INTERNET-DRAFT
Intended Status: Informational

N. Elkins
Inside Products
M. Ackermann
BCBS Michigan
K. Haining
US Bank
S. Perdomo
DTCC
W. Jouris
Inside Products
D. Boyes
Sine Nomine

Expires: November 30, 2013

May 30, 2013

IPv6 Packet Sequence Number Needed
draft-elkins-v6ops-ipv6-packet-sequence-needed-00

Abstract

For a number of Enterprise Data Center Operators (EDCO) both real-time and after the fact problem resolution is critical. Two metrics are critical for timely end-to-end problem resolution, without impacting an operational production network. They are: packet sequence number and packet timestamp. Packet sequence number is required for diagnostics. Packet timestamp is required to calculate end-to-end response time. Current methods are inadequate for these purposes because they assume unreasonable access to intermediate devices, are cost prohibitive, require infeasible changes to a running production network, or do not provide timely data. This document provides the background and rationale for the packet sequence number which is a part of the IPv6 Performance and Diagnostic Metrics Destination Option (PDM).

The list of Internet-Draft Shadow Directories can be accessed at
http://www.ietf.org/shadow.html

Table of Contents

1 Background

To diagnose problems for a number of Enterprise Data Center Operators (EDCO) two metrics are critical for timely end-to-end problem resolution, both real-time and after the fact, without impacting an operational production network. They are: packet sequence number and packet timestamp. Packet sequence number is required for diagnostics. Packet timestamp is required to calculate end-to-end response time.

This document provides the background and rationale for the packet sequence number which is a part of the IPv6 Performance and Diagnostic Metrics destination option (PDM).

For background, please see Draft-Elkins-6MAN-IPv6-PDM-Dest-Option-00 [PDMELK], Draft-Elkins-End-To-End-Response-Time-00 [RSPELK], and Draft-Elkins-PDM-Recommended-Usage-00 [USEELK]. These drafts are companion documents to this document. All four documents should be read together.

As discussed in the above Internet Drafts, current methods are inadequate for these purposes because they assume unreasonable access to intermediate devices, are cost prohibitive, require infeasible changes to a running production network, or do not provide timely data. The IPv6 Performance and Diagnostic Metrics destination option (PDM) provides a solution to these problems. This document will detail the background and need for the packet sequence number.

1.1 Why Packet Sequence Number

In many EDCO networks, during network diagnostics of an end-to-end connection, it becomes necessary to find the device along the network path creating problems. Diagnostic data may be collected at multiple places along the path (if possible), or at the source and destination. Then, the diagnostic data must be matched. Packet sequence number is critical in this matching process. The timestamp or even the IP addresses may be different at different devices. In IPv4 networks, the IPID field was used as a de facto sequence number. This will be discussed at greater length in section 1.2.

This method of data collection along the path is of special use on large multi-tier networks to determine where packet loss or packet corruption is happening. Multi-tier networks are those which have multiple routers or switches on the path between the sender and the receiver.

1.2 IPv4 IPID : DeFacto Sequence Number

With IPv4 networks, on many stack implementations, but not all, the IPID field has the property of sequentiality.

1.2.1 Description of IPID in IPv4

In IPv4, the 16 bit IP Identification (IPID) field is located at an offset of 4 bytes into the IPv4 header and is described in RFC0791 [RFC0791]. In IPv6, the IPID field is a 32-bit field contained in the Fragment Header defined by section 4.5 of RFC2460 [RFC2460]. Unfortunately, unless fragmentation is being done by the source node, the IPv6 packet will not contain this Fragment Header, and therefore will have no Identification field.

The intended purpose of the IPID field, in both IPv4 and IPv6, is to enable fragmentation and reassembly, and as currently specified is required to be unique within the maximum segment lifetime (MSL) on all datagrams. The MSL is often 2 minutes.

1.2.2 DeFacto Use of IPID

In many EDCO networks, the IPID field is used for more than fragmentation. During network diagnostics, packet traces may be taken at multiple places along the path, or at the source and destination. Then, packets can be matched by looking at the IPID.

The inclusion of the IPID makes it easier for a device(s) in the middle of the network, or on the receiving end of the network, to identify flows belonging to a single node, even if that node might have a different IP address. For example, in the case of sessions going through a NAT or proxy server.

For its de-facto diagnostic mode usage, the IPID field needs to be available whether or not fragmentation occurs. It also needs to be unique in the context of the session, and across all the connections controlled by the stack. In IPv4, the IPID is in the main header, so it is available for all packets. As it is a 16-bit field, it wrapped during the course of the session and thus had some limitations.

Even with these limitations, the IPID has been valuable and useful in IPv4 for diagnostics and problem resolution. It is a practical solution that is 'good enough' in many instances. Not having it available in IPv6, may be a major detriment to new IPv6 deployments and contribute to protracted downtimes in existing IPv6 operations.

1.2.3 Merits of DeFacto Usage

As network technology evolves, the uses to which fields are put can change as well. De-facto use is powerful, and should not be lightly ignored. In fact, it is a testament to the power and pervasiveness of the protocol that users create new uses for the original technology.

For example, the use of the IPID goes beyond the vision of the original authors. This sort of thing has happened with numerous other technologies and protocols.

The implementation of the traceroute command sends ICMP echo packets with a varying TTL. This is a very useful for diagnostics yet departs from the original purpose of TTL.

Similarly, cell phones have evolved to be more than just a means of vocal communication, including Internet communications, photo-sharing, stock exchange transactions, etc. Indeed, the Internet itself has evolved, from a small network for researchers and the military to share files into the pervasive global information superhighway that it is today.

1.2.4 Use Cases of IPv4 IPID in Diagnostics

Use Case # 1 --- Large Insurance Company

- (estimated time saved by use of IPID: 7 hours)

Performance Tool produces extraneous packets

- Issue was whether a performance tool was accurately replicating session flow during performance testing.
- Trace IPIDs showed more unique packets within same flow from performance tool compared to IE Browser.
- Having the clear IPID sequence numbers also showed where and why the extra packets were being generated.
- Solution: Problem rectified in subsequent version of performance tool.
- Without IPID, it was not clear if there was an issue at all.

Use Case #2 --- Large Bank

- (estimated time saved by use of IPID: 4 hours)

Batch transfer duration increases 12x

- A data transfer which formerly took 30 minutes to complete started taking 6-8 hours to complete.
- Was there packet loss? All the vendors said no.
- The other applications on the network did not report any

problems.
- 4 trace points were used, and the IPIDs in the packets were compared.
- The comparison showed 7% packet loss.
- Solution: WAN hardware was replaced and problem fixed.
- Without IPID, no one would agree a problem existed

Use Case #3 --- Large Bank
- (estimated time saved by use of IPID: 6 hours)

Very slow interactive performance

- All network links looked good.
- Traces showed duplicated small packets (which can be OK).
- We saw that the IPID was the same in both packets but the TTL was always + 1.
- A network device was "splitting" only small packets over two interfaces.
- The small packets were control info, telling other side to slow down.
- It erroneously looked like network congestion.
- Solution: Network device replaced and good interactive performance restored.
- Without IPID, flows would have appeared OK.

Use Case #4 --- Large Government Agency
- (estimated time saved by use of IPID: 9 hours)

VPN drops

- Cell phone connections to law enforcement were being dropped. The connections were going through a VPN.
- All parties (both sides of VPN connection, application, etc.) said it was not their problem. The problem went on for weeks.
- Finally, we took a trace which showed packets with IPID and TTL that did not match others in the flow AT ALL coming from the router nearest the application server end of VPN.
- Solution: Provider for VPN for application server changed. Problem resolved.
- Without IPID, much harder to diagnose problem.
- (Same case also happened with large corporation. Again, all parties saying not their fault until proven via packet trace.)

1.3 TCP sequence number is not enough

TCP Sequence number is defined in RFC0793 [RFC0793]. Some have proposed that this field will meet the needs of EDCO networks for a packet sequence number. Indeed, the TCP Sequence Number along with

the TCP Acknowledgment number can be used to calculate dropped packets, duplicate packets, out-of-order packets etc. That is, IF the packet flow itself reflects accurately what happened on the wire!

See Scenario 1 (Section 1.5.2) and Scenario 2 (Section 1.5.3) for what happens with packet trace capture in real networks.

The TCP Sequence Number is, obviously, available only for TCP and not other transport protocols.

1.4 Inadequacy of current measurement techniques

The question arises of whether current methods of instrumentation cannot be used without a change to the protocol. Current methods of measuring network data, other than packet traces, are inadequate because they assume unreasonable access to intermediate devices, are cost prohibitive, require infeasible changes to a running production network, or do not provide timely data. This section will discuss each of these in detail.

Current methods include both instrumentation and third party products. These include SNMP, CMIP, router logs, and firewall logs.

1.4.1 SNMP / CMIP Counters

The traditional network performance counters measured by SNMP or CMIP do not provide information at the granularity desired on the behavior of application flows across the network. The problem is that such counters do not contain enough data be able to provide a detailed and realistic view of the end-to-end behavior of a connection.

1.4.2 Router / Firewall Logs

Router and firewall logs may provide some information for diagnostics But as discussed in section 1.4.5, routers and firewalls in a production network are generally set to do minimal logging and diagnostics to allow maximum efficiency and throughput. Such devices cannot be asked to collect detailed data for an operational problem, as this requires a change to a production network.

1.4.3 Netflow

Netflow is instrumentation which is available from some middle devices. As discussed in detail in section 1.4.5, such devices are generally set to do minimal logging and diagnostics to allow maximum efficiency and throughput.

Correlations to produce some level of response time data may be possible from Netflow. But, it is not an adequate picture of end-to-end response time as Netflow is in an intermediate device and is not in a position to know what has happened at a client.

1.4.4 Access to Intermediate Devices

The above current methods require access to the transport infrastructure - that is, the routers, switches or other intermediate devices. In some cases, this is possible; in others, the connections in question may cross a number of administrative entities (both in the transport and in the endpoints). When it is the enterprise at the endpoint which is interested in the diagnostics, the administrative entities who own the devices in the middle of the path have no stake in operational measurement at the enterprise or application level. They have no reason to provide the necessary data or to impact the basic transport with the instrumentation necessary to capture flow-oriented data as a continuous stream suitable for general consumption.

In other words, if you don't own the path end-to-end, you will not be able to get the data you need if you are required to get it from the devices in the middle. Not only that, the devices in the middle do not have the instrumentation necessary to make it easy to do end-to-end diagnostics because they are not responsible for that and so do not want to burden their devices with doing those kind of functions.

Many EDCO networks may not own the path end-to-end. They may be working with a business partner's network or crossing the Internet.

1.4.5 Modifications to an Operational Production Network

Even when the enterprise does own all the devices along the entire path, to get enough data to adequately resolve a problem means changing the device configuration to do detailed diagnostics. In a production network, devices are generally set to do minimal logging and diagnostics. This is to allow maximum efficiency and throughput. The more logging and diagnostics such devices do, the fewer resources they have for actually transmitting traffic across the network.

So, if devices are to be asked to collect more data for an operational problem, this requires a change to a production network. This is generally not possible as it destabilizes a critical network during business hours, thus potentially disrupting many customers. Making changes is usually a lengthy process requiring change control, testing on a test network, etc. On networks which are critical to the business function, such as the networks we are discussing, it is

hardly likely that changing configuration "in flight" is an option.

2 Solution Parameters

What is needed is:

1) A method to identify and/or track the behavior of a connection without assuming access to the transport devices.

2) A method to observe a connection in flight without introducing agents at endpoints.

3) a method to observe arbitrary flows at multiple points within a network and correlate the results of those observations in a consistent manner.

4) A method to signal and correlate transport issues to application end-to-end behavior.

5) A method which does not require changes to a production network in real time.

6) Adequate granularity in the measurement technique to provide the needed metrics.

2.1 Packet Trace Meets Criteria

The only instrumentation which provides enough detail to diagnose end-to-end problems is a packet trace. Packet traces do not require changes to devices in production mode because in many large EDCO networks, products are available to capture packets in passive mode. Such products continuously monitor network traffic. Often, they are used not for diagnostic reasons but for regulatory reasons. For example, there may be legal requirements to log all stock exchange transactions.

Products for packet tracing are available freely and can be used at a client host without disrupting major portions of the network.

2.1.1 Limitations of Packet Capture

Even though packets are the only reliable way to provide data at the needed granularity, there are limitations with collecting packet traces in some situations. They are as follows:

2.1.2 Problem Scenario 1

1. Packets are captured for analysis at places like large core

switches. All packets are kept. Again, not necessarily for diagnostic reasons but for regulatory ones. For example, records of all stock trades may need to be kept for a certain number of years.

2. When there is a problem, an analyst extracts the needed information.

3. If the extract is done incorrectly, as often happens, or the packet capture itself is incorrect, then there may be false duplicate packets which can be quite misleading and can lead to wrong conclusions. Are these real TCP duplicates? Is there congestion on the subnet? Are these retransmissions? Situations have been seen where routers incorrectly send two packets instead of one - is this such a situation?

2.1.3 Problem Scenario 2

1. In this scenario, packets are captured for analysis at places like a middleware box. It may be because problems are suspected with the box itself or it is a central point of the suspected failure.

2. The box may not offer any way to tailor the packet capture. "You will get what we give you, how we give it to you!" is their philosophy.

3. The packet capture incorrectly duplicates only packets going to certain nodes.

4. Again, there are false duplicate packets which can be misleading and can lead to wrong conclusions. Are these real TCP duplicates? Is there congestion on the subnet? Situations have been seen where routers incorrectly send two packets instead of one - is this such a situation?

3 Rationale for Proposed Solution (PDM)

The current IPv6 specification does not provide a packet sequence number or similar field in the IPv6 main header. One option might be to force all IPv6 packets to contain a Fragment Header. In packets which are entire in and of themselves, the fragment ID would be zero - that is, an atomic fragment. Why was a new destination option header defined rather than recommending that Fragment Header be used?

Our reasoning was that the PDM destination option header would provide multiple benefits : the packet sequence number and the timestamp to calculate response time. See Draft-Elkins-End-To-End-Response-Time-Needed-00 [RSPELK].

4 Backward Compatibility

The scheme proposed in this document is backward compatible with all the currently defined IPv6 extension headers. According to RFC2460 [RFC2460], if the destination node does not recognize this option, it should skip over this option and continue processing the header.

5 Security Considerations

No security considerations are seen.

6 IANA Considerations

There are no IANA considerations.

7 References

7.1 Normative References

[RFC0791] Postel, J., "Internet Protocol", STD 5, RFC 791, September 1981.

[RFC0793] Postel, J., "Transmission Control Protocol", STD 7, RFC 793, September 1981.

[RFC2460] Deering, S. and R. Hinden, "Internet Protocol, Version 6 (IPv6) Specification", RFC 2460, December 1998.

[PDMELK] Elkins, N., "Draft-Elkins-IPv6-PDM-Dest-Option-00", Internet Draft, May 2013.

[RSPELK] Elkins, N., "Draft-Elkins-End-To-End-Response-Time-00", Internet Draft, May 2013

[USEELK] Elkins, N., "Draft-Elkins-PDM-Recommended-Usage-00", Internet Draft, May 2013

8 Acknowledgments

The authors would like to thank Rick Troth and Fred Baker for their comments.

Authors' Addresses

Nalini Elkins
Inside Products, Inc.
36A Upper Circle
Carmel Valley, CA 93924
United States
Phone: +1 831 659 8360
Email: nalini.elkins@insidethestack.com
http://www.insidethestack.com

Michael S. Ackermann
Blue Cross Blue Shield of Michigan
P.O. Box 2888
Detroit, Michigan 48231
United States
Phone: +1 310 460 4080
Email: mackermann@bcbsmi.com
http://www.bcbsmi.com

Keven Haining
US Bank
16900 W Capitol Drive
Brookfield, WI 53005
United States
Phone: +1 262 790 3551
Email: keven.haining@usbank.com
http://www.usbank.com

Sigfrido Perdomo
Depository Trust and Clearing Corporation
55 Water Street
New York, NY 10055
United States
Phone: +1 917 842 7375
Email: s.perdomo@dtcc.com
http://www.dtcc.com

William Jouris
Inside Products, Inc.
36A Upper Circle
Carmel Valley, CA 93924
United States
Phone: +1 925 855 9512
Email: bill.jouris@insidethestack.com

http://www.insidethestack.com

David Boyes
Sine Nomine Associates
43596 Blacksmith Square
Ashburn, VA 20147
United States
Phone: +1 703 723 6673
dboyes@sinenomine.net
http://www.sinenomine.net