Attribution of Internet Probes

Abstract

   Active measurements over the public Internet can target either
   collaborating parties or non-collaborating ones. Sometimes these
   measurements, also called "probes", are viewed as unwelcome or
   aggressive.

   This document suggests some simple techniques for a source to
   identify its probes. This allows any party or organization to
   understand what an unsolicited probe packet is, what its purpose is,
   and, most importantly, who to contact. The technique relies on
   offline analysis of the probe; therefore, it does not require any
   change in the data or control plane. It has been designed mainly for
   layer 3 measurements.

Status of This Memo

   This document is not an Internet Standards Track specification; it is
   published for informational purposes.

   This document is a product of the Internet Engineering Task Force
   (IETF). It represents the consensus of the IETF community. It has
   received public review and has been approved for publication by the
   Internet Engineering Steering Group (IESG). Not all documents
   approved by the IESG are candidates for any level of Internet
   Standard; see Section 2 of RFC 7841.

   Information about the current status of this document, any errata,
   and how to provide feedback on it may be obtained at
   https://www.rfc-editor.org/info/rfc9511.

Table of Contents

1. Introduction

   Most measurement research (e.g., [LARGE_SCALE], [RFC7872], and
   [JAMES]) is about sending IP packets (sometimes with extension
   headers or layer 4 headers) over the public Internet, and those
   packets can be destined to either collaborating parties or non-
   collaborating ones. Such packets are called "probes" in this
   document.

   Sending unsolicited probes should obviously be done at a rate low
   enough to not unduly impact the other parties' resources. But even
   at a low rate, those probes could trigger an alarm that will request
   some investigations by either the party receiving the probe (i.e.,
   when the probe destination address is one address assigned to the
   receiving party) or a third party having some devices through which
   those probes are transiting (e.g., an Internet transit router). The
   investigation will be done offline by using packet captures;
   therefore, probe attribution does not require any change in the data
   or control planes.

   This document suggests some simple techniques for a source to
   identify its probes. This allows any party or organization to
   understand:

   * what an unsolicited probe packet is,

   * what its purpose is, and

   * most importantly, who to contact for further information.

   It is expected that only researchers with good intentions will use
   these techniques, although anyone might use them. This is discussed
   in Section 7.

   While the technique could be used to mark measurements done at any
   layer of the protocol stack, it is mainly designed to work for

measurements done at layer 3 (and its associated options or extension headers).

2. Probe Description

This section provides a way for a source to describe (i.e., to identify) its probes.

2.1. Probe Description URI

This document defines a Probe Description URI as a URI pointing to one of the following:

* a Probe Description File (see Section 2.2) as defined in Section 8, e.g., "https://example.net/.well-known/probing.txt";

* an email address, e.g., "mailto:lab@example.net"; or

* a phone number, e.g., "tel:+1-201-555-0123".

2.2. Probe Description File

As defined in Section 8, the Probe Description File must be made available at "/.well-known/probing.txt". The Probe Description File must follow the format defined in Section 4 of [RFC9116] and should contain the following fields defined in Section 2 of [RFC9116]:

* Canonical

* Contact

* Expires

* Preferred-Languages

A new field "Description" should also be included to describe the measurement. To match the format defined in Section 4 of [RFC9116], this field must be a one-line string with no line break.

2.2.1. Example

```
# Canonical URI (if any)
Canonical: https://example.net/measurement.txt

# Contact address
Contact: mailto:lab@example.net

# Validity
Expires: 2023-12-31T18:37:07z

# Languages
Preferred-Languages: en, es, fr

# Probes description
Description: This is a one-line string description of the probes.
```

3. Out-of-Band Probe Attribution

    A possibility for probe attribution is to build a specific URI based
    on the source address of the probe packet, following [RFC8615]. For
    example, with a probe source address 2001:db8:dead::1, the following
    URI is built:

    *   If the reverse DNS record for 2001:db8:dead::1 exists, e.g.,
        "example.net", then the Probe Description URI is
        "https://example.net/.well-known/probing.txt". There should be
        only one reverse DNS record; otherwise, the Probe Description File
        should also exist for all reverse DNS records and be identical.

    *   Else (or in addition), the Probe Description URI is
        "https://[2001:db8:dead::1]/.well-known/probing.txt".

    The built URI must be a reference to the Probe Description File (see
    Section 2.2).

    As an example, the UK National Cyber Security Centre [NCSC] uses a
    similar attribution. They scan for vulnerabilities across Internet-
    connected systems in the UK and publish information on their scanning
    [NCSC_SCAN_INFO], providing the address of the web page in reverse
    DNS.

4. In-Band Probe Attribution

    Another possibility for probe attribution is to include a Probe
    Description URI in the probe itself. Here is a non-exhaustive list
    of examples:

    *   For an ICMPv6 echo request [RFC4443], include it in the data
        field.

    *   For an ICMPv4 echo request [RFC0792], include it in the data
        field.

    *   For a UDP datagram [RFC0768], include it in the data payload if
        there is no upper-layer protocol after the transport layer.

    *   For a TCP segment [RFC9293], include it in the data payload if
        there is no upper-layer protocol after the transport layer.

    *   For an IPv6 packet [RFC8200], include it in a PadN option inside
        either a Hop-by-Hop or Destination Options header.

    The Probe Description URI must start at the first octet of the
    payload and must be terminated by an octet of 0x00, i.e., it must be
    null terminated. If the Probe Description URI cannot be placed at
    the beginning of the payload, then it must be preceded by an octet of
    0x00. Inserting the Probe Description URI could obviously bias the
    measurement itself if the probe packet becomes larger than the path
    MTU. Some examples are given in Appendix A.

    Using a magic string (i.e., a unique, special opaque marker) to
    signal the presence of the Probe Description URI is not recommended

as some transit nodes could apply different processing for packets containing this magic string.

For the record, in-band probe attribution was used in [JAMES].

5. Operational and Technical Considerations

Using either the out-of-band or in-band technique, or even both combined, highly depends on intent or context. This section describes the upsides and downsides of each technique so that probe owners or probe makers can freely decide what works best for their cases.

The advantages of using the out-of-band technique are that the probing measurement is not impacted by probe attribution and that it is easy to set up, i.e., by running a web server on a probe device to describe the measurements. Unfortunately, there are some disadvantages too. In some cases, using the out-of-band technique might not be possible due to several conditions: the presence of a NAT, too many endpoints to run a web server on, the probe source IP address cannot be known (e.g., RIPE Atlas [RIPE_ATLAS] probes are sent from IP addresses not owned by the probe owner), dynamic source addresses, etc.

The primary advantage of using the in-band technique is that it covers the cases where the out-of-band technique is not feasible (as described above). The primary disadvantage is that it could potentially bias the measurements, since packets with the Probe Description URI might be discarded. For example, data is allowed in TCP segments with the SYN flag [RFC9293] but may change the way they are processed, i.e., TCP segments with the SYN flag containing the Probe Description URI might be discarded. Another example is the Probe Description URI included in a Hop-by-Hop or Destination Options header inside a PadN option. Section 2.1.9.5 of [RFC4942] (an Informational RFC) suggests that a PadN option should only contain 0s and be smaller than 8 octets, thus limiting its use for probe attribution. If a PadN option does not respect the recommendation, it is suggested that one may consider dropping such packets. For example, since version 3.5, the Linux Kernel follows these recommendations and discards such packets.

Having both the out-of-band and in-band techniques combined also has a big advantage, i.e., it could be used as an indirect means of "authenticating" the Probe Description URI in the in-band probe, thanks to a correlation with the out-of-band technique (e.g., a reverse DNS lookup). While the out-of-band technique alone is less prone to spoofing, the combination with the in-band technique offers a more complete solution.

6. Ethical Considerations

Executing measurement experiences over the global Internet obviously requires ethical consideration, which is discussed in [ANRW_PAPER], especially when unsolicited transit or destination parties are involved.

This document proposes a common way to identify the source and the purpose of active probing in order to reduce the potential burden on the unsolicited parties.

But there are other considerations to be taken into account, from the payload content (e.g., is the encoding valid?) to the transmission rate (see also [IPV6_TOPOLOGY] and [IPV4_TOPOLOGY] for some probing speed impacts). Those considerations are out of scope of this document.

7. Security Considerations

This document proposes simple techniques for probe attribution. It is expected that only ethical researchers would use them, which would simplify and reduce the time to identify probes across the Internet. In fact, these techniques could be used by anyone, malicious or not, which means that the information obtained cannot be blindly trusted. Using these techniques should not mean that a probe can be trusted. Instead, third parties should use this solution to potentially understand the origin and context of such probes. This solution is not perfect, but it provides a way for probe attribution, which is better than no solution at all.

Probe attribution is provided to identify the source and intent of specific probes, but there is no authentication possible for the inline information. Therefore, a malevolent actor could provide false information while conducting the probes or spoof them so that the action is attributed to a third party. In that case, not only would this third party be wrongly accused, but it might also be exposed to unwanted solicitations (e.g., angry emails or phone calls if the malevolent actor used someone else's email address or phone number). As a consequence, the recipient of this information cannot trust it without confirmation. If a recipient cannot confirm the information or does not wish to do so, it should treat the flows as if there were no probe attribution. Note that using probe attribution does not create a new DDoS vector since there is no expectation that third parties would automatically confirm the information obtained.

As the Probe Description URI is transmitted in the clear and as the Probe Description File is publicly readable, Personally Identifiable Information (PII) should not be used for an email address and a phone number; a generic or group email address and phone number should be preferred. Also, the Probe Description File could contain malicious data (e.g., links) and therefore should not be blindly trusted.

8. IANA Considerations

IANA has added the following URI suffix to the "Well-Known URIs" registry in accordance with [RFC8615]:

URI Suffix: probing.txt

Change Controller: IETF

Reference: RFC 9511

Status: permanent

9. References

9.1. Normative References

   [RFC0768] Postel, J., "User Datagram Protocol", STD 6, RFC 768,
              DOI 10.17487/RFC0768, August 1980,
              <https://www.rfc-editor.org/info/rfc768>.

   [RFC0792] Postel, J., "Internet Control Message Protocol", STD 5,
              RFC 792, DOI 10.17487/RFC0792, September 1981,
              <https://www.rfc-editor.org/info/rfc792>.

   [RFC4443] Conta, A., Deering, S., and M. Gupta, Ed., "Internet
              Control Message Protocol (ICMPv6) for the Internet
              Protocol Version 6 (IPv6) Specification", STD 89,
              RFC 4443, DOI 10.17487/RFC4443, March 2006,
              <https://www.rfc-editor.org/info/rfc4443>.

   [RFC8200] Deering, S. and R. Hinden, "Internet Protocol, Version 6
              (IPv6) Specification", STD 86, RFC 8200,
              DOI 10.17487/RFC8200, July 2017,
              <https://www.rfc-editor.org/info/rfc8200>.

   [RFC8615] Nottingham, M., "Well-Known Uniform Resource Identifiers
              (URIs)", RFC 8615, DOI 10.17487/RFC8615, May 2019,
              <https://www.rfc-editor.org/info/rfc8615>.

   [RFC9116] Foudil, E. and Y. Shafranovich, "A File Format to Aid in
              Security Vulnerability Disclosure", RFC 9116,
              DOI 10.17487/RFC9116, April 2022,
              <https://www.rfc-editor.org/info/rfc9116>.

   [RFC9293] Eddy, W., Ed., "Transmission Control Protocol (TCP)",
              STD 7, RFC 9293, DOI 10.17487/RFC9293, August 2022,
              <https://www.rfc-editor.org/info/rfc9293>.

9.2. Informative References

   [ANRW_PAPER]
              Fiebig, T., "Crisis, Ethics, Reliability & a
              measurement.network - Reflections on Active Network
              Measurements in Academia", DOI 10.1145/3606464.3606483,
              July 2023,
              <https://pure.mpg.de/rest/items/item_3517635/component/
              file_3517636/content>.

   [IPV4_TOPOLOGY]
              Beverly, R., "Yarrp'ing the Internet: Randomized High-
              Speed Active Topology Discovery",
              DOI 10.1145/2987443.2987479, November 2016,
              <http://www.cmand.org/papers/yarrp-imc16.pdf>.

   [IPV6_TOPOLOGY]

Beverly, R., Durairajan, R., Plonka, D., and J. Rohrer, "In the IP of the Beholder: Strategies for Active IPv6 Topology Discovery", DOI 10.1145/3278532.3278559, October 2018, <http://www.cmand.org/papers/beholder-imc18.pdf>.

[JAMES]   Vyncke, É., Léas, R., and J. Iurman, "Just Another Measurement of Extension header Survivability (JAMES)", Work in Progress, Internet-Draft, draft-vyncke-v6ops-james-03, 9 January 2023, <https://datatracker.ietf.org/doc/html/draft-vyncke-v6ops-james-03>.

[LARGE_SCALE]
            Donnet, B., Raoult, P., Friedman, T., and M. Crovella, "Efficient Algorithms for Large-Scale Topology Discovery", DOI 10.1145/1071690.1064256, DOI 10.1145/1071690.1064256, June 2005, <https://dl.acm.org/doi/pdf/10.1145/1071690.1064256>.

[NCSC]   UK NCSC, "The National Cyber Security Centre", <https://www.ncsc.gov.uk/>.

[NCSC_SCAN_INFO]
            UK NCSC, "NCSC Scanning information", <https://www.ncsc.gov.uk/information/ncsc-scanning-information>.

[RFC4942]   Davies, E., Krishnan, S., and P. Savola, "IPv6 Transition/ Co-existence Security Considerations", RFC 4942, DOI 10.17487/RFC4942, September 2007, <https://www.rfc-editor.org/info/rfc4942>.

[RFC7872]   Gont, F., Linkova, J., Chown, T., and W. Liu, "Observations on the Dropping of Packets with IPv6 Extension Headers in the Real World", RFC 7872, DOI 10.17487/RFC7872, June 2016, <https://www.rfc-editor.org/info/rfc7872>.

[RIPE_ATLAS]
            RIPE Network Coordination Centre (RIPE NCC), "RIPE Atlas", <https://atlas.ripe.net/>.

[SCAPY]   "Scapy", <https://scapy.net/>.

Appendix A. Examples of In-Band Attribution

Here are several examples generated by [SCAPY] and displayed in the 'tcpdump' format:

IP packet with Probe Description URI inside a Destination Options extension header:

IP6 2001:db8:dead::1 > 2001:db8:beef::1: DSTOPT 60878 > traceroute: Flags [S], seq 0, win 8192, length 0

0x0000: 6000 0000 0044 3c40 2001 0db8 dead 0000 `....D<@........

```
0x0010:  0000 0000 0000 0001 2001 0db8 beef 0000   ................
0x0020:  0000 0000 0000 0001 0605 012c 6874 7470   ...........,http
0x0030:  733a 2f2f 6578 616d 706c 652e 6e65 742f   s://example.net/
0x0040:  2e77 656c 6c2d 6b6e 6f77 6e2f 7072 6f62   .well-known/prob
0x0050:  696e 672e 7478 7400 edce 829a 0000 0000   ing.txt.........
0x0060:  0000 0000 5002 2000 2668 0000             ....P...&h..
```

IP packet with the URI in the data payload of a TCP SYN:

```
IP6 2001:db8:dead::1.15581 > 2001:db8:beef::1.traceroute:
Flags [S], seq 0:23, win 8192, length 23

0x0000:  6000 0000 002b 0640 2001 0db8 dead 0000   `....+.@........
0x0010:  0000 0000 0000 0001 2001 0db8 beef 0000   ................
0x0020:  0000 0000 0000 0001 3cdd 829a 0000 0000   ........<.......
0x0030:  0000 0000 5002 2000 c9b7 0000 6d61 696c   ....P.......mail
0x0040:  746f 3a6c 6162 4065 7861 6d70 6c65 2e6e   to:lab@example.n
0x0050:  6574 00                                   et.
```

IP echo request with another URI in the data part of the ICMP
ECHO_REQUEST:

```
IP6 2001:db8:dead::1 > 2001:db8:beef::1: ICMP6, echo request, id 0,
seq 0, length 28

0x0000:  6000 0000 001c 3a40 2001 0db8 dead 0000   `.....:@........
0x0010:  0000 0000 0000 0001 2001 0db8 beef 0000   ................
0x0020:  0000 0000 0000 0001 8000 2996 0000 0000   ..........).....
0x0030:  7465 6c3a 2b31 2d32 3031 2d35 3535 2d30   tel:+1-201-555-0
0x0040:  3132 3300                                 123.
```

IPv4 echo request with a URI in the data part of the ICMP
ECHO_REQUEST:

```
IP 192.0.2.1 > 198.51.10.1: ICMP echo request, id 0, seq 0, length 31

0x0000:  4500 0033 0001 0000 4001 8e93 c000 0201   E..3....@.......
0x0010:  c633 0a01 0800 ea74 0000 0000 6d61 696c   .3d....t....mail
0x0020:  746f 3a6c 6162 4065 7861 6d70 6c65 2e6e   to:lab@example.n
0x0030:  6574 00                                   et.
```

Acknowledgments

Authors' Addresses

Éric Vyncke
Cisco
De Kleetlaan 6A
1831 Diegem
Belgium
Email: evyncke@cisco.com

Benoît Donnet
Université de Liège
Belgium
Email: benoit.donnet@uliege.be

Justin Iurman
Université de Liège
Belgium
Email: justin.iurman@uliege.be