Volume 40, Number 4
October 2010

Published by the Association for Computing Machinery
Special Interest Group on Data Communication

An ACM SIGCOMM Publication

COMPUTER COMMUNICATION *review*

Proceedings of the

Association for Computing Machinery

Advancing Computing as a Science & Profession

acm sigcomm

Computer Communication Review

a publication of the acm special interest group on data communication

SIGCOMM Mission Statement: SIGCOMM is ACM's professional forum for the discussion of topics in the field of communications and computer networks, including technical design and engineering, regulation and operations, and the social implications of computer networking.

The SIG's members are particularly interested in the systems engineering and architectural questions of communication.

Feedback on our mission is always welcome. Please email suggestions to any member of the SIGCOMM Executive Committee.

Computer Communication Review (CCR) is a publication of the ACM Special Interest Group on Data Communication and publishes articles on topics within the SIG's field of interest. CCR serves as a forum for interesting and novel ideas at an early stage in their development. The focus is on timely dissemination of new ideas that may help trigger additional investigations.

Items attributed to persons will ordinarily be interpreted as personal rather than organizational opinions. Submissions to CCR can be made online at http://blizzard.cs.uwaterloo.ca/ccr

SIGCOMM mailing list: sigcomm@postel.org

This mailing list is for discussions of ACM SIGCOMM, including technical and administrative discussions, CCR, and conference posts. Information available at http://www.postel.org/sigcomm/

COMPUTER COMMUNICATION REVIEW (ISSN 0146-4833) is published five times a year (January, April, July and two issues in October) by the ACM, Inc., 2 Penn Plaza, Suite 701, NY, NY 10121-0701 USA. Annual subscription cost of $14.44 is included in SIGCOMM's member dues of $25.00 (for students, the cost is included in the student membership fee of $15.00); the non-member annual subscription is $35.00. Periodicals postage paid at New York, NY 10001, and at additional mailing offices.

POSTMASTER: Send address changes to ACM COMPUTER COMMUNICATION REVIEW, ACM, 2 Penn Plaza, Suite 701, NY, NY 10121-0701 USA

Passive IP Traceback: Capturing the Origin of Anonymous Traffic through Network Telescopes

Guang Yao, Jun Bi, Zijian Zhou
Network Research Center, Tsinghua University, Beijing 100084, China
{yaog, junbi, zhouzj}@netarchlab.tsinghua.edu.cn

ABSTRACT

IP traceback can be used to find the origin of anonymous traffic; however, Internet-scale IP traceback systems have not been deployed due to a need for cooperation between Internet Service Providers (ISPs). This article presents an Internet-scale Passive IP Trackback (PIT) mechanism that does not require ISP deployment. PIT analyzes the ICMP messages that may scattered to a network telescope as spoofed packets travel from attacker to victim. An Internet route model is then used to help re-construct the attack path. Applying this mechanism to data collected by Cooperative Association for Internet Data Analysis (CAIDA), we found PIT can construct a trace tree from at least one intermediate router in 55.4% the fiercest packet spoofing attacks, and can construct a tree from at least 10 routers in 23.4% of attacks. This initial result shows PIT is a promising mechanism.

Categories and Subject Descriptors: C.2.0 [**Computer-Communication Networks**]: General– *Security and protection.*

General Terms: Design

Keywords: IP traceback, network telescope

1. INTRODUCTION

Distributed Denial of Service (DDoS) [1] attacks are one of the most pressing security threats on the Internet. Attackers can launch attacks from corners of the Internet, and the aggregated traffic can exhaust the bandwidth (or other resources) at the victim. DDoS attack flows can be shut off by configuring IP source based filters if one can identify the malicious clients. However, attack flows can use forged source addresses and thus source based filtering will fail. In addition to sending forged packets to the victim, IP spoofing can play a critical role in reflection based DDoS attacks. Examples of notorious DDoS attacks that make use of forged source address include SYN flooding, SMURF, DNS amplification, etc. [2] reported a DNS amplification attack against a Top Level Domain (TLD) name server which severely degraded the service of the TLD name server for a long period.

The objective of IP traceback [3] is to find the origin of spoofing traffic. If the origin of spoofing traffic is found, the attacker can be deterred from launching further attacks. Most IP traceback approaches trace the spoofed traffic to the edge of region where traceback is deployed. Unfortunately, the non-cooperation nature of Internet Service Providers (ISPs) means IP traceback approaches are only be deployed in a domain controlled by the single ISP, and can only trace to the edge of this domain. To the best of our knowledge, there has been no Internet-scale IP traceback system deployed.

In this article, authors propose a natural Internet-scale Passive IP traceback (PIT) approach that requires no ISP deployment. PIT analyzes packets collected from network telescopes, and infers the locations of spoofed traffic. The intuition behind PIT is that spoofing flows may trigger ICMP error messages at routers on the path to victim. These ICMP messages are sent to the spoofed nodes. Under the assumption that attackers use randomly forged addresses, some of these ICMP messages will be received by the network telescopes. The addresses of routers sending the message can be combined with an Internet route model to re-construct the attack path and find the locations of the spoofers.

SIGCOMM 2010, August 30–September 3, 2010, New Delhi, India.
ACM 978-1-4503-0201-2/10/08.

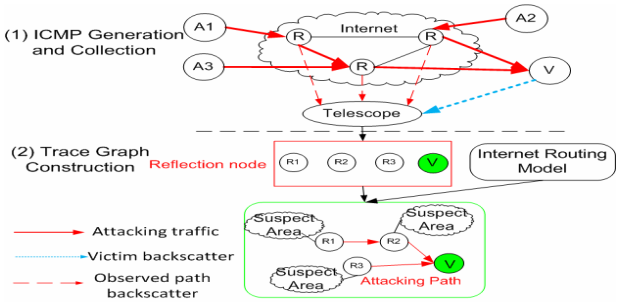

Figure 1. Passive IP Traceback

2. PASSIVE IP TRACEBACK

Figure 1. shows the structure of our Passive IP traceback (PIT) approach. It consists of the existing ICMP generation mechanisms on routers, a network telescope to collect the ICPMP backscatter, and route model used to construct a trace graph.

2.1 ICMP Generation

Previous work [4] has observed that ICMP messages are generated by routers on the path from the attacker to the victim. CAIDA makes use of 1/256 of entire IP address space to collect passive traffic and supplies the dataset used in this study. Through analyzing the 2008 CAIDA data set [6], we found that the ICMP messages reflected on path (*path backscatter*) are non-negligible both in number and reflection locations as shown in Table 1. Since the ICMP messages are reflected to the telescope, we refer to the routers sending these messages as ***reflection routers***.

Table 1. Message and router number per ICMP type

Type	Packets Reflected	Routers
TTL exceeded in transit	167041824	70144
Communication administratively prohibited	810246	16312
Destination host unreachable	5066141	183933
Host administratively prohibited	103395	678
Destination network unreachable	1874413	7235
Redirect Datagram for the Host	688902	7456
Fragment reassembly time exceeded	44606	1667
Network administratively prohibited	55326	332

Figure 2 shows the rough "distance" between reflection routers and the victim. The distance is calculated by comparing the reflection router's address and the victim's address. The results show routers sending the ICMP messages are generally away from victim. Further evaluation work will be done using a router level topology.

Figure 2. Address "distance" proportion

Figure 3 shows the number of routers per attack for the top 100 attacks each month. Attacks are ranked by the number of ICMP messages (including reflection from victim and intermediate router). 55.4% of the attacks triggered ICMP reflection on at least 1 router, and 23.4% triggered reflection on at least 10 routers. Though obviously this mechanism is not effective for all attacks, it works in a steady proportion of attacks. Given it requires no new deployment at any ISP or router, this result is acceptable. Heuristic link testing can be performed to trigger more path backscatter.

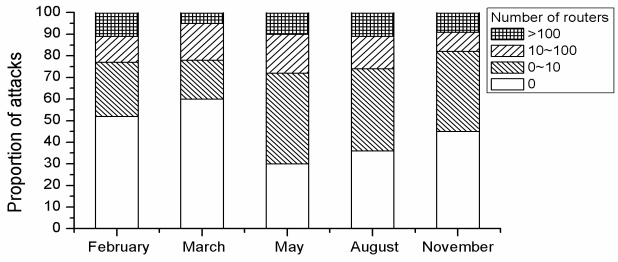

Figure 3. Router number per attack proportion

Figure 4 shows the locations of reflection routers identified during a February 2008 attack against a victim in Taiwan. Reflection router IP addresses were mapped to locations using http://www.ipaddresslocation.org/.

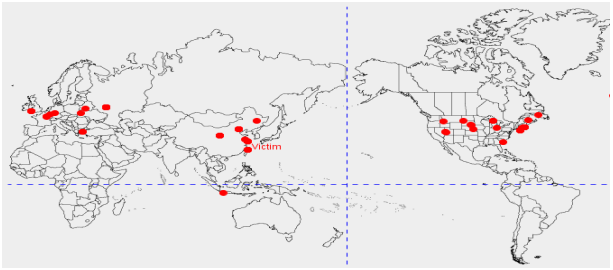

Figure 4. Reflection locations of an actual attack in Feb. 2008

2.2 Trace Graph Construction

PIT faces a new problem of inferring the origin and path of spoofed traffic from the incomplete set of reflection routers. Some estimates of the Internet topology have been published, but the actual path between two nodes can be affected by the BGP policies, traffic engineering and non-negligible accidental factors [5].

2.2.1 Path from Reflection Router to Victim

If all Internet routes were known, determining the path from reflection router to victim would be a trivial problem. However, it is impossible to achieve an Internet route map that covers all the ISPs. "Predicting" the route from one AS to another is not a new problem. Recent research [5] relies on training data model from known routes, and inferring unknown routes. This model can greatly reduce the variety of all the valley-free paths between two nodes based on inferred AS relationships; however, false negatives may occur because the policy and topology predictions may not be accurate. A more suitable model should balance false positives and false negatives.

To construct finer granularity path is even more challenging as it requires ISPs to expose their inner policies. However, considering the re-active ability of the victim, finer granularity path may be not of important value.

2.2.2 Identifying Suspect Areas

Finding the suspect area that contains the spoofed origin is equivalent to finding the nodes from which the spoofing traffic to victim will pass at least one of the reflection routers, which is actually a similar problem with determining the path between two nodes.

Because the reflection router may be near the victim, a probabilistic method should be used to filter the results; otherwise the resulting set can become too large to be meaningful. One simple method to is mark all suspect nodes with the same probability on each (intermediate router, victim) pair in an attack, and accumulate the probability values to find the most suspect nodes. The accumulated result on all attacks can be used to help reduce bias that may be present in one attack.

3. CONCLUSION

This paper introduced a novel Passive IP traceback mechanism (PIT) that can help identify the actual origin of spoofed traffic. A major advantage of PIT is that it requires no new deployment at any router or ISP. Given the set of reflection routers observed at a telescope, a method to construct an attack path is also proposed. Initial results show it is practical though not perfect.

4. ACKNOWLEDGMENTS

Special thanks to CAIDA for supplying us with dataset and tools.

5. REFERENCES

[1] CERT, "CERT Advisory CA-1997-28 IP Denial-of-Service Attacks", http://www.cert.org/advisories/CA-1997-28.html , 1997.

[2] ICANN. (March 2006). "SSAC Advisory SAC008 DNS Distributed Denial of Service (DDoS) Attacks.".

[3] S. Savage, D.Wetherall, A. R. Karlin, and T. Anderson, "Practical network support for IP traceback," in *Proc. ACM SIGCOMM*, 2000, pp.295–306.

[4] David Moore , Colleen Shannon , Douglas J. Brown , Geoffrey M. Voelker , Stefan Savage, Inferring Internet denial-of-service activity, ACM Transactions on Computer Systems (TOCS), v.24 n.2, p.115-139, May 2006.

[5] Wolfgang Mühlbauer, Anja Feldmann, Olaf Maennel, Matthew Roughan, Steve Uhlig: Building an AS-topology model that captures route diversity. SIGCOMM 2006: 195-206.

[6] The CAIDA Backscatter-2008 Dataset, Colleen Shannon, David Moore, Emile Aben, and kc claffy, http://www.caida.org/data/passive/backscatter_2008_dataset.xml.