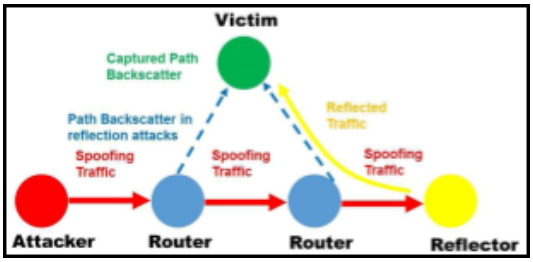

Fig. 4. The victim captures path backscatter in reflection attacks.

Reflection attacks, e.g., DNS amplification, are the most dominant IP spoofing attacks in current years. The victim in a reflection attack is the host who owns the spoofed address. The victim itself is able to capture all the path backscatter messages in reflection attacks. As shown in Fig. 4, because all the spoofing packets are set the address of the victim, all the path backscatter messages will be sent to the victim. Then the victim can grow the path backscatter messages complete checking if it has sent messages to the original destination IP address field in received ICMP messages.

3) Multiple Sources, Multiple Destinations:

Spoofing attacks can be thrown against multiple destination IP addresses going to the same website or service provider e.g., cloud. Normally, such attacks can be observed as the combination of multiple attacks going to the above two types.

4) Single Source, Single Destination:

Such attacks are often used to hijack a session among the source and the destination, e.g., Man-in-the-Middle, TCP hijack, replay attack. The spoofed origin is able to imprisonment the path backscatter messages. Though, because the spoofing packets are normally rare compared with the other types of attacks, path backscatter messages could be fairly infrequent. On the other hand, because the attacker and the spoofed origin frequently exist in the same network, it is possible to track the attacker more efficiently than using path

backscatter. In summary, path backscatter messages can be successfully composed in random spoofing attacks, reflection attacks and their combinations, which cover the mainstream of IP spoofing attacks

D. SECURITY ISSUES WITH PATH BACKSCATTER MESSAGES :

It should be noted it is almost as easy to construct a path backscatter message as to produce a spoofing data packet. Thus, the gatherer should filter out the forged packet backscatter messages to avoid false positive. For reflection attack, the victim can get the valid hop count from the routers to itself through tracing or passive learning. Then the mapping from router to hop count can be used to filter out a large part of spoofing packets based on the mechanism projected in [7]. The attacker must get the correct hop count from each router to the victim to bypass such a filtering mechanism. However, it is difficult and costly for the attacker to achieve this information, as it cannot get the hop count among the victim and the router from tracing directly. Hop count based filtering can not remove all the spoofing messages, but anyway it makes the spoofing of path backscatter message harder. Besides, to bypass such filtering, the spoofers have to send some trace or provisional messages, and such messages may uncover their locations and objectives. Another approach of the attackers is sending forged path backscatter messages with all the possible TTL values, but the victim can check whether there are path backscatter messages from a node but with several TTL values and/or hop counts to identify such attacks. For path backscatter messages took by network tele- scope in random spoofing, hop count based filtering can also be used by the network telescope itself. Though, a third- party who is performing tracing does not know the hop count from each router to the network telescope. In this article, we make use of clustering based apparatus to filter out forged path backscatter messages. We abstract all the prefixes from the BGP dataset. For backscatter messages from each prefix,

1) We divided the whole dataset into 1-hour slices. The routing and address assignment are dynamic on the Internet. We chose one hour as the time recess in order to make a trade-off between getting enough data and mitigating the effects of network changes. Note that because network telescopes collect all the non-solicited messages, the dataset contains all the messages, in which only a small portion are path backscatter messages.

2) We inferred the addresses of NAT from each slice. First, we inferred the initial TTL value and hop count of each message (not only path backscatter messages) based on the mechanism proposed in. If an address has multiple initial TTL values but approximate hop counts, the address is considered as a NAT address.

3) For each address other than NAT addresses, we got the mode of hop count value. We filtered out the path backscatter messages whose hop count is deviated from the mode more than 1. We didn't use exact match for taking the network changes into account. For NAT addresses, we got multiple modes of hop count, and filtered messages whose hop count is deviated from the nearest mode more than 1. The rationale of this mechanism is the address space of network telescope is hidden. Thus, it will not be targeted, and the received forged path backscatter messages are rare. On the other hand, we make use backscatter messages from hosts together with path backscatter messages. Thus, the dataset is large and the messages are from almost every corner of the Internet. Then the majority of learned hop count should be valid, avoiding pollution from forged path backscatter messages. Although forged path backscatter messages may be of matched hop count accidentally, the possibility will be quite low. To effectively pollute the captured dataset, an attacker will have to send messages with all the possible TTL values to every corner of the Internet. This requires tremendous effort, but the forged messages can still be effectively identified through checking whether messages from a node are with wide-ranged hop counts.

A misunderstanding about the packet backscatter messages is that the normal ICMP error messages, e.g., Time-exceeded messages generated in traceroute, may be regarded as triggered by spoofing packets. However, hosts are using the real source IP address in normal behaviors, and the normal ICMP error messages will go to the hosts themselves rather than the collectors. Thus, innocent hosts triggering normal ICMP error messages will not be regarded as spoofers.

V. CONCLUSION

We try to scatter the mist on the the locations of spoofers based on investigating the path backscatter messages. In this article, we proposed Passive IP Traceback (PIT) which tracks spoofers based on path backscatter messages and public available information. We illustrate causes, group, 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 actual algorithms to apply PIT in large scale networks and proofed their perfection. 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 understood.

ACKNOWLEDGMENT

We are very much grateful to Prof. Kirti Panmand for reviewing the paper.

REFERENCES

[1] ICANN Security and Stability Advisory Committee, "Distributed denial of service (DDOS) attacks," SSAC, Tech. Rep. SSAC Advisory SAC008, Mar. 2006.

[2] C. Labovitz, "Bots, DDoS and ground truth," presented at the 50th NANOG, Oct. 2010.

[3] S. Savage, D. Wetherall, A. Karlin, and T. Anderson, "Practical network support for IP traceback," in Proc. Conf. Appl., Technol., Archit., Protocols Comput. Commun. (SIGCOMM), 2000, pp. 295–306.

[4] D. X. Song and A. Perrig, "Advanced and authenticated marking schemes for IP traceback," in Proc. IEEE 20th Annu. Joint Conf. IEEE Comput. Commun. Soc. (INFOCOM), vol. 2. Apr. 2001, pp. 878–886.

[5] Y. Xiang, W. Zhou, and M. Guo, "Flexible deterministic packet marking: An IP traceback system to find the real source of attacks," IEEE Trans. Parallel Distrib. Syst., vol. 20, no. 4, pp. 567–580, Apr. 2009.

[6] A. C. Snoerenet al., "Hash-based IP traceback," SIGCOMM Comput. Commun. Rev., vol. 31, no. 4, pp. 3–14, Aug. 2001. 484 IEEE TRANSACTIONS ON INFORMATION FORENSICS AND SECURITY, VOL. 10, NO. 3, MARCH 2015

[7] H. Wang, C. Jin, and K. G. Shin, "Defense against spoofed IP traffic using hop-count filtering," IEEE/ACM Trans. Netw., vol. 15, no. 1, pp. 40–53, Feb. 2007.