Network Working Group
INTERNET-DRAFT
Expires in: May 2008
Intended Status: Informational

Scott Poretsky
Reef Point Systems

Brent Imhoff
Juniper Networks

November 2007

Terminology for Benchmarking Link-State IGP Data Plane Route Convergence

<draft-ietf-bmwg-igp-dataplane-conv-term-14.txt>

ABSTRACT
This document describes the terminology for benchmarking Interior Gateway Protocol (IGP) Route Convergence. The terminology is to be used for benchmarking IGP convergence time through externally observable (black box) data plane measurements. The terminology can be applied to any link-state IGP, such as ISIS and OSPF.

Table of Contents

1. Introduction

This draft describes the terminology for benchmarking Interior Gateway Protocol (IGP) Route Convergence. The motivation and applicability for this benchmarking is provided in [Po07a]. The methodology to be used for this benchmarking is described in [Po07m]. The methodology and terminology to be used for benchmarking Route Convergence can be applied to any link-state IGP such as ISIS [Ca90] and OSPF [Mo98]. The data plane is measured to obtain black-box (externally observable) convergence benchmarking metrics. The purpose of this document is to introduce new terms required to complete execution of the IGP Route Convergence Methodology [Po07m]. These terms apply to IPv4 and IPv6 traffic and IGPs.

An example of Route Convergence as observed and measured from the data plane is shown in Figure 1. The graph in Figure 1 shows Forwarding Rate versus Time. Time 0 on the X-axis is on the far right of the graph. The Offered Load to the ingress interface of the DUT SHOULD equal the measured maximum Throughput [Ba99][Ma98] of the DUT and the Forwarding Rate [Ma98] is measured at the egress interfaces of the DUT. The components of the graph and the metrics are defined in the Term Definitions section.

```
                     Convergence    Convergence
                     Recovery       Event
                     Instant        Instant  Time = 0sec
Forwarding Rate =       ^              ^       ^     Offered Load =
  Offered Load --> ------\     Packet   /-------- <---Max Throughput
                          \     Loss   /<----Convergence
        Convergence------->\          /      Event Transition
     Recovery Transition    \        /
                             \_____/<------Maximum Packet Loss

     X-axis = Time
     Y-axis = Forwarding Rate
```

Figure 1. Convergence Graph

2. Existing definitions

This document uses existing terminology defined in other BMWG work. Examples include, but are not limited to:

Latency	[Ref.[Ba91], section 3.8]
Frame Loss Rate	[Ref.[Ba91], section 3.6]
Throughput	[Ref.[Ba91], section 3.17]
Device Under Test (DUT)	[Ref.[Ma98], section 3.1.1]
System Under Test (SUT)	[Ref.[Ma98], section 3.1.2]
Out-of-order Packet	[Ref.[Po06], section 3.3.2]
Duplicate Packet	[Ref.[Po06], section 3.3.3]
Packet Reordering	[Ref.[Mo06], section 3.3]

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in BCP 14, RFC 2119 [Br97]. RFC 2119 defines the use of these key words to help make the intent of standards track documents as clear as possible. While this document uses these keywords, this document is not a standards track document.

3. Term Definitions

3.1 Convergence Event

Definition:
The occurrence of a planned or unplanned action in the network that results in a change in the egress interface of the Device Under Test (DUT) for routed packets.

Discussion:
Convergence Events include link loss, routing protocol session loss, router failure, configuration change, and better next-hop learned via a routing protocol.

Measurement Units:
N/A

Issues:
None

See Also:
Convergence Packet Loss
Convergence Event Instant

3.2 Route Convergence

Definition:
Route Convergence is the action to update all components of the router with the most recent route change(s) including the Routing Information Base (RIB) and Forwarding Information Base (FIB), along with software and hardware tables, such that forwarding is successful for one or more destinations.

Discussion:
Route Convergence MUST occur after a Convergence Event. Route Convergence can be observed externally by the rerouting of data traffic to the Next-best Egress Interface. Also, Route Convergence may or may not be sustained over time.

Measurement Units:
N/A

Issues:
None

See Also:
Network Convergence
Full Convergence
Convergence Event

3.3 Network Convergence

Definition:
The completion of updating of all routing tables, including distributed FIBs, in all routers throughout the network.

Discussion:
Network Convergence requires completion of all Route Convergenceoperations for all routers in the network following a Convergence Event. Network Convergence can be observed by recovery of System Under Test (SUT) Throughput to equal the offered load, with no Stale Forwarding, and no Blenders [Ca01][Ci03].

Measurement Units:
N/A

Issues:
None

See Also:
Route Convergence
Stale Forwarding

3.4 Full Convergence

Definition:
Route Convergence for an entire FIB in which complete recovery from the Convergence Event is indicated by the DUT Throughput equal to the offered load.

Discussion:
When benchmarking convergence, it is useful to measure the time to converge an entire FIB. For example, a Convergence Event can be produced for an OSPF table of 5000 routes so that the time to converge routes 1 through 5000 is measured. Full Convergence is externally observable from the data plane when the Throughput of the data plane traffic on the Next-Best Egress Interface equals the offered load.

Measurement Units:
N/A

Issues:
None

See Also:
Network Convergence

Route Convergence
Convergence Event

3.5 Packet Loss

Definition:
The number of packets that should have been forwarded by a DUT under a constant offered load that were not forwarded due to lack of resources.

Discussion:
Packet Loss is a modified version of the term "Frame Loss Rate" as defined in [Ba91]. The term "Frame Loss" is intended for Ethernet Frames while "Packet Loss" is intended for IP packets. Packet Loss can be measured as a reduction in forwarded traffic from the Throughput [Ba91] of the DUT.

Measurement units:
Number of offered packets that are not forwarded.

Issues: None

See Also:
Convergence Packet Loss

3.6 Convergence Packet Loss

Definition:
The number of packets lost due to a Convergence Event until Full Convergence occurs.

Discussion:
Convergence Packet Loss includes packets that were lost and packets that were delayed due to buffering. The Convergence Packet Loss observed in a Packet Sampling Interval may or may not be equal to the number of packets in the offered load during the interval following a Convergence Event (see Figure 1).

Measurement Units:
number of packets

Issues: None

See Also:
Packet Loss
Route Convergence
Convergence Event
Packet Sampling Interval

3.7 Convergence Event Instant

Definition:
The time instant that a Convergence Event becomes observable in the data plane.

Discussion:
Convergence Event Instant is observable from the data plane as the precise time that the device under test begins to exhibit packet loss.

Measurement Units:
hh:mm:ss:nnn:uuu,
where 'nnn' is milliseconds and 'uuu' is microseconds.

Issues:
None

See Also:
Convergence Event
Convergence Packet Loss
Convergence Recovery Instant

3.8 Convergence Recovery Instant

Definition:
The time instant that Full Convergence is measured and then maintained for an interval of duration equal to the Sustained Forwarding Convergence Time

Discussion:
Convergence Recovery Instant is measurable from the data plane as the precise time that the device under test achieves Full Convergence.

Measurement Units:
hh:mm:ss:nnn:uuu,
where 'nnn' is milliseconds and 'uuu' is microseconds.

Issues:
None

See Also:
Sustained Forwarding Convergence Time
Convergence Packet Loss
Convergence Event Instant

3.9 First Prefix Convergence Instant

Definition:
The time instant for convergence of a first route entry following a Convergence Event, as observed by receipt of the first packet from the Next-Best Egress Interface.

Discussion:
The First Prefix Convergence Instant is an indication that the process to achieve Full Convergence has begun. Any route may be the first to converge for First Convergence. Measurement on the data-plane enables First Convergence to be observed without any white-box information from the DUT.

Measurement Units:
N/A

Issues:
None

See Also:
Route Convergence
Full Convergence
Stale Forwarding

3.10 Convergence Event Transition

Definition:
A time interval observed following a Convergence Event in which Throughput gradually reduces to zero.

Discussion:
The Convergence Event Transition is best observed for Full Convergence. The egress packet rate observed during a Convergence Event Transition may not decrease linearly. Both the offered load and the Packet Sampling Interval influence the observations of the Convergence Event Transition. For example, even if the Convergence Event were to cause the Throughput [Ba91] to drop to zero there would be some number of packets observed, unless the Packet Sampling Interval is exactly aligned with the Convergence Event. This is further discussed with the term "Packet Sampling Interval".

Measurement Units:
seconds

Issues:
None

See Also:
Convergence Event
Full Convergence
Packet Sampling Interval

3.11 Convergence Recovery Transition

Definition:
The characteristic of the DUT in which Throughput gradually increases to equal the offered load.

Discussion:
The Convergence Recovery Transition is best observed for Full Convergence. The egress packet rate observed during a Convergence Recovery Transition may not increase linearly. Both the offered load and the Packet Sampling Interval influence the observations of the Convergence Recovery Transition. This is further discussed with the term "Packet Sampling Interval".

Measurement Units:
seconds

Issues: None

See Also:
Full Convergence
Packet Sampling Interval

3.12 Rate-Derived Convergence Time

Definition:
The amount of time for Convergence Packet Loss to persist upon occurrence of a Convergence Event until measurement of Full Convergence.

Rate-Derived Convergence Time can be measured as the time difference from the Convergence Event Instant to the Convergence Recovery Instant, as shown with Equation 1.

(Equation 1)
Rate-Derived Convergence Time =
Convergence Recovery Instant - Convergence Event Instant.

Discussion:
Rate-Derived Convergence Time should be measured at the maximum Throughput of the DUT. At least one packet per route in the FIB for all routes in the FIB MUST be offered to the DUT per second. Failure to achieve Full Convergence results in a Rate-Derived Convergence Time benchmark of infinity.

Measurement Units:
seconds

Issues:
None

See Also:
Convergence Packet Loss
Convergence Recovery Instant
Convergence Event Instant
Full Convergence

3.13 Loss-Derived Convergence Time

Definition:
The amount of time it takes for Full Convergence to be achieved as calculated from the amount of Convergence Packet Loss. Loss-Derived Convergence Time can be calculated from Convergence Packet Loss that occurs due to a Convergence Event and Route Convergence as shown with Equation 2.

```
Equation 2 -
  Loss-Derived Convergence Time =
        Convergence Packets Loss / Offered Load
        NOTE: Units for this measurement are
        packets / packets/second = seconds
```

Discussion:
Loss-Derived Convergence Time gives a better than actual result when converging many routes simultaneously. Rate-Derived Convergence Time takes the Convergence Recovery Transition into account, but Loss-Derived Convergence Time ignores the Route Convergence Recovery Transition because it is obtained from the measured Convergence Packet Loss.

Ideally, the Convergence Event Transition and Convergence Recovery Transition are instantaneous so that the Rate-Derived Convergence Time = Loss-Derived Convergence Time. However, router implementations are less than ideal. For these reasons the preferred reporting benchmark for IGP

Route Convergence is the Rate-Derived Convergence Time. Guidelines for reporting Loss-Derived Convergence Time are provided in [Po07m].

Measurement Units:
seconds

Issues:
None

See Also:
Convergence Event
Convergence Packet Loss
Rate-Derived Convergence Time
Convergence Event Transition
Convergence Recovery Transition

3.14 Sustained Forwarding Convergence Time

Definition:
The amount of time for which Full Convergence is maintained without additional packet loss.

Discussion:
The purpose of the Sustained Forwarding Convergence Time is to produce Convergence benchmarks protected against fluctuation in Throughput after Full Convergence is observed. The Sustained Forwarding Convergence Time to be used is calculated as shown in Equation 3.

Equation 3 -
Sustained Forwarding Convergence Time =
    C*(Convergence Packet Loss/Offered Load)

where,
a. units are packets/pps = sec and

b. C is a constant. The RECOMMENDED value for C is 5 as selected from working group consensus. This is similar to RFC 2544 [Ba99] which recommends waiting 2 seconds for residual frames to arrive and 5 seconds for DUT restabilization.

c. at least one packet per route in the FIB for all routes in the FIB MUST be offered to the DUT per second.

Measurement Units:
seconds

Issues: None

See Also:
Full Convergence

Convergence Recovery Instant

3.15 First Prefix Convergence Time

Definition:
The amount of time for Convergence Packet Loss until the convergence of a first route entry on the Next-Best Egress Interface, as indicated by the First Prefix Convergence Instant.

First Prefix Convergence Time can be measured as the time difference from the Convergence Event Instant and the First Prefix Convergence Instant, as shown with Equation 4.

```
(Equation 4)
  First Prefix Convergence Time =
      First Prefix Convergence Instant -
                Convergence Event Instant.
```

Discussion:
First Prefix Convergence Time should be measured at the maximum Throughput of the DUT. At least one packet per route in the FIB for all routes in the FIB MUST be offered to the DUT per second. Failure to achieve the First Prefix Convergence Instant results in a First Prefix Convergence Time benchmark of infinity.

Measurement Units:
hh:mm:ss:nnn:uuu,
   where 'nnn' is milliseconds and 'uuu' is microseconds.

Issues:
None

See Also:
Convergence Packet Loss
First Prefix Convergence Instant

3.16 Reversion Convergence Time

Definition:
The amount of time for the DUT to forward traffic from the Preferred Egress Interface, instead of the Next-Best Egress Interface, upon recovery from a Convergence Event.

Discussion:
Reversion Convergence Time is the amount of time for routes to converge to the original outbound port. This is achieved by recovering from the Convergence Event, such as restoring the failed link. Reversion Convergence Time is measured using the Rate-Derived Convergence Time calculation technique, as provided in Equation 1. It is possible to have the

Reversion Convergence Time differ from the Rate-Derived Convergence Time.

Measurement Units:
seconds

Issues:
None

See Also:
Preferred Egress Interface
Convergence Event
Rate-Derived Convergence Time

3.17 Packet Sampling Interval

Definition:
The interval at which the tester (test equipment) polls to make measurements for arriving packet flows.

Discussion:
Metrics measured at the Packet Sampling Interval MUST include Forwarding Rate and Convergence Packet Loss.

Measurement Units:
seconds

Issues:
Packet Sampling Interval can influence the Convergence Graph. This is particularly true when implementations achieve Full Convergence in less than 1 second. The Convergence Event Transition and Convergence Recovery Transition can become exaggerated when the Packet Sampling Interval is too long. This will produce a larger than actual Rate-Derived Convergence Time. The recommended value for configuration of the Packet Sampling Interval is provided in [Po07m].

See Also:
Convergence Packet Loss
Convergence Event Transition
Convergence Recovery Transition

3.18 Local Interface

Definition:
An interface on the DUT.

Discussion:
A failure of the Local Interface indicates that the failure occured directly on the DUT.

Measurement Units:
N/A

Issues:
None

See Also:
Neighbor Interface
Remote Interface

3.19 Neighbor Interface

Definition:
The interface on the neighbor router or tester that is directly linked to the DUT's Local Interface.

Discussion:
None

Measurement Units:
N/A

Issues:
None

See Also:
Local Interface
Remote Interface

3.20 Remote Interface

Definition:
An interface on a neighboring router that is not directly connected to any interface on the DUT.

Discussion:
A failure of a Remote Interface indicates that the failure occurred on an interface that is not directly connected to the DUT.

Measurement Units:
N/A

Issues:
None

See Also:
Local Interface
Neighbor Interface

3.21 Preferred Egress Interface

Definition:
The outbound interface from the DUT for traffic routed to the preferred next-hop.

Discussion:
The Preferred Egress Interface is the egress interface prior to a Convergence Event.

Measurement Units:
N/A

Issues:
None

See Also:
Next-Best Egress Interface

3.22 Next-Best Egress Interface

Definition:
The outbound interface from the DUT for traffic routed to the second-best next-hop. It is the same media type and link speed as the Preferred Egress Interface

Discussion:
The Next-Best Egress Interface becomes the egress interface after a Convergence Event.

Measurement Units:
N/A

Issues:
None

See Also:
Preferred Egress Interface

3.23 Stale Forwarding

Definition:
Forwarding of traffic to route entries that no longer exist or to route entries with next-hops that are no longer preferred.

Discussion:
Stale Forwarding can be caused by a Convergence Event and is also known as a "black-hole" or microloop since it may produce

packet loss. Stale Forwarding exists until Network Convergence
is achieved.

Measurement Units:
N/A

Issues:
None

See Also:
Network Convergence

3.24 Nested Convergence Events

Definition:
The occurrence of a Convergence Event while the route table is converging from a prior Convergence Event.

Discussion:
The Convergence Events for a Nested Convergence Event MUST occur with different neighbors. A common observation from a Nested Convergence Event will be the withdrawal of routes from one neighbor while the routes of another neighbor are being installed.

Measurement Units:
N/A

Issues:
None

See Also:
Convergence Event

4. IANA Considerations

This document requires no IANA considerations.

5. Security Considerations

Documents of this type do not directly affect the security of Internet or corporate networks as long as benchmarking is not performed on devices or systems connected to production networks.

6. Acknowledgements

Thanks to Sue Hares, Al Morton, Kevin Dubray, Ron Bonica, David Ward, and the BMWG for their contributions to this work.

7. References

7.1 Normative References

[Ba91] Bradner, S. "Benchmarking Terminology for Network Interconnection Devices", RFC1242, July 1991.

[Ba99] Bradner, S. and McQuaid, J., "Benchmarking Methodology for Network Interconnect Devices", RFC 2544, March 1999.

[Br97] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", RFC 2119, March 1997

[Ca90] Callon, R., "Use of OSI IS-IS for Routing in TCP/IP and Dual Environments", RFC 1195, December 1990.

[Ma98] Mandeville, R., "Benchmarking Terminology for LAN Switching Devices", RFC 2285, February 1998.

[Mo98] Moy, J., "OSPF Version 2", RFC 2328, IETF, April 1998.

[Mo06] Morton, A., et al, "Packet Reordering Metrics", RFC 4737, November 2006.

[Po06] Poretsky, S., et al., "Terminology for Benchmarking Network-layer Traffic Control Mechanisms", RFC 4689, November 2006.

[Po07a] Poretsky, S., "Benchmarking Applicability for Link-State IGP Data Plane Route Convergence", draft-ietf-bmwg-igp-dataplane-conv-app-14, work in progress, November 2007.

[Po07m] Poretsky, S. and Imhoff, B., "Benchmarking Methodology for Link-State IGP Data Plane Route Convergence", draft-ietf-bmwg-igp-dataplane-conv-meth-14, work in progress, November 2007.

7.2 Informative References

[Ca01] S. Casner, C. Alaettinoglu, and C. Kuan, "A Fine-Grained View of High Performance Networking", NANOG 22, June 2001.

[Ci03] L. Ciavattone, A. Morton, and G. Ramachandran, "Standardized Active Measurements on a Tier 1 IP Backbone", IEEE Communications Magazine, pp90-97, May 2003.

8. Author's Address

Scott Poretsky
Reef Point Systems
3 Federal Street
Billerica, MA 01821
USA
Phone: + 1 508 439 9008
EMail: sporetsky@reefpoint.com

Brent Imhoff
Juniper Networks
1194 North Mathilda Ave
Sunnyvale, CA 94089
USA
Phone: + 1 314 378 2571
EMail: bimhoff@planetspork.com

such proprietary rights by implementers or users of this specification can be obtained from the IETF on-line IPR repository at http://www.ietf.org/ipr.

Acknowledgement

Funding for the RFC Editor function is currently provided by the Internet Society.