

Network Working Group
Internet-Draft
Intended status: Informational
Expires: November 30, 2012

R. Papneja
Huawei Technologies
S. Vapiwala
J. Karthik
Cisco Systems
S. Poretsky
Allot Communications
S. Rao
Qwest Communications
JL. Le Roux
France Telecom
May 29, 2012

Methodology for Benchmarking MPLS-TE Fast Reroute Protection
draft-ietf-bmwg-protection-meth-10.txt

Abstract

This document describes the methodology for benchmarking MPLS Protection mechanisms for link and node protection as defined in [RFC4090]. This document provides test methodologies and testbed setup for measuring failover times while considering all dependencies that might impact faster recovery of real-time applications bound to MPLS traffic engineered (MPLS-TE) tunnels.

Status of this Memo

This Internet-Draft is submitted in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at <http://datatracker.ietf.org/drafts/current/>.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

This Internet-Draft will expire on November 30, 2012.

Copyright Notice

Copyright (c) 2012 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust's Legal Provisions Relating to IETF Documents (<http://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

This document may contain material from IETF Documents or IETF Contributions published or made publicly available before November 10, 2008. The person(s) controlling the copyright in some of this material may not have granted the IETF Trust the right to allow modifications of such material outside the IETF Standards Process. Without obtaining an adequate license from the person(s) controlling the copyright in such materials, this document may not be modified outside the IETF Standards Process, and derivative works of it may not be created outside the IETF Standards Process, except to format it for publication as an RFC or to translate it into languages other than English.

Table of Contents

1. Introduction	5
2. Document Scope	6
3. Existing Definitions and Requirements	6
4. General Reference Topology	7
5. Test Considerations	8
5.1. Failover Events	8
5.2. Failure Detection	9
5.3. Use of Data Traffic for MPLS Protection benchmarking	9
5.4. LSP and Route Scaling	10
5.5. Selection of IGP	10
5.6. Restoration and Reversion	10
5.7. Offered Load	11
5.8. Tester Capabilities	11
5.9. Failover Time Measurement Methods	12
6. Reference Test Setup	12
6.1. Link Protection	13
6.1.1. Link Protection - 1 hop primary (from PLR) and 1 hop backup TE tunnels	13
6.1.2. Link Protection - 1 hop primary (from PLR) and 2 hop backup TE tunnels	14
6.1.3. Link Protection - 2+ hops (from PLR) primary and 1 hop backup TE tunnels	14
6.1.4. Link Protection - 2+ hop (from PLR) primary and 2 hop backup TE tunnels	15
6.2. Node Protection	16
6.2.1. Node Protection - 2 hop primary (from PLR) and 1 hop backup TE tunnels	16
6.2.2. Node Protection - 2 hop primary (from PLR) and 2 hop backup TE tunnels	17
6.2.3. Node Protection - 3+ hop primary (from PLR) and 1 hop backup TE tunnels	18
6.2.4. Node Protection - 3+ hop primary (from PLR) and 2 hop backup TE tunnels	19
7. Test Methodology	20
7.1. MPLS FRR Forwarding Performance	21
7.1.1. Headend PLR Forwarding Performance	21
7.1.2. Mid-Point PLR Forwarding Performance	22
7.2. Headend PLR with Link Failure	23
7.3. Mid-Point PLR with Link Failure	25
7.4. Headend PLR with Node Failure	26
7.5. Mid-Point PLR with Node Failure	28
8. Reporting Format	29
9. Security Considerations	30
10. IANA Considerations	31
11. Acknowledgements	31
12. References	31

12.1. Informative References	31
12.2. Normative References	31
Appendix A. Fast Reroute Scalability Table	32
Appendix B. Abbreviations	35
Authors' Addresses	36

1. Introduction

This document describes the methodology for benchmarking MPLS based protection mechanisms. This document uses much of the terminology defined in [RFC6414].

MPLS based protection mechanisms provide fast recovery of real-time services from a planned or an unplanned link or node failures. MPLS protection mechanisms are generally deployed in a network infrastructure where MPLS is used for provisioning of point-to-point traffic engineered tunnels (tunnel). MPLS based protection mechanisms promise to reduce service disruption period by minimizing recovery time from most common failures.

Network elements from different manufacturers behave differently to network failures, which impacts the network's ability and failure recovery performance. It therefore becomes imperative for service providers to have a common benchmark to verify the performance behaviors of these network elements.

There are two factors impacting service availability: frequency of failures and duration for which the failures persist. Failures can be classified into two types: correlated and uncorrelated. Correlated or uncorrelated failures may be planned or unplanned.

Planned failures are predictable. Network implementations should be able to handle both planned and unplanned failures and recover gracefully within a time period acceptable to maintain service assurance. Hence, failover recovery time is one of the most important benchmark that a service provider considers in choosing a the building blocks for their network infrastructure.

A correlated failure is the simultaneous occurrence of two or more failures. A typical example is failure of a logical resource (e.g. layer-2 links) due to a dependency on a common physical resource (e.g. common conduit) that fails. Within the context of MPLS-TE protection mechanisms, failures that arise due to Shared Risk Link Groups (SRLG) [RFC4090] can be considered as correlated failures.

MPLS Fast Re-Route (MPLS-FRR) allows for the possibility that the Label Switched Paths tunnels can be re-optimized following the Failover. IP Traffic would be re-routed according to the preferred path according to the post-failure topology. Hence, MPLS-FRR may include additional steps following the occurrence of the failure detection [RFC6414] and failover event [RFC6414].

- (1) Failover Event - Primary Path (Working Path) fails
- (2) Failure Detection- Failover Event is detected
- (3)
 - a. Failover - Working Path switched to Backup path
 - b. Re-Optimization of Working Path (possible change from Backup Path)
- (4) Restoration [RFC6414]
- (5) Reversion [RFC6414]

2. Document Scope

This document provides detailed test cases along with different topologies and scenarios that should be considered to effectively benchmark MPLS-TE protection mechanisms and failover times. Different failover events and scaling considerations are also provided in this document.

All benchmarking test-cases defined in this document apply to facility backup method [RFC4090]. The test cases cover all possible failure scenarios to benchmark the performance of the Device Under Test (DUT) to recover from failures. Data plane traffic is used to benchmark failover times.

Benchmarking of correlated failures is out of scope of this document. Faster failure detection using Bi-directional Forwarding Detection (BFD) is outside the scope of this document, but is mentioned in the discussion sections.

The Performance benchmarking of control plane is outside the scope of this benchmarking.

As described above, MPLS-FRR may include a Re-optimization of the Working Path. Characterization of Re-optimization is beyond the scope of this memo.

3. Existing Definitions and Requirements

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 [RFC2119]. 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.

The reader is assumed to be familiar with the commonly used MPLS terminology, some of which is defined in [RFC4090].

This document uses much of the terminology defined in [RFC6414]. This document also uses existing terminology defined in other BMWG Work [RFC1242], [RFC2285], [RFC4689].

4. General Reference Topology

Figure 1 illustrates the basic reference testbed and is applicable to all the test cases defined in this document. Tester comprises a Traffic Generator (TG), Test Analyzer (TA) and Emulator. The Tester is connected to the test network and based on test case, the DUT role could vary. The Tester (TG) sends and receives (TA) IP traffic to the tunnel ingress and performs signaling protocol emulation to simulate real network scenarios in a lab environment. The Tester may also support MPLS-TE signaling to act as the ingress/egress node.

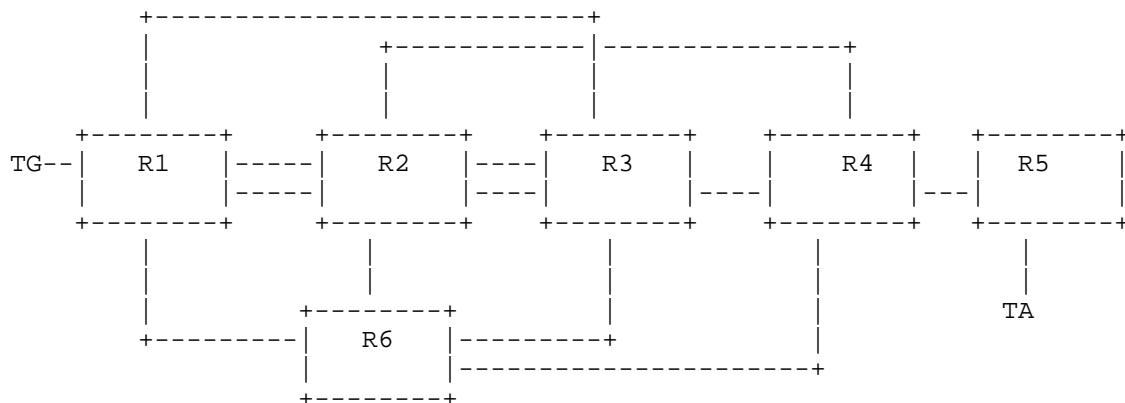


Fig. 1 Fast Reroute Topology

The tester must be able to record the number of lost, duplicate, and reordered packets. It should further record arrival and departure times so that Failover Time, Additive Latency, and Reversion Time can be measured. The tester may be a single device or a test system emulating different roles along a primary or backup path.

The label stack is dependent on the following 3 entities:

- (1) Type of protection (Link Vs Node)
- (2) # of remaining hops of the primary tunnel from the Point of Local Repair (PLR)[RFC6414]
- (3) # of remaining hops of the backup tunnel from the PLR

Due to this dependency, it is RECOMMENDED that the benchmarking of failover times be performed on all the topologies provided in section 6.

5. Test Considerations

This section discusses the fundamentals of MPLS Protection testing:

- (1) The types of network events that causes failover
- (2) Indications for failover
- (3) the use of data traffic
- (4) Traffic generation
- (5) LSP Scaling
- (6) Reversion of LSP
- (7) IGP Selection

5.1. Failover Events

The failover to the backup tunnel is primarily triggered by either link or node failures observed downstream of the PLR. Some of these failure events [RFC6414] are listed below.

Link Failure Events

- Interface Shutdown on PLR side with POS Alarm
- Interface Shutdown on remote side with POS Alarm
- Interface Shutdown on PLR side with RSVP hello enabled
- Interface Shutdown on remote side with RSVP hello enabled
- Interface Shutdown on PLR side with BFD
- Interface Shutdown on remote side with BFD
- Fiber Pull on the PLR side (Both TX & RX or just the TX)
- Fiber Pull on the remote side (Both TX & RX or just the RX)
- Online insertion and removal (OIR) on PLR side
- OIR on remote side
- Sub-interface failure on PLR side (e.g. shutting down of a VLAN)
- Sub-interface failure on remote side
- Parent interface shutdown on PLR side (an interface bearing multiple sub-interfaces)
- Parent interface shutdown on remote side

Node Failure Events

- A System reload initiated either by a graceful shutdown or by a power failure.
- A system crash due to a software failure or an assert.

5.2. Failure Detection

Link failure detection [RFC6414] time depends on the link type and failure detection techniques enabled. For SONET/SDH, the alarm type (such as LOS, AIS, or RDI) can be used. Other link types have layer-two alarms, but they may not provide a short enough failure detection time. Ethernet based links do not have layer 2 failure indicators, and therefore relies on layer 3 signaling for failure detection. However for directly connected devices, remote fault indication in the Ethernet auto-negotiation scheme could be considered as a type of layer 2 link failure indicator.

BFD and RSVP-hellos may be used as failure detection techniques. These methods can be used for the layer 3 failure indicators required by Ethernet based links, or for some other non- Ethernet based links to help improve failure detection time. However, these fast failure detection mechanisms are out of scope of this document.

The test procedures in this document can be used for MPLS-TE protection benchmarking due to either a local failure or remote failure.

5.3. Use of Data Traffic for MPLS Protection benchmarking

Currently end customers use packet loss as a key metric for Failover Time [RFC6414]. Failover Packet Loss [RFC6414] is an externally

observable event and has direct impact on application performance. MPLS-TE protection is expected to minimize the packet loss in the event of a failure. For this reason it is important to develop a standard router benchmarking methodology for measuring MPLS protection that uses packet loss as a metric. At a known rate of forwarding, packet loss can be measured and the failover time can be determined. Measurement of control plane recovery and establishing backup paths is not enough to verify a timely failover. Failover performance is best determined when packets are actually switched to the backup path.

Benefit of using packet loss for calculation of failover time is that it allows use of a black-box test environment. Data traffic is offered at line-rate to the device under test (DUT) an emulated network failure event is forced to occur, and packet loss is externally measured to calculate the convergence time. This setup is independent of the DUT architecture.

The methodology considers lost, packet in error, out-of-order [RFC4689] and duplicate packets as impaired packets that contribute to the Failover Time.

5.4. LSP and Route Scaling

Failover time performance may vary with the number of established primary and backup tunnel label switched paths (LSP) and installed routes. However, the procedure outlined here should be used for any number of LSPs (L) and number of routes protected by the headend as the PLR(R). The amount of L and R must be recorded. The recommended table is provided in appendix A.

5.5. Selection of IGP

The underlying IGP could be ISIS-TE or OSPF-TE for the methodology proposed here. See [RFC6412] for IGP options to consider and report. At least one of the IGP is required to be enabled for the procedures discussed in the document.

5.6. Restoration and Reversion

Path restoration [RFC6414] provides a method to restore an alternate primary LSP upon failure and to switch traffic from the Backup Path to the restored Primary Path (Reversion). In MPLS-FRR, Reversion can be implemented as Global Reversion or Local Reversion. It is important to include Restoration and Reversion as a step in each test case to measure the amount of packet loss, out of order packets, or duplicate packets that occurs in this process.

Note: In addition to restoration and reversion, re-optimization can take place while the failure is still not recovered but it depends on the user configuration, and re-optimization timers.

5.7. Offered Load

It is recommended that there be three or more traffic streams configured with steady and constant rate of flow for all the streams. In order to monitor the DUT performance for recovery times, a set of route prefixes should be advertised before traffic is sent. The traffic should be configured to target the advertised routes.

For better accuracy, one may consider provisioning 16 flows, or more if possible. IP Prefix-dependency behaviors are key and tests with route-specific flows spread across the routing table reveals such dependency. Sending traffic to all of the prefixes reachable by the protected tunnel in a round-robin fashion is not recommended as the time interval between two subsequent packets destined to one prefix may be higher than the failover time being measured resulting in inaccurate failover measurements.

5.8. Tester Capabilities

It is RECOMMENDED that the Tester used to execute each test case have the following capabilities:

- 1.Ability to establish MPLS-TE tunnels and push/pop labels.
- 2.Ability to produce Failover Event [RFC6414].
- 3.Ability to insert a timestamp in each data packet's IP payload.
- 4.An internal time clock to control timestamping, time measurements, and time calculations.
- 5.Ability to disable or tune specific Layer-2 and Layer-3 protocol functions on any interface(s) such as disabling or enabling interface IP addresses, auto-negotiation on ethernet interfaces or scrambling on Packet over SONET interfaces.
6. In a case, if the tester is the headend, it should be able to react upon the receipt of path error from the PLR

The Tester MAY be capable to make non-data plane convergence observations and use those observations for measurements.

5.9. Failover Time Measurement Methods

Failover Time is calculated using one of the following three methods

1. Packet-Loss Based method (PLBM): (Number of packets dropped/ packets per second * 1000) milliseconds. This method could also be referred as Loss-Derived method.
2. Time-Based Loss Method (TBLM): This method relies on the ability of the Traffic generators to provide statistics which reveal the duration of failure in milliseconds based on when the packet loss occurred (interval between non-zero packet loss and zero loss).
3. Timestamp Based Method (TBM): This method of failover calculation is based on the timestamp that gets transmitted as payload in the packets originated by the generator. The Traffic Analyzer records the timestamp of the last packet received before the failover event and the first packet after the failover and derives the time based on the difference between these 2 timestamps. Note: The payload could also contain sequence numbers for out-of-order packet calculation and duplicate packets.

The timestamp based method would be able to detect Reversion impairments beyond loss, thus it is RECOMMENDED method as a Failover Time method.

6. Reference Test Setup

In addition to the general reference topology shown in figure 1, this section provides detailed insight into various proposed test setups that should be considered for comprehensively benchmarking the failover time in different roles along the primary tunnel

This section proposes a set of topologies that covers all the scenarios for local protection. All of these topologies can be mapped to the reference topology shown in Figure 1. Topologies provided in this section refer to the testbed required to benchmark failover time when the DUT is configured as a PLR in either Headend or midpoint role. Provided with each topology below is the label stack at the PLR. Penultimate Hop Popping (PHP) MAY be used and must be reported when used.

Figures 2 thru 9 use the following convention and are subset of figure 1:

- a) HE is Headend
- b) TE is Tail-End
- c) MID is Mid point
- d) MP is Merge Point
- e) PLR is Point of Local Repair
- f) PRI is Primary Path
- g) BKP denotes Backup Path and Nodes
- h) UR is Upstream Router

6.1. Link Protection

6.1.1. Link Protection - 1 hop primary (from PLR) and 1 hop backup TE tunnels

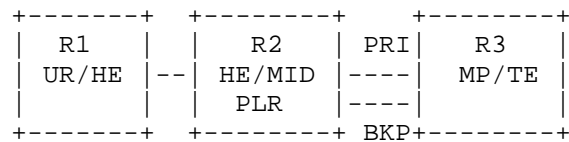


Figure 2.

Traffic	Num of Labels before failure	Num of labels after failure
IP TRAFFIC (P-P)	0	0
Layer3 VPN (PE-PE)	1	1
Layer3 VPN (PE-P)	2	2
Layer2 VC (PE-PE)	1	1
Layer2 VC (PE-P)	2	2
Mid-point LSPs	0	0

Note: Please note the following:

- a) For P-P case, R2 and R3 acts as P routers
- b) For PE-PE case, R2 acts as PE and R3 acts as a remote PE
- c) For PE-P case, R2 acts as a PE router, R3 acts as a P router and R5 acts as remote PE router (Please refer to figure 1 for complete setup)
- d) For Mid-point case, R1, R2 and R3 act as shown in above figure HE, Midpoint/PLR and TE respectively

6.1.2. Link Protection - 1 hop primary (from PLR) and 2 hop backup TE tunnels

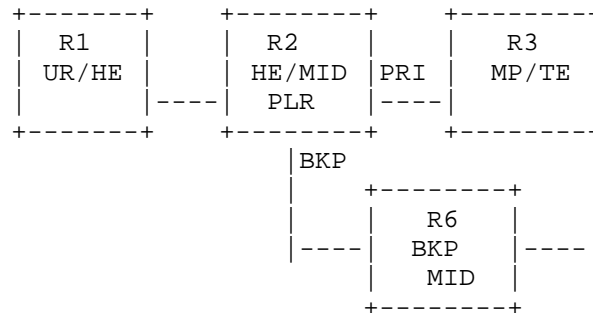


Figure 3.

Traffic	Num of Labels before failure	Num of labels after failure
IP TRAFFIC (P-P)	0	1
Layer3 VPN (PE-PE)	1	2
Layer3 VPN (PE-P)	2	3
Layer2 VC (PE-PE)	1	2
Layer2 VC (PE-P)	2	3
Mid-point LSPs	0	1

Note: Please note the following:

- For P-P case, R2 and R3 acts as P routers
- For PE-PE case, R2 acts as PE and R3 acts as a remote PE
- For PE-P case, R2 acts as a PE router, R3 acts as a P router and R5 acts as remote PE router (Please refer to figure 1 for complete setup)
- For Mid-point case, R1, R2 and R3 act as shown in above figure HE, Midpoint/PLR and TE respectively

6.1.3. Link Protection - 2+ hops (from PLR) primary and 1 hop backup TE tunnels

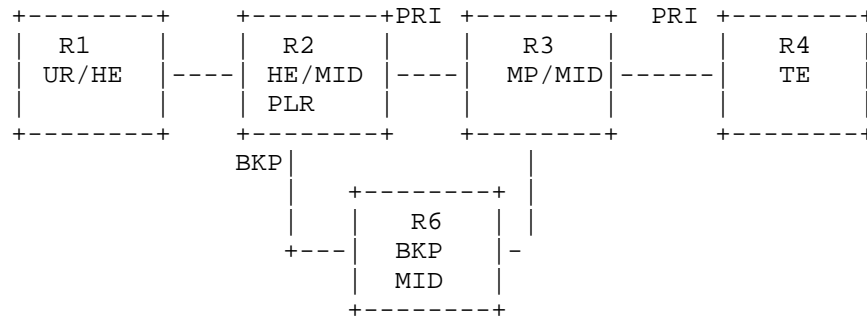


Figure 5.

Traffic	Num of Labels before failure	Num of labels after failure
IP TRAFFIC (P-P)	1	2
Layer3 VPN (PE-PE)	2	3
Layer3 VPN (PE-P)	3	4
Layer2 VC (PE-PE)	2	3
Layer2 VC (PE-P)	3	4
Mid-point LSPs	1	2

Note: Please note the following:

- For P-P case, R2, R3 and R4 acts as P routers
- For PE-PE case, R2 acts as PE and R4 acts as a remote PE
- For PE-P case, R2 acts as a PE router, R3 acts as a P router and R5 acts as remote PE router (Please refer to figure 1 for complete setup)
- For Mid-point case, R1, R2, R3 and R4 act as shown in above figure HE, Midpoint/PLR and TE respectively

6.2. Node Protection

6.2.1. Node Protection - 2 hop primary (from PLR) and 1 hop backup TE tunnels

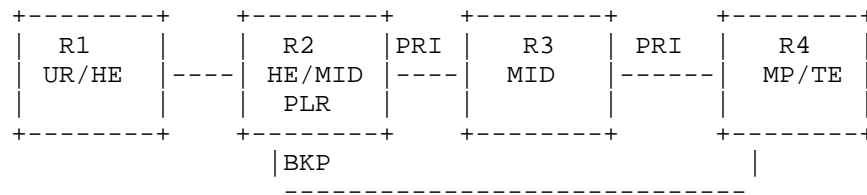


Figure 6.

Traffic	Num of Labels before failure	Num of labels after failure
IP TRAFFIC (P-P)	1	0
Layer3 VPN (PE-PE)	2	1
Layer3 VPN (PE-P)	3	2
Layer2 VC (PE-PE)	2	1
Layer2 VC (PE-P)	3	2
Mid-point LSPs	1	0

Note: Please note the following:

- For P-P case, R2, R3 and R3 acts as P routers
- For PE-PE case, R2 acts as PE and R4 acts as a remote PE
- For PE-P case, R2 acts as a PE router, R4 acts as a P router and R5 acts as remote PE router (Please refer to figure 1 for complete setup)
- For Mid-point case, R1, R2, R3 and R4 act as shown in above figure HE, Midpoint/PLR and TE respectively

6.2.2. Node Protection - 2 hop primary (from PLR) and 2 hop backup TE tunnels

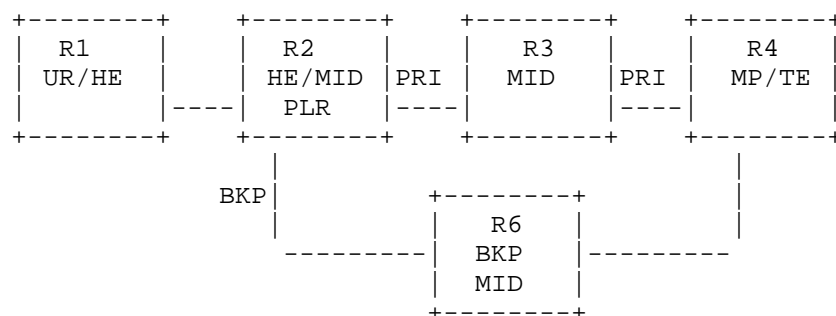


Figure 7.

Traffic	Num of Labels before failure	Num of labels after failure
IP TRAFFIC (P-P)	1	1
Layer3 VPN (PE-PE)	2	2
Layer3 VPN (PE-P)	3	3
Layer2 VC (PE-PE)	2	2
Layer2 VC (PE-P)	3	3
Mid-point LSPs	1	1

Note: Please note the following:

- a) For P-P case, R2, R3 and R4 acts as P routers
- b) For PE-PE case, R2 acts as PE and R4 acts as a remote PE
- c) For PE-P case, R2 acts as a PE router, R4 acts as a P router and R5 acts as remote PE router (Please refer to figure 1 for complete setup)
- d) For Mid-point case, R1, R2, R3 and R4 act as shown in above figure HE, Midpoint/PLR and TE respectively

6.2.3. Node Protection - 3+ hop primary (from PLR) and 1 hop backup TE tunnels

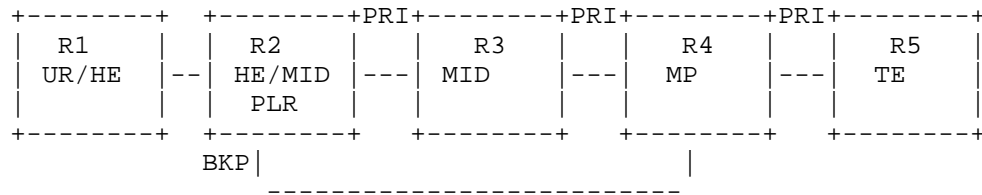


Figure 8.

Traffic	Num of Labels before failure	Num of labels after failure
IP TRAFFIC (P-P)	1	1
Layer3 VPN (PE-PE)	2	2
Layer3 VPN (PE-P)	3	3
Layer2 VC (PE-PE)	2	2
Layer2 VC (PE-P)	3	3
Mid-point LSPs	1	1

Note: Please note the following:

- a) For P-P case, R2, R3, R4 and R5 acts as P routers
- b) For PE-PE case, R2 acts as PE and R5 acts as a remote PE
- c) For PE-P case, R2 acts as a PE router, R4 acts as a P router and R5 acts as remote PE router (Please refer to figure 1 for complete setup)
- d) For Mid-point case, R1, R2, R3, R4 and R5 act as shown in above figure HE, Midpoint/PLR and TE respectively

6.2.4. Node Protection - 3+ hop primary (from PLR) and 2 hop backup TE tunnels

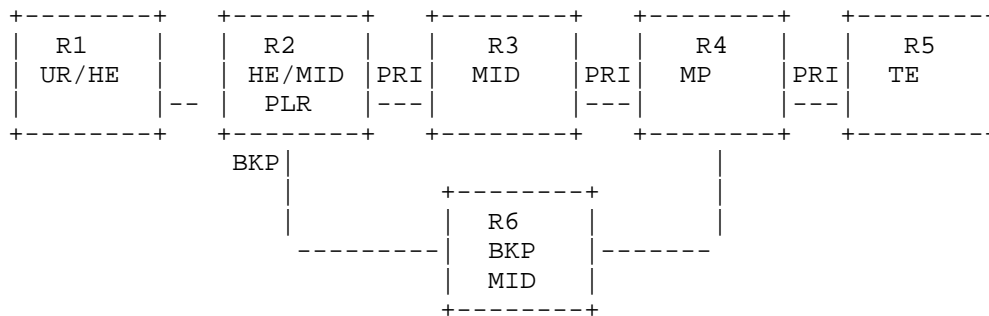


Figure 9.

Traffic	Num of Labels before failure	Num of labels after failure
IP TRAFFIC (P-P)	1	2
Layer3 VPN (PE-PE)	2	3
Layer3 VPN (PE-P)	3	4
Layer2 VC (PE-PE)	2	3
Layer2 VC (PE-P)	3	4
Mid-point LSPs	1	2

Note: Please note the following:

- For P-P case, R2, R3, R4 and R5 acts as P routers
- For PE-PE case, R2 acts as PE and R5 acts as a remote PE
- For PE-P case, R2 acts as a PE router, R4 acts as a P router and R5 acts as remote PE router (Please refer to figure 1 for complete setup)
- For Mid-point case, R1, R2, R3, R4 and R5 act as shown in above figure HE, Midpoint/PLR and TE respectively

7. Test Methodology

The procedure described in this section can be applied to all the 8 base test cases and the associated topologies. The backup as well as the primary tunnels are configured to be alike in terms of bandwidth usage. In order to benchmark failover with all possible label stack depth applicable as seen with current deployments, it is RECOMMENDED to perform all of the test cases provided in this section. The forwarding performance test cases in section 7.1 MUST be performed prior to performing the failover test cases.

The considerations of Section 4 of [RFC2544] are applicable when evaluating the results obtained using these methodologies as well.

7.1. MPLS FRR Forwarding Performance

Benchmarking Failover Time [RFC6414] for MPLS protection first requires baseline measurement of the forwarding performance of the test topology including the DUT. Forwarding performance is benchmarked by the Throughput as defined in [RFC5695] and measured in units packet per second (pps). This section provides two test cases to benchmark forwarding performance. These are with the DUT configured as a Headend PLR, Mid-Point PLR, and Egress PLR.

7.1.1. Headend PLR Forwarding Performance

Objective:

To benchmark the maximum rate (pps) on the PLR (as headend) over primary LSP and backup LSP.

Test Setup:

- A. Select any one topology out of the 8 from section 6.
- B. Select or enable IP, Layer 3 VPN or Layer 2 VPN services with DUT as Headend PLR.
- C. The DUT will also have 2 interfaces connected to the traffic Generator/analyzer. (If the node downstream of the PLR is not a simulated node, then the Ingress of the tunnel should have one link connected to the traffic generator and the node downstream to the PLR or the egress of the tunnel should have a link connected to the traffic analyzer).

Procedure:

1. Establish the primary LSP on R2 required by the topology selected.
2. Establish the backup LSP on R2 required by the selected topology.
3. Verify primary and backup LSPs are up and that primary is protected.

4. Verify Fast Reroute protection is enabled and ready.
5. Setup traffic streams as described in section 5.7.
6. Send the required MPLS traffic over the primary LSP to achieve the throughput supported by the DUT (section 6, RFC 2544).
7. Record the Throughput over the primary LSP.
8. Trigger a link failure as described in section 5.1.
9. Verify that the offered load gets mapped to the backup tunnel and measure the Additive Backup Delay (RFC 6414).
10. 30 seconds after Failover, stop the offered load and measure the Throughput, Packet Loss, Out-of-Order Packets, and Duplicate Packets over the Backup LSP.
11. Adjust the offered load and repeat steps 6 through 10 until the Throughput values for the primary and backup LSPs are equal.
12. Record the final Throughput, which corresponds to the offered load that will be used for the Headend PLR failover test cases.

7.1.2. Mid-Point PLR Forwarding Performance

Objective:

To benchmark the maximum rate (pps) on the PLR (as mid-point) over primary LSP and backup LSP.

Test Setup:

- A. Select any one topology out of the 8 from section 6.
- B. The DUT will also have 2 interfaces connected to the traffic generator.

Procedure:

1. Establish the primary LSP on R1 required by the topology selected.
2. Establish the backup LSP on R2 required by the selected topology.
3. Verify primary and backup LSPs are up and that primary is protected.
4. Verify Fast Reroute protection is enabled and ready.
5. Setup traffic streams as described in section 5.7.
6. Send MPLS traffic over the primary LSP at the Throughput supported by the DUT (section 6, RFC 2544).
7. Record the Throughput over the primary LSP.
8. Trigger a link failure as described in section 5.1.
9. Verify that the offered load gets mapped to the backup tunnel and measure the Additive Backup Delay (RFC 6414).
10. 30 seconds after Failover, stop the offered load and measure the Throughput, Packet Loss, Out-of-Order Packets, and Duplicate Packets over the Backup LSP.
11. Adjust the offered load and repeat steps 6 through 10 until the Throughput values for the primary and backup LSPs are equal.
12. Record the final Throughput which corresponds to the offered load that will be used for the Mid-Point PLR failover test cases.

7.2. Headend PLR with Link Failure

Objective:

To benchmark the MPLS failover time due to link failure events described in section 5.1 experienced by the DUT which is the Headend PLR.

Test Setup:

- A. Select any one topology out of the 8 from section 6.
- B. Select or enable IP, Layer 3 VPN or Layer 2 VPN services with DUT as Headend PLR.
- C. The DUT will also have 2 interfaces connected to the traffic Generator/analyzer. (If the node downstream of the PLR is not a simulated node, then the Ingress of the tunnel should have one link connected to the traffic generator and the node downstream to the PLR or the egress of the tunnel should have a link connected to the traffic analyzer).

Test Configuration:

- 1. Configure the number of primaries on R2 and the backups on R2 as required by the topology selected.
- 2. Configure the test setup to support Reversion.
- 3. Advertise prefixes (as per FRR Scalability Table described in Appendix A) by the tail end.

Procedure:

Test Case "7.1.1. Headend PLR Forwarding Performance" MUST be completed first to obtain the Throughput to use as the offered load.

- 1. Establish the primary LSP on R2 required by the topology selected.
- 2. Establish the backup LSP on R2 required by the selected topology.
- 3. Verify primary and backup LSPs are up and that primary is protected.
- 4. Verify Fast Reroute protection is enabled and ready.
- 5. Setup traffic streams for the offered load as described in section 5.7.
- 6. Provide the offered load from the tester at the Throughput [RFC1242] level obtained from test case 7.1.1.

7. Verify traffic is switched over Primary LSP without packet loss.
8. Trigger a link failure as described in section 5.1.
9. Verify that the offered load gets mapped to the backup tunnel and measure the Additive Backup Delay.
10. 30 seconds after Failover [RFC6414], stop the offered load and measure the total Failover Packet Loss [RFC6414].
11. Calculate the Failover Time [RFC6414] benchmark using the selected Failover Time Calculation Method (TBLM, PLBM, or TBM) [RFC6414].
12. Restart the offered load and restore the primary LSP to verify Reversion [RFC6414] occurs and measure the Reversion Packet Loss [RFC6414].
13. Calculate the Reversion Time [RFC6414] benchmark using the selected Failover Time Calculation Method (TBLM, PLBM, or TBM) [RFC6414].
14. Verify Headend signals new LSP and protection should be in place again.

IT is RECOMMENDED that this procedure be repeated for each of the link failure triggers defined in section 5.1.

7.3. Mid-Point PLR with Link Failure

Objective:

To benchmark the MPLS failover time due to link failure events described in section 5.1 experienced by the DUT which is the Mid-Point PLR.

Test Setup:

- A. Select any one topology out of the 8 from section 6.
- B. The DUT will also have 2 interfaces connected to the traffic generator.

Test Configuration:

1. Configure the number of primaries on R1 and the backups on R2 as required by the topology selected.
2. Configure the test setup to support Reversion.
3. Advertise prefixes (as per FRR Scalability Table described in Appendix A) by the tail end.

Procedure:

Test Case "7.1.2. Mid-Point PLR Forwarding Performance" MUST be completed first to obtain the Throughput to use as the offered load.

1. Establish the primary LSP on R1 required by the topology selected.
2. Establish the backup LSP on R2 required by the selected topology.
3. Perform steps 3 through 14 from section 7.2 Headend PLR with Link Failure.

IT is RECOMMENDED that this procedure be repeated for each of the link failure triggers defined in section 5.1.

7.4. Headend PLR with Node Failure

Objective:

To benchmark the MPLS failover time due to Node failure events described in section 5.1 experienced by the DUT which is the Headend PLR.

Test Setup:

- A. Select any one topology out of the 8 from section 6.
- B. Select or enable IP, Layer 3 VPN or Layer 2 VPN services with DUT as Headend PLR.

- C. The DUT will also have 2 interfaces connected to the traffic generator/analyzer.

Test Configuration:

1. Configure the number of primaries on R2 and the backups on R2 as required by the topology selected.
2. Configure the test setup to support Reversion.
3. Advertise prefixes (as per FRR Scalability Table described in Appendix A) by the tail end.

Procedure:

Test Case "7.1.1. Headend PLR Forwarding Performance" MUST be completed first to obtain the Throughput to use as the offered load.

1. Establish the primary LSP on R2 required by the topology selected.
2. Establish the backup LSP on R2 required by the selected topology.
3. Verify primary and backup LSPs are up and that primary is protected.
4. Verify Fast Reroute protection is enabled and ready.
5. Setup traffic streams for the offered load as described in section 5.7.
6. Provide the offered load from the tester at the Throughput [RFC1242] level obtained from test case 7.1.1.
7. Verify traffic is switched over Primary LSP without packet loss.
8. Trigger a node failure as described in section 5.1.
9. Perform steps 9 through 14 in 7.2 Headend PLR with Link Failure.

IT is RECOMMENDED that this procedure be repeated for each of the node failure triggers defined in section 5.1.

7.5. Mid-Point PLR with Node Failure

Objective:

To benchmark the MPLS failover time due to Node failure events described in section 5.1 experienced by the DUT which is the Mid-Point PLR.

Test Setup:

- A. Select any one topology from section 6.1 to 6.2.
- B. The DUT will also have 2 interfaces connected to the traffic generator.

Test Configuration:

1. Configure the number of primaries on R1 and the backups on R2 as required by the topology selected.
2. Configure the test setup to support Reversion.
3. Advertise prefixes (as per FRR Scalability Table described in Appendix A) by the tail end.

Procedure:

Test Case "7.1.1. Mid-Point PLR Forwarding Performance" MUST be completed first to obtain the Throughput to use as the offered load.

1. Establish the primary LSP on R1 required by the topology selected.
2. Establish the backup LSP on R2 required by the selected topology.
3. Verify primary and backup LSPs are up and that primary is protected.

4. Verify Fast Reroute protection is enabled and ready.
5. Setup traffic streams for the offered load as described in section 5.7.
6. Provide the offered load from the tester at the Throughput [RFC1242] level obtained from test case 7.1.1.
7. Verify traffic is switched over Primary LSP without packet loss.
8. Trigger a node failure as described in section 5.1.
9. Perform steps 9 through 14 in 7.2 Headend PLR with Link Failure.

IT is RECOMMENDED that this procedure be repeated for each of the node failure triggers defined in section 5.1.

8. Reporting Format

For each test, it is RECOMMENDED that the results be reported in the following format.

Parameter	Units
IGP used for the test	ISIS-TE/ OSPF-TE
Interface types	Gige,POS,ATM,VLAN etc.
Packet Sizes offered to the DUT	Bytes (at layer 3)
Offered Load (Throughput)	packets per second
IGP routes advertised	Number of IGP routes
Penultimate Hop Popping	Used/Not Used
RSVP hello timers	Milliseconds
Number of Protected tunnels	Number of tunnels
Number of VPN routes installed on the Headend	Number of VPN routes

Number of VC tunnels	Number of VC tunnels
Number of mid-point tunnels	Number of tunnels
Number of Prefixes protected by Primary	Number of LSPs
Topology being used	Section number, and figure reference
Failover Event	Event type
Re-optimization	Yes/No

Benchmarks (to be recorded for each test case):

Failover-

Failover Time	seconds
Failover Packet Loss	packets
Additive Backup Delay	seconds
Out-of-Order Packets	packets
Duplicate Packets	packets
Failover Time Calculation Method	Method Used

Reversion-

Reversion Time	seconds
Reversion Packet Loss	packets
Additive Backup Delay	seconds
Out-of-Order Packets	packets
Duplicate Packets	packets
Failover Time Calculation Method	Method Used

9. Security Considerations

Benchmarking activities as described in this memo are limited to technology characterization using controlled stimuli in a laboratory environment, with dedicated address space and the constraints specified in the sections above.

The benchmarking network topology will be an independent test setup and MUST NOT be connected to devices that may forward the test traffic into a production network, or misroute traffic to the test management network.

Further, benchmarking is performed on a "black-box" basis, relying solely on measurements observable external to the DUT/SUT.

Special capabilities SHOULD NOT exist in the DUT/SUT specifically for benchmarking purposes. Any implications for network security arising from the DUT/SUT SHOULD be identical in the lab and in production networks.

10. IANA Considerations

This document does not require any new allocations by IANA.

11. Acknowledgements

We would like to thank Jean Philip Vasseur for his invaluable input to the document, Curtis Villamizar for his contribution in suggesting text on definition and need for benchmarking Correlated failures and Bhavani Parise for his textual input and review. Additionally we would like to thank Al Morton, Arun Gandhi, Amrit Hanspal, Karu Ratnam, Raveesh Janardan, Andrey Kiselev, and Mohan Nanduri for their formal reviews of this document.

12. References

12.1. Informative References

- [RFC2285] Mandeville, R., "Benchmarking Terminology for LAN Switching Devices", RFC 2285, February 1998.
- [RFC4689] Poretsky, S., Perser, J., Erramilli, S., and S. Khurana, "Terminology for Benchmarking Network-layer Traffic Control Mechanisms", RFC 4689, October 2006.

12.2. Normative References

- [RFC1242] Bradner, S., "Benchmarking terminology for network interconnection devices", RFC 1242, July 1991.
- [RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.
- [RFC2544] Bradner, S. and J. McQuaid, "Benchmarking Methodology for Network Interconnect Devices", RFC 2544, March 1999.
- [RFC4090] Pan, P., Swallow, G., and A. Atlas, "Fast Reroute Extensions to RSVP-TE for LSP Tunnels", RFC 4090, May 2005.

- [RFC5695] Akhter, A., Asati, R., and C. Pignataro, "MPLS Forwarding Benchmarking Methodology for IP Flows", RFC 5695, November 2009.
- [RFC6412] Poretsky, S., Imhoff, B., and K. Michielsen, "Terminology for Benchmarking Link-State IGP Data-Plane Route Convergence", RFC 6412, November 2011.
- [RFC6414] Poretsky, S., Papneja, R., Karthik, J., and S. Vapiwala, "Benchmarking Terminology for Protection Performance", RFC 6414, November 2011.

Appendix A. Fast Reroute Scalability Table

This section provides the recommended numbers for evaluating the scalability of fast reroute implementations. It also recommends the typical numbers for IGP/VPNv4 Prefixes, LSP Tunnels and VC entries. Based on the features supported by the device under test (DUT), appropriate scaling limits can be used for the test bed.

A1. FRR IGP Table

No. of Headend TE Tunnels (L)	IGP Prefixes (R)
1	100
1	500
1	1000
1	2000
1	5000
2 (Load Balance)	100
2 (Load Balance)	500
2 (Load Balance)	1000
2 (Load Balance)	2000
2 (Load Balance)	5000
100	100
500	500
1000	1000
2000	2000

A2. FRR VPN Table

No. of Headend TE Tunnels (L)	VPNv4 Prefixes (R)
1	100
1	500
1	1000
1	2000
1	5000
1	10000
1	20000
1	Max
2 (Load Balance)	100
2 (Load Balance)	500
2 (Load Balance)	1000
2 (Load Balance)	2000
2 (Load Balance)	5000
2 (Load Balance)	10000
2 (Load Balance)	20000
2 (Load Balance)	Max

A3. FRR Mid-Point LSP Table

No of Mid-point TE LSPs could be configured at recommended levels - 100, 500, 1000, 2000, or max supported number.

A2. FRR VC Table

No. of Headend TE Tunnels (L)	VC entries (R)
1	100
1	500
1	1000
1	2000
1	Max
100	100
500	500
1000	1000
2000	2000

Appendix B. Abbreviations

AIS	- Alarm Indication Signal
BFD	- Bidirectional Fault Detection
BGP	- Border Gateway protocol
CE	- Customer Edge
DUT	- Device Under Test
FRR	- Fast Reroute
IGP	- Interior Gateway Protocol
IP	- Internet Protocol
LOS	- Loss of Signal
LSP	- Label Switched Path
MP	- Merge Point
MPLS	- Multi Protocol Label Switching
N-Nhop	- Next - Next Hop
Nhop	- Next Hop
OIR	- Online Insertion and Removal
P	- Provider
PE	- Provider Edge
PHP	- Penultimate Hop Popping
PLR	- Point of Local Repair
RSVP	- Resource reSerVation Protocol
SRLG	- Shared Risk Link Group
TA	- Traffic Analyzer
TE	- Traffic Engineering
TG	- Traffic Generator
VC	- Virtual Circuit
VPN	- Virtual Private Network

Authors' Addresses

Rajiv Papneja
Huawei Technologies
2330 Central Expressway
Santa Clara, CA 95050
USA

Email: rajiv.papneja@huawei.com

Samir Vapiwala
Cisco Systems
300 Beaver Brook Road
Boxborough, MA 01719
USA

Email: svapiwal@cisco.com

Jay Karthik
Cisco Systems
300 Beaver Brook Road
Boxborough, MA 01719
USA

Email: jkarthik@cisco.com

Scott Poretsky
Allot Communications
USA

Email: sporetsky@allot.com

Shankar Rao
Qwest Communications
950 17th Street
Suite 1900
Denver, CO 80210
USA

Email: shankar.rao@du.edu

JL. Le Roux
France Telecom
2 av Pierre Marzin
22300 Lannion
France

Email: jeanlouis.leroux@orange.com

This Internet-Draft, draft-ietf-bmwg-sip-bench-meth-03.txt, has expired, and has been deleted from the Internet-Drafts directory. An Internet-Draft expires 185 days from the date that it is posted unless it is replaced by an updated version, or the Secretariat has been notified that the document is under official review by the IESG or has been passed to the RFC Editor for review and/or publication as an RFC. This Internet-Draft was not published as an RFC.

Internet-Drafts are not archival documents, and copies of Internet-Drafts that have been deleted from the directory are not available. The Secretariat does not have any information regarding the future plans of the authors or working group, if applicable, with respect to this deleted Internet-Draft. For more information, or to request a copy of the document, please contact the authors directly.

Draft Authors:

Carol Davids<davids@iit.edu>

Vijay Gurbani<vkg@bell-labs.com>

Scott Poretsky<sporetsky@allot.com>

This Internet-Draft, draft-ietf-bmwg-sip-bench-term-03.txt, has expired, and has been deleted from the Internet-Drafts directory. An Internet-Draft expires 185 days from the date that it is posted unless it is replaced by an updated version, or the Secretariat has been notified that the document is under official review by the IESG or has been passed to the RFC Editor for review and/or publication as an RFC. This Internet-Draft was not published as an RFC.

Internet-Drafts are not archival documents, and copies of Internet-Drafts that have been deleted from the directory are not available. The Secretariat does not have any information regarding the future plans of the authors or working group, if applicable, with respect to this deleted Internet-Draft. For more information, or to request a copy of the document, please contact the authors directly.

Draft Authors:

Carol Davids<davids@iit.edu>

Vijay Gurbani<vkg@bell-labs.com>

Scott Poretsky<sporetsky@allot.com>

Benchmarking Working Group
Internet-Draft
Intended status: Standards Track
Expires: September 27, 2012

R. Papneja
Huawei Technologies
B. Parise
Cisco Systems
S. Hares
Huawei Technologies
I. Varlashkin
Easynet Global Services
March 26, 2012

Basic BGP Convergence Benchmarking Methodology for Data Plane
Convergence
draft-papneja-bgp-basic-dp-convergence-03.txt

Abstract

BGP is widely deployed and used by several service providers as the default Inter AS routing protocol. It is of utmost importance to ensure that when a BGP peer or a downstream link of a BGP peer fails, the alternate paths are rapidly used and routes via these alternate paths are installed. This document provides the basic BGP Benchmarking Methodology using existing BGP Convergence Terminology, RFC 4098.

Status of this Memo

This Internet-Draft is submitted in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at <http://datatracker.ietf.org/drafts/current/>.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

This Internet-Draft will expire on September 27, 2012.

Copyright Notice

Copyright (c) 2012 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust's Legal

Provisions Relating to IETF Documents
(<http://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

This document may contain material from IETF Documents or IETF Contributions published or made publicly available before November 10, 2008. The person(s) controlling the copyright in some of this material may not have granted the IETF Trust the right to allow modifications of such material outside the IETF Standards Process. Without obtaining an adequate license from the person(s) controlling the copyright in such materials, this document may not be modified outside the IETF Standards Process, and derivative works of it may not be created outside the IETF Standards Process, except to format it for publication as an RFC or to translate it into languages other than English.

Table of Contents

1. Introduction	4
1.1. Precise Benchmarking Definition	4
1.2. Purpose of BGP FIB (Data Plane) Convergence	4
1.3. Control Plane Convergence	5
1.4. Benchmarking Testing	5
2. Existing Definitions and Requirements	5
3. Test Topologies	6
3.1. General Reference Topologies	6
4. Test Considerations	8
4.1. Number of Peers	9
4.2. Number of Routes per Peer	9
4.3. Policy Processing/Reconfiguration	9
4.4. Configured Parameters (Timers, etc..)	9
4.5. Interface Types	11
4.6. Measurement Accuracy	11
4.7. Measurement Statistics	11
4.8. Authentication	12
4.9. Convergence Events	12
4.10. High Availability	12
5. Test Cases	12
5.1. Basic Convergence Tests	12
5.1.1. RIB-IN Convergence	13
5.1.2. RIB-OUT Convergence	14
5.1.3. eBGP Convergence	16
5.1.4. iBGP Convergence	16
5.1.5. eBGP Multihop Convergence	16
5.2. BGP Failure/Convergence Events	18
5.2.1. Physical Link Failure on DUT End	18
5.2.2. Physical Link Failure on Remote/Emulator End	19
5.2.3. ECMP Link Failure on DUT End	19
5.3. BGP Adjacency Failure (Non-Physical Link Failure) on Emulator	19
5.4. BGP Hard Reset Test Cases	21
5.4.1. BGP Non-Recovering Hard Reset Event on DUT	21
5.5. BGP Soft Reset	22
5.6. BGP Route Withdrawal Convergence Time	23
5.7. BGP Path Attribute Change Convergence Time	25
5.8. BGP Graceful Restart Convergence Time	26
6. Reporting Format	28
7. IANA Considerations	31
8. Security Considerations	31
9. References	31
9.1. Normative References	31
9.2. Informative References	32
Authors' Addresses	32

1. Introduction

This document defines the methodology for benchmarking data plane FIB convergence performance of BGP in router and switches for simple topologies of 3 or 4 nodes. The methodology proposed in this document applies to both IPv4 and IPv6 and if a particular test is unique to one version, it is marked accordingly. For IPv6 benchmarking the device under test will require the support of Multi-Protocol BGP (MP-BGP) [RFC4760, RFC2545].

The scope of this companion document is limited to basic BGP protocol FIB convergence measurements. BGP extensions outside of carrying IPv6 in (MP-BGP) [RFC4760, RFC2545] are outside the scope of this document. Interaction with IGPs (IGP interworking) is outside the scope of this document.

1.1. Precise Benchmarking Definition

Since benchmarking is science of precision, let us restate the purpose of this document in benchmarking terms. This document defines methodology to test

- data plane convergence on a single BGP device that supports the BGP [RFC4271] functionality
- in test topology of 3 or 4 nodes
- using Basic BGP

Data plane convergence is defined as the completion of all FIB changes so that all forwarded traffic now takes the new proposed route. RFC 4098 defines the terms BGP device, FIB and the forwarded traffic. Data plane convergence is different than control plane convergence within a node.

Basic BGP is defined as RFC 4271 functional with Multi-Protocol BGP (MP-BGP) [RFC4760, RFC2545] for IPv6. The use of other extensions of BGP to support layer-2, layer-3 virtual private networks (VPN) are out of scope of this document.

The terminology used in this document is defined in [RFC4098]. One additional term is defined in this draft: FIB (Data plane) BGP Convergence.

1.2. Purpose of BGP FIB (Data Plane) Convergence

In the current Internet architecture the Inter-Autonomous System (inter-AS) transit is primarily available through BGP. To maintain a

reliable connectivity within intra-domains or across inter-domains, fast recovery from failures remains most critical. To ensure minimal traffic losses, many service providers are requiring BGP implementations to converge the entire Internet routing table within sub-seconds at FIB level.

Furthermore, to compare these numbers amongst various devices, service providers are also looking at ways to standardize the convergence measurement methods. This document offers test methods for simple topologies. These simple tests will provide a quick high-level check, of the BGP data plane convergence across multiple implementations.

1.3. Control Plane Convergence

The convergence of BGP occurs at two levels: RIB and FIB convergence. RFC 4098 defines terms for BGP control plane convergence. Methodologies which test control plane convergence are out of scope for this draft.

1.4. Benchmarking Testing

In order to ensure that the results obtained in tests are repeatable, careful setup of initial conditions and exact steps are required.

This document proposes these initial conditions, test steps, and result checking. To ensure uniformity of the results all optional parameters SHOULD be disabled and all settings SHOULD be changed to default, these may include BGP timers as well.

2. Existing Definitions and Requirements

RFC 1242, "Benchmarking Terminology for Network Interconnect Devices" [RFC1242] and RFC 2285, "Benchmarking Terminology for LAN Switching Devices" [RFC2285] SHOULD be reviewed in conjunction with this document. WLAN-specific terms and definitions are also provided in Clauses 3 and 4 of the IEEE 802.11 standard [802.11]. Commonly used terms may also be found in RFC 1983 [RFC1983].

For the sake of clarity and continuity, this document adopts the general template for benchmarking terminology set out in Section 2 of RFC 1242. Definitions are organized in alphabetical order, and grouped into sections for ease of reference. The following terms are assumed to be taken as defined in RFC 1242 [RFC1242]: Throughput, Latency, Constant Load, Frame Loss Rate, and Overhead Behavior. In addition, the following terms are taken as defined in [RFC2285]: Forwarding Rates, Maximum Forwarding Rate, Loads, Device Under Test

(DUT), and System Under Test (SUT).

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 RFC 2119 [RFC2119].

3. Test Topologies

This section describes simple test setups for use in BGP benchmarking tests measuring convergence of the FIB (data plane) after the BGP updates has been received.

These simple test nodes have 3 or 4 nodes with the following configuration:

1. Basic Test Setup
2. Three node setup for iBGP or eBGP convergence
3. Setup for eBGP multihop test scenario
4. Four node setup for iBGP or eBGP convergence

Individual tests refer to these topologies.

Figures 1-4 use the following conventions

- o AS-X: Autonomous System X
- o Loopback Int: Loopback interface on the BGP enabled device
- o R2: Helper router

3.1. General Reference Topologies

Emulator acts as 1 or more BGP peers for different testcases.

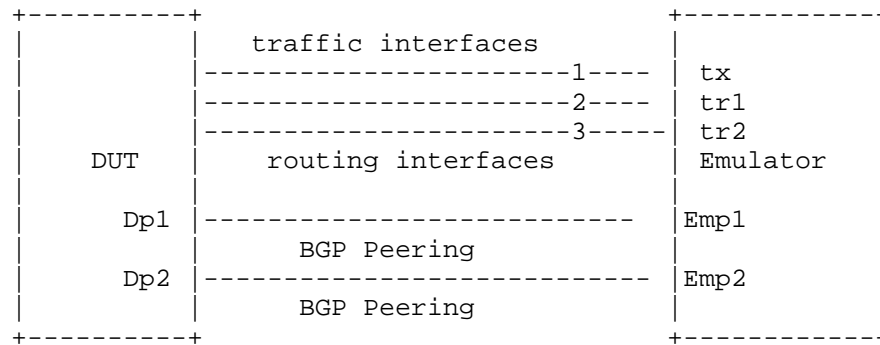


Figure 1 Basic Test Setup

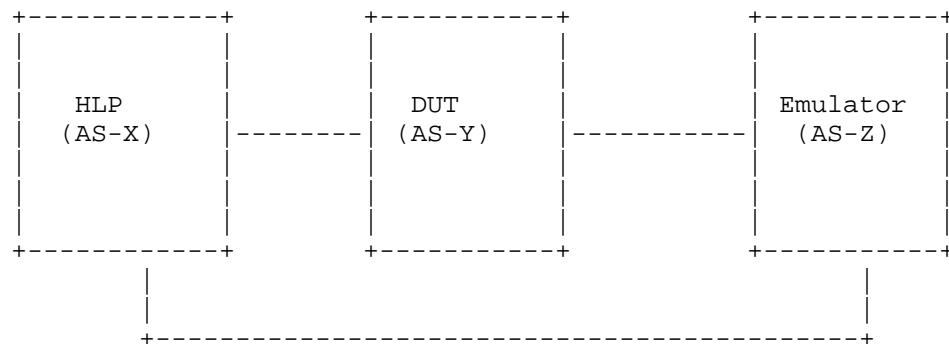


Figure 2 Three Node Setup for eBGP and iBGP Convergence

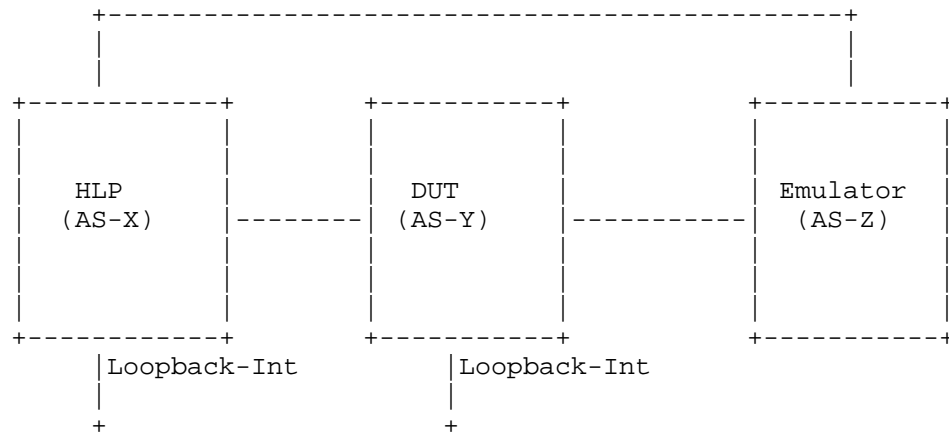


Figure 3 BGP Convergence for eBGP Multihop Scenario

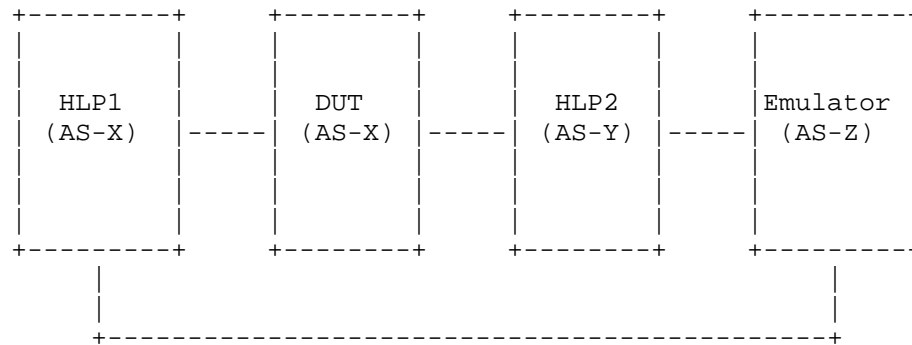


Figure 4 Four Node Setup for EBGP and IBGP Convergence

4. Test Considerations

The test cases for measuring convergence for iBGP and eBGP are different. Both iBGP and eBGP use different mechanisms to advertise, install and learn the routes. Typically, an iBGP route on the DUT is installed and exported when the next-hop is valid. For eBGP the

route is installed on the DUT with the remote interface address as the next-hop with the exception of the multihop case.

4.1. Number of Peers

Number of Peers is defined as the number of BGP neighbors or sessions the DUT has at the beginning of the test. The peers are established before the tests begin. The relationship could be either, iBGP or eBGP peering depending upon the test case requirement.

The DUT establishes one or more BGP sessions with one more emulated routers or helper nodes. Additional peers can be added based on the testing requirements. The number of peers enabled during the testing should be well documented in the report matrix.

4.2. Number of Routes per Peer

Number of Routes per Peer is defined as the number of routes advertized or learnt by the DUT per session or through neighbor relationship with an emulator or helper node. The tester, emulating as neighbor MUST advertise at least one route per peer.

Each test run must identify the route stream in terms of route packing, route mixture, and number of routes. This route stream must be well documented in the reporting stream. RFC 4098 defines these terms.

It is RECOMMENDED that the user may consider advertizing the entire current Internet routing table per peering session using an Internet route mixture with unique or non-unique routes. If multiple peers are used, it is important to precisely document the timing sequence between the peer sending routes (as defined in RFC 4098).

4.3. Policy Processing/Reconfiguration

The DUT MUST run one baseline test where policy is Minimal policy as defined in RFC 4098. Additional runs may be done with policy set-up before the tests begin. Exact policy settings should be documented as part of the test.

4.4. Configured Parameters (Timers, etc..)

There are configured parameters and timers that may impact the measured BGP convergence times.

The benchmark metrics MAY be measured at any fixed values for these configured parameters.

It is RECOMMENDED these configure parameters have the following settings: a) default values specified by the respective RFC b) platform-specific default parameters and c) values as expected in the operational network. All optional BGP settings MUST be kept consistent across iterations of any specific tests

Examples of the configured parameters that may impact measured BGP convergence time include, but are not limited to:

1. Interface failure detection timer
2. BGP Keepalive timer
3. BGP Holdtime
4. BGP update delay timer
5. ConnectRetry timer
6. TCP Segment Size
7. Minimum Route Advertisement Interval (MRAI)
8. MinASOriginationInterval (MAOI)
9. Route Flap Dampening parameters
10. TCP MD5
11. Maximum TCP Window Size
12. MTU

The basic-test settings for the parameters should be:

1. Interface failure detection timer (0 ms)
2. BGP Keepalive timer (1 min)
3. BGP Holdtime (3 min)
4. BGP update delay timer (0 s)

5. ConnectRetry timer (1 s)
6. TCP Segment Size (4096)
7. Minimum Route Advertisement Interval (MRAI) (0 s)
8. MinASOriginationInterval (MAOI)(0 s)
9. Route Flap Dampening parameters (off)
10. TCP MD5 (off)

4.5. Interface Types

The type of media dictate which test cases may be executed, each interface type has unique mechanism for detecting link failures and the speed at which that mechanism operates will influence the measurement results. All interfaces MUST be of the same media and throughput for each test case.

4.6. Measurement Accuracy

Since observed packet loss is used to measure the route convergence time, the time between two successive packets offered to each individual route is the highest possible accuracy of any packet-loss based measurement. When packet jitter is much less than the convergence time, it is a negligible source of error and hence it will be treated as within tolerance.

Other options to measure convergence are the Time-Based Loss Method (TBLM) and Timestamp Based Method(TBM)[MPLSProt]

An exterior measurement on the input media (such Ethernet)is defined by this specification.

4.7. Measurement Statistics

The benchmark measurements may vary for each trial, due to the statistical nature of timer expirations, CPU scheduling, etc. It is recommended to repeat the test multiple times. Evaluation of the test data must be done with an understanding of generally accepted testing practices regarding repeatability, variance and statistical significance of a small number of trials.

For any repeated tests that are averaged to remove variance, all parameters MUST remain the same.

4.8. Authentication

Authentication in BGP is done using the TCP MD5 Signature Option [RFC5925]. The processing of the MD5 hash, particularly in devices with a large number of BGP peers and a large amount of update traffic, can have an impact on the control plane of the device. If authentication is enabled, it SHOULD be documented correctly in the reporting format

4.9. Convergence Events

Convergence events or triggers are defined as abnormal occurrences in the network, which initiate route flapping in the network, and hence forces the re-convergence of a steady state network. In a real network, a series of convergence events may cause convergence latency operators desire to test.

These convergence events must be defined in terms of the sequences defined in RFC 4098. This basic document begins all tests with a router initial set-up. Additional documents will define BGP data plane convergence based on peer initialization.

The convergence events may or may not be tied to the actual failure. A Soft Reset (RFC 4098) does not clear the RIB or FIB tables. A Hard reset clears the BGP peer sessions, the RIB tables, and FIB tables.

4.10. High Availability

Due to the different Non-Stop-Routing (sometimes referred to High-Availability) solutions available from different vendors, it is RECOMMENDED that any redundancy available in the routing processors should be disabled during the convergence measurements.

5. Test Cases

All tests defined under this section assume the following:

- a. BGP peers should be brought to BGP Peer established state
- b. Furthermore the traffic generation and routing should be verified in the topology

5.1. Basic Convergence Tests

These test cases measure characteristics of a BGP implementation in non-failure scenarios like:

1. RIB-IN Convergence
2. RIB-OUT Convergence
3. eBGP Convergence
4. iBGP Convergence

5.1.1. RIB-IN Convergence

Objective:

This test measures the convergence time taken to receive and install a route in RIB using BGP

Reference Test Setup:

This test uses the setup as shown in figure 1

Procedure:

- A. All variables affecting Convergence should be set to a basic test state (as defined in section 4-4).
- B. Establish BGP adjacency between DUT and peer x of Emulator.
- C. To ensure adjacency establishment, wait for 3 KeepAlives from the DUT or a configurable delay before proceeding with the rest of the test.
- D. Start the traffic from the Emulator peer-x towards the DUT targeted at a routes specified in route mixture (ex. route A) Initially no traffic SHOULD be observed on the egress interface as the route A is not installed in the forwarding database of the DUT.
- E. Advertise route A from the Peer-x to the DUT and record the time.

This is $T_{up}(EMx, Rt-A)$ also named 'XMT-Rt-time'.

- F. Record the time when the route-A from Peer-x is received at the DUT.

This $Tup(DUT, Rt-A)$ also named 'RCV-Rt-time'.

- G. Record the time when the traffic targeted towards route A is received by Emulator on appropriate traffic egress interface.

This is $TR(TDr, Rt-A)$. This is also named DUT-XMT-Data-Time.

- H. The difference between the $Tup(DUT, RT-A)$ and traffic received time ($TR(TDr, Rt-A)$) is the FIB Convergence Time for route-A in the route mixture. A full convergence for the route update is the measurement between the 1st route (Route-A) and the last route ($Rt-last$)

Route update convergence is

$TR(TDr, RT-last) - Tup(DUT, Rt-A)$ or

$(DUT-XMT-Data-Time - RCV-Rt-Time)(Rt-A)$

Note: It is recommended that a single test with the same route mixture be repeated several times. A report should provide the Standard Deviation of all tests and the Average.

Running tests with a varying number of routes and route mixtures is important to get a full characterization of a single peer.

5.1.2. RIB-OUT Convergence

Objective:

This test measures the convergence time taken by an implementation to receive, install and advertise a route using BGP

Reference Test Setup:

This test uses the setup as shown in figure 2

Procedure:

- A. The Helper node (HLP) run same version of BGP as DUT.

- B. All devices MUST be synchronized using NTP or some local reference clock.
- C. All configuration variables for HLP, DUT, and Emulator SHOULD be set to the same values. These values MAY be basic-test or a unique set completely described in the test set-up.
- D. Establish BGP adjacency between DUT and Emulator.
- E. Establish BGP adjacency between DUT and Helper Node.
- F. To ensure adjacency establishment, wait for 3 KeepAlives from the DUT or a configurable delay before proceeding with the rest of the test
- G. Start the traffic from the Emulator towards the Helper Node targeted at a specific route say route A. Initially no traffic SHOULD be observed on the egress interface as the route-A is not installed in the forwarding database of the DUT.
- H. Advertise routeA from the Emulator to the DUT and note the time.

This is $Tup(EMx, Route-A)$. (also named EM-XMT-Rt-Time)

- I. Record when Route-A is received by DUT.

This is $Tup(DUTr, Route-A)$. (also named DUT-RCV-Rt-Time)

- J. Record the time when the ROUTE is forwarded by DUT towards the Helper node.

This is $Tup(DUTx, Route-A)$. (also named DUT-XMT-Rt-Time)

- K. Record the time when the traffic targeted towards route-A is received on the Route Egress Interface. This is $TR(EMr, Route-A)$. (also named DUT-XMT-Data Time).

$FIB\ convergence = (DUT-RCV-Rt-Time - DUT-XMT-Data-Time)$

$RIB\ convergence = (DUT-RCV-Rt-Time - DUT-XMT-Rt-Time)$

Convergence for a route stream is characterized by

a) Individual route convergence for FIB, RIB

b) All route convergence of

FIB-convergence =DUT-RCV-Rt-Time(A)-DUT-XMT-Data-Time(last)

RIB-convergence =DUT-RCV-Rt-Time(A)-DUT-XMT-Rt-Time(last)

5.1.3. eBGP Convergence

Objective:

This test measures the convergence time taken by an implementation to receive, install and advertise a route in an eBGP Scenario

Reference Test Setup:

This test uses the setup as shown in figure 2 and the scenarios described in RIB-IN and RIB-OUT are applicable to this test case.

5.1.4. iBGP Convergence

Objective:

This test measures the convergence time taken by an implementation to receive, install and advertise a route in an iBGP Scenario

Reference Test Setup:

This test uses the setup as shown in figure 2 and the scenarios described in RIB-IN and RIB-OUT are applicable to this test case.

5.1.5. eBGP Multihop Convergence

Objective:

This test measures the convergence time taken by an implementation to receive, install and advertise a route in an eBGP Multihop Scenario

Reference Test Setup:

This test uses the setup as shown in figure 3. DUT is used along with a helper node.

Procedure:

- A. The Helper Node (HLP) runs the same BGP version as DUT
- B. All devices to be synchronized using NTP
- C. All variables affecting Convergence like authentication, policies, timers should be set to basic-settings
- D. All 3 devices, DUT, Emulator and Helper Node are configured with different Autonomous Systems
- E. Loopback Interfaces are configured on DUT and Helper Node and connectivity is established between them using any config options available on the DUT
- F. Establish BGP adjacency between DUT and Emulator
- G. Establish BGP adjacency between DUT and Helper Node
- H. To ensure adjacency establishment, wait for 3 KeepAlives from the DUT or a configurable delay before proceeding with the rest of the test
- I. Start the traffic from the Emulator towards the DUT targeted at a specific route say routeA
- J. Initially no traffic SHOULD be observed on the egress interface as the routeA is not installed in the forwarding database of the DUT
- K. Advertise routeA from the Emulator to the DUT and note the time (Tup(EMx,RouteA) also named (Route-Tx-time)
- L. Record the time when the route is received by the DUT. This is Tup(EMr,DUT) named (Route-Rcv-time)
- M. Record the time when the traffic targeted towards routeA is received from Egress Interface of DUT on emulator. This is Tup(EMd,DUT) named (Data-Rcv-time)
- N. Record the time when the routeA is forwarded by DUT towards the Helper node. This is Tup(EMf,DUT) also named (Route-Fwd-time)

FIB Convergence = (Data-Rcv-time - Route-Rcv-time)

RIB Convergence = (Route-Fwd-time - Route-Rcv-time)

Note: It is recommended that the test be repeated with varying number

of routes and route mixtures. With each set route mixture, the test should be repeated multiple times. The results should record average, mean, Standard Deviation

5.2. BGP Failure/Convergence Events

5.2.1. Physical Link Failure on DUT End

Objective:

This test measures the route convergence time due to local link failure event at DUT's Local Interface

Reference Test Setup:

This test uses the setup as shown in figure 1. Shutdown event is defined as an administrative shutdown event on the DUT

Procedure:

- A. All variables affecting Convergence like authentication, policies, timers should be set to basic-test policy
- B. Establish 2 BGP adjacencies from DUT to Emulator, one over the peer interface and the other using a second peer interface
- C. Advertise the same route, route A over both the adjacencies and (Tx1)Interface to be the preferred next hop
- D. To ensure adjacency establishment, wait for 3 KeepAlives from the DUT or a configurable delay before proceeding with the rest of the test
- E. Start the traffic from the Emulator towards the DUT targeted at a specific route say route A. Initially traffic would be observed on the best egress route (Emp1) instead of Trr2
- F. Trigger the shutdown event of Best Egress Interface on DUT (Drr1)
- G. Measure the Convergence Time for the event to be detected and traffic to be forwarded to Next-Best Egress Interface (rr2)

Time = Data-detect(rr2) - Shutdown time

H. Stop the offered load and wait for the queues to drain and Restart

I. Bring up the link on DUT Best Egress Interface

J. Measure the convergence time taken for the traffic to be rerouted from (rr2) to Best Interface (rr1)

Time = Data-detect(rr1) - Bring Up time

K. It is recommended that the test be repeated with varying number of routes and route mixtures or with number of routes & route mixtures closer to what is deployed in operational networks

5.2.2. Physical Link Failure on Remote/Emulator End

Objective:

This test measures the route convergence time due to local link failure event at Tester's Local Interface

Reference Test Setup:

This test uses the setup as shown in figure 1. Shutdown event is defined as shutdown of the local interface of Tester via logical shutdown event. The procedure used in 5.2.1 is used for the termination

5.2.3. ECMP Link Failure on DUT End

Objective:

This test measures the route convergence time due to local link failure event at ECMP Member. The FIB configuration and BGP is set to allow two ECMP routes to be installed. However, policy directs the routes to be sent only over one of the paths

Reference Test Setup:

This test uses the setup as shown in figure 1 and the procedure uses 5.2.1

5.3. BGP Adjacency Failure (Non-Physical Link Failure) on Emulator

Objective:

This test measures the route convergence time due to BGP Adjacency Failure on Emulator

Reference Test Setup:

This test uses the setup as shown in figure 1

Procedure:

- A. All variables affecting Convergence like authentication, policies, timers should be basic-policy set
- B. Establish 2 BGP adjacencies from DUT to Emulator, one over the Best Egress Interface and the other using the Next-Best Egress Interface
- C. Advertise the same route, routeA over both the adjacencies and make Best Egress Interface to be the preferred next hop
- D. To ensure adjacency establishment, wait for 3 KeepAlives from the DUT or a configurable delay before proceeding with the rest of the test
- E. Start the traffic from the Emulator towards the DUT targeted at a specific route say routeA. Initially traffic would be observed on the Best Egress interface
- F. Remove BGP adjacency via a software adjacency down on the Emulator on the Best Egress Interface. This time is called BGPadj-down-time also termed BGPpeer-down
- G. Measure the Convergence Time for the event to be detected and traffic to be forwarded to Next-Best Egress Interface. This time is Tr-rr2 also called TR2-traffic-on

$$\text{Convergence} = \text{TR2-traffic-on} - \text{BGPpeer-down}$$

- H. Stop the offered load and wait for the queues to drain and Restart
- I. Bring up BGP adjacency on the Emulator over the Best Egress Interface. This time is BGP-adj-up also called BGPpeer-up
- J. Measure the convergence time taken for the traffic to be rerouted to Best Interface. This time is BGP-adj-up also called BGPpeer-up

5.4. BGP Hard Reset Test Cases

5.4.1. BGP Non-Recovering Hard Reset Event on DUT

Objective:

This test measures the route convergence time due to Hard Reset on the DUT

Reference Test Setup:

This test uses the setup as shown in figure 1

Procedure:

- A. The requirement for this test case is that the Hard Reset Event should be non-recovering and should affect only the adjacency between DUT and Emulator on the Best Egress Interface
- B. All variables affecting SHOULD be set to basic-test values
- C. Establish 2 BGP adjacencies from DUT to Emulator, one over the Best Egress Interface and the other using the Next-Best Egress Interface
- D. Advertise the same route, routeA over both the adjacencies and make Best Egress Interface to be the preferred next hop
- E. To ensure adjacency establishment, wait for 3 KeepAlives from the DUT or a configurable delay before proceeding with the rest of the test
- F. Start the traffic from the Emulator towards the DUT targeted at a specific route say routeA. Initially traffic would be observed on the Best Egress interface
- G. Trigger the Hard Reset event of Best Egress Interface on DUT
- H. Measure the Convergence Time for the event to be detected and traffic to be forwarded to Next-Best Egress Interface

Time of convergence = time-traffic flow - time-reset

- I. Stop the offered load and wait for the queues to drain and Restart
- J. It is recommended that the test be repeated with varying number of routes and route mixtures or with number of routes & route mixtures closer to what is deployed in operational networks
- K. When varying number of routes are used, convergence Time is measured using the Loss Derived method [IGPData]
- L. Convergence Time in this scenario is influenced by Failure detection time on Tester, BGP Keep Alive Time and routing, forwarding table update time

5.5. BGP Soft Reset

Objective:

This test measures the route convergence time taken by an implementation to service a BGP Route Refresh message and advertise a route

Reference Test Setup:

This test uses the setup as shown in figure 2

Procedure:

- A. The BGP implementation on DUT & Helper Node needs to support BGP Route Refresh Capability [RFC2918]
- B. All devices to be synchronized using NTP
- C. All variables affecting Convergence like authentication, policies, timers should be set to basic-test defaults
- D. DUT and Helper Node are configured in the same Autonomous System whereas Emulator is configured under a different Autonomous System
- E. Establish BGP adjacency between DUT and Emulator
- F. Establish BGP adjacency between DUT and Helper Node

- G. To ensure adjacency establishment, wait for 3 KeepAlives from the DUT or a configurable delay before proceeding with the rest of the test
- H. Configure a policy under BGP on Helper Node to deny routes received from DUT
- I. Advertise routeA from the Emulator to the DUT
- J. The DUT will try to advertise the route to Helper Node will be denied
- K. Wait for 3 KeepAlives
- L. Start the traffic from the Emulator towards the Helper Node targeted at a specific route say routeA. Initially no traffic would be observed on the Egress interface, as routeA is not present
- M. Remove the policy on Helper Node and issue a Route Refresh request towards DUT. Note the timestamp of this event. This is the RefreshTime
- N. Record the time when the traffic targeted towards routeA is received on the Egress Interface. This is RecTime
- O. The following equation represents the Route Refresh Convergence Time per route

$$\text{Route Refresh Convergence Time} = (\text{RecTime} - \text{RefreshTime})$$

5.6. BGP Route Withdrawal Convergence Time

Objective:

This test measures the route convergence time taken by an implementation to service a BGP Withdraw message and advertise the withdraw

Reference Test Setup:

This test uses the setup as shown in figure 2

Procedure:

- A. This test consists of 2 steps to determine the Total Withdraw Processing Time
- B. Step 1:
- (1) All devices to be synchronized using NTP
 - (2) All variables should be set to basic-test parameters
 - (3) DUT and Helper Node are configured in the same Autonomous System whereas Emulator is configured under a different Autonomous System
 - (4) Establish BGP adjacency between DUT and Emulator
 - (5) To ensure adjacency establishment, wait for 3 KeepAlives from the DUT or a configurable delay before proceeding with the rest of the test
 - (6) Start the traffic from the Emulator towards the DUT targeted at a specific route say routeA. Initially no traffic would be observed on the Egress interface as the routeA is not present on DUT
 - (7) Advertise routeA from the Emulator to the DUT
 - (8) The traffic targeted towards routeA is received on the Egress Interface
 - (9) Now the Tester sends request to withdraw routeA to DUT, TRx(Awith) also called WdrawTime1
 - (10) Record the time when no traffic is observed on the Egress Interface. This is the RouteRemoveTime1(A)

WdrawConvTime1 = RouteRemoveTime1(A)
 - (11) The difference between the RouteRemoveTime1 and WdrawTime1 is the WdrawConvTime1
- C. Step 2:
- (1) Continuing from Step 1, re-advertise routeA back to DUT from Tester

- (2) The DUT will try to advertise the routeA to Helper Node (assumption there exists a session between DUT and helper node)
- (3) Start the traffic from the Emulator towards the Helper Node targeted at a specific route say routeA. Traffic would be observed on the Egress interface after routeA is received by the Helper Node

WATime=time traffic first flows

- (4) Now the Tester sends a request to withdraw routeA to DUT. This is the WdrawTime2

WAWtime-TRx(RouteA) = WdrawTime2

- (5) DUT processes the withdraw and sends it to Helper Node
- (6) Record the time when no traffic is observed on the Egress Interface of Helper Node. This is

TR-WAW(DUT,RouteA) = RouteRemoveTime2

- (7) Total withdraw processing time is

TotalWdrawTime = ((RouteRemoveTime2 - WdrawTime2) - WdrawConvTime1)

5.7. BGP Path Attribute Change Convergence Time

Objective:

This test measures the convergence time taken by an implementation to service a BGP Path Attribute Change

Reference Test Setup:

This test uses the setup as shown in figure 1

Procedure:

- A. This test only applies to Well-Known Mandatory Attributes like Origin, AS Path, Next Hop
- B. In each iteration of test only one of these mandatory attributes need to be varied whereas the others remain the

same

- C. All devices to be synchronized using NTP
- D. All variables should be set to basic-test parameters
- E. Advertise the route, routeA over the Best Egress Interface only, making it the preferred named Tbest
- F. To ensure adjacency establishment, wait for 3 KeepAlives from the DUT or a configurable delay before proceeding with the rest of the test
- G. Start the traffic from the Emulator towards the DUT targeted at the specific route say routeA. Initially traffic would be observed on the Best Egress interface
- H. Now advertise the same route routeA on the Next-Best Egress Interface but by varying one of the well-known mandatory attributes to have a preferred value over that interface. We call this Tbetter. The other values need to be same as what was advertised on the Best-Egress adjacency

$TRx(\text{Path-Change}) = \text{Path Change Event Time}$

- I. Measure the Convergence Time for the event to be detected and traffic to be forwarded to Next-Best Egress Interface

$DUT(\text{Path-Change}, \text{RouteA}) = \text{Path-switch time}$

$\text{Convergence} = \text{Path-switch time} - \text{Path Change Event Time}$

- J. Stop the offered load and wait for the queues to drain and Restart
- K. Repeat the test for various attributes

5.8. BGP Graceful Restart Convergence Time

Objective:

This test measures the route convergence time taken by an implementation during a Graceful Restart Event

Reference Test Setup:

This test uses the setup as shown in figure 4

Procedure:

- A. It measures the time taken by an implementation to service a BGP Graceful Restart Event and advertise a route
- B. The Helper Nodes are the same model as DUT and run the same BGP implementation as DUT
- C. The BGP implementation on DUT & Helper Node needs to support BGP Graceful Restart Mechanism [RFC4724]
- D. All devices to be synchronized using NTP
- E. All variables are set to basic-test values
- F. DUT and Helper Node-1(HLP1) are configured in the same Autonomous System whereas Emulator and Helper Node-2(HLP2) are configured under different Autonomous Systems
- G. Establish BGP adjacency between DUT and Helper Nodes
- H. Establish BGP adjacency between Helper Node-2 and Emulator
- I. To ensure adjacency establishment, wait for 3 KeepAlives from the DUT or a configurable delay before proceeding with the rest of the test
- J. Configure a policy under BGP on Helper Node-1 to deny routes received from DUT
- K. Advertise routeA from the Emulator to Helper Node-2
- L. Helper Node-2 advertises the route to DUT and DUT will try to advertise the route to Helper Node-1 which will be denied
- M. Wait for 3 KeepAlives
- N. Start the traffic from the Emulator towards the Helper Node-1 targeted at the specific route say routeA. Initially no traffic would be observed on the Egress interface as the routeA is not present
- O. Perform a Graceful Restart Trigger Event on DUT and note the time. This is the GREventTime

- P. Remove the policy on Helper Node-1
- Q. Record the time when the traffic targeted towards routeA is received on the Egress Interface

TRr(DUT, routeA). This is also called RecTime
- R. The following equation represents the Graceful Restart Convergence Time

$$\text{Graceful Restart Convergence Time} = ((\text{RecTime} - \text{GREventTime}) - \text{RIB-IN})$$
- S. It is assumed in this test case that after a Switchover is triggered on the DUT, it will not have any cycles to process BGP Refresh messages. The reason for this assumption is that there is a narrow window of time where after switchover when we remove the policy from Helper Node -1, implementations might generate Route-Refresh automatically and this request might be serviced before the DUT actually switches over and reestablishes BGP adjacencies with the peers

6. Reporting Format

For each test case, it is recommended that the reporting tables below are completed and all time values SHOULD be reported with resolution as specified in [RFC4098]

Parameter	Units
Test case	Test case number
Test topology	1,2,3 or 4
Parallel links	Number of parallel links
Interface type	GigE, POS, ATM, other
Convergence Event	Hard reset, Soft reset, link failure, or other defined
eBGP sessions	Number of eBGP sessions
iBGP sessions	Number of iBGP sessions
eBGP neighbor	Number of eBGP neighbors
iBGP neighbor	Number of iBGP neighbors
Routes per peer	Number of routes
Total unique routes	Number of routes
Total non-unique routes	Number of routes
IGP configured	ISIS, OSPF, static, or other
Route Mixture	Description of Route mixture
Route Packing	Number of routes in an update
Policy configured	Yes, No
Packet size offered to the DUT	Bytes
Offered load	Packets per second
Packet sampling interval on tester	Seconds
Forwarding delay threshold	Seconds
Timer Values configured on DUT	
Interface failure indication delay	Seconds
Hold time	Seconds
MinRouteAdvertisementInterval (MRAI)	Seconds
MinASOriginationInterval (MAOI)	Seconds
Keepalive Time	Seconds
ConnectRetry	Seconds
TCP Parameters for DUT and tester	
MSS	Bytes
Slow start threshold	Bytes
Maximum window size	Bytes

Test Details:

- a. If the Offered Load matches a subset of routes, describe how this subset is selected
- b. Describe how the Convergence Event is applied; does it cause instantaneous traffic loss or not

c. If there is any policy configured, describe the configured policy

Complete the table below for the initial Convergence Event and the reversion Convergence Event

Parameter	Unit
Convergence Event	Initial or reversion
Traffic Forwarding Metrics	
Total number of packets offered to DUT	Number of packets
Total number of packets forwarded by DUT	Number of packets
Connectivity Packet Loss	Number of packets
Convergence Packet Loss	Number of packets
Out-of-order packets	Number of packets
Duplicate packets	Number of packets
Convergence Benchmarks	
Rate-derived Method [IGP-Data]:	
First route convergence time	Seconds
Full convergence time	Seconds
Loss-derived Method [IGP-Data]:	
Loss-derived convergence time	Seconds
Route-Specific Loss-Derived Method:	
Minimum R-S convergence time	Seconds
Maximum R-S convergence time	Seconds
Median R-S convergence time	Seconds
Average R-S convergence time	Seconds
Loss of Connectivity Benchmarks	
Loss-derived Method:	
Loss-derived loss of connectivity period	Seconds
Route-Specific loss-derived Method:	
Minimum LoC period [n]	Array of seconds
Minimum Route LoC period	Seconds
Maximum Route LoC period	Seconds
Median Route LoC period	Seconds

Average Route LoC period Seconds

7. IANA Considerations

This draft does not require any new allocations by IANA.

8. Security Considerations

Benchmarking activities as described in this memo are limited to technology characterization using controlled stimuli in a laboratory environment, with dedicated address space and the constraints specified in the sections above.

The benchmarking network topology will be an independent test setup and MUST NOT be connected to devices that may forward the test traffic into a production network, or misroute traffic to the test management network.

Further, benchmarking is performed on a "black-box" basis, relying solely on measurements observable external to the DUT/SUT.

Special capabilities SHOULD NOT exist in the DUT/SUT specifically for benchmarking purposes. Any implications for network security arising from the DUT/SUT SHOULD be identical in the lab and in production networks.

9. References

9.1. Normative References

- [I-D.ietf-bmwg-igp-dataplane-conv-term]
Poretsky, S., Imhoff, B., and K. Michielsen, "Terminology for Benchmarking Link-State IGP Data Plane Route Convergence", draft-ietf-bmwg-igp-dataplane-conv-term-23 (work in progress), February 2011.
- [RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.
- [RFC2918] Chen, E., "Route Refresh Capability for BGP-4", RFC 2918, September 2000.
- [RFC4271] Rekhter, Y., Li, T., and S. Hares, "A Border Gateway Protocol 4 (BGP-4)", RFC 4271, January 2006.

9.2. Informative References

- [RFC1242] Bradner, S., "Benchmarking terminology for network interconnection devices", RFC 1242, July 1991.
- [RFC1983] Malkin, G., "Internet Users' Glossary", RFC 1983, August 1996.
- [RFC2285] Mandeville, R., "Benchmarking Terminology for LAN Switching Devices", RFC 2285, February 1998.
- [RFC2545] Marques, P. and F. Dupont, "Use of BGP-4 Multiprotocol Extensions for IPv6 Inter-Domain Routing", RFC 2545, March 1999.
- [RFC4098] Berkowitz, H., Davies, E., Hares, S., Krishnaswamy, P., and M. Lepp, "Terminology for Benchmarking BGP Device Convergence in the Control Plane", RFC 4098, June 2005.
- [RFC4724] Sangli, S., Chen, E., Fernando, R., Scudder, J., and Y. Rekhter, "Graceful Restart Mechanism for BGP", RFC 4724, January 2007.
- [RFC4760] Bates, T., Chandra, R., Katz, D., and Y. Rekhter, "Multiprotocol Extensions for BGP-4", RFC 4760, January 2007.
- [RFC5925] Touch, J., Mankin, A., and R. Bonica, "The TCP Authentication Option", RFC 5925, June 2010.

Authors' Addresses

Rajiv Papneja
Huawei Technologies

Email: rajiv.papneja@huawei.com

Bhavani Parise
Cisco Systems

Email: bhavani@cisco.com

Susan Hares
Huawei Technologies (USA)

Email: shares@huawei.com

Ilya Varlashkin
Easynet Global Services

Email: ilya.varlashkin@easynet.com

Dean Lee
Ixia

Email: dlee@ixiacom.com

Eric Brendel
Independent Consultant

Email: brendel@pektel.com

Mohan Nanduri
Microsoft

Email: mnanduri@microsoft.com

Jay Karthik
Cisco Systems

Email: jkarthik@cisco.com

Internet Engineering Task Force
Internet-Draft
Intended status: Informational
Expires: April 22, 2012

I. Varlashkin
Easynet Global Services
R. Papneja
Huawei Technologies (USA)
B. Parise
Cisco
T. Van Unen
Ixia
October 20, 2011

Convergence benchmarking on contemporary routers
draft-varlashkin-router-conv-bench-00

Abstract

This document specifies methodology for benchmarking convergence of routers without making assumptions about relation and dependencies between data- and control-planes. Provided methodology is primary intended for testing routers running BGP and some form of link-state IGP with or without MPLS. It may also be applicable for environments using MPLS-TE or GRE, however they're beyond scope of this document and such application is left for further study.

Status of this Memo

This Internet-Draft is submitted in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at <http://datatracker.ietf.org/drafts/current/>.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

This Internet-Draft will expire on April 22, 2012.

Copyright Notice

Copyright (c) 2011 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust's Legal Provisions Relating to IETF Documents

(<http://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

This document may contain material from IETF Documents or IETF Contributions published or made publicly available before November 10, 2008. The person(s) controlling the copyright in some of this material may not have granted the IETF Trust the right to allow modifications of such material outside the IETF Standards Process. Without obtaining an adequate license from the person(s) controlling the copyright in such materials, this document may not be modified outside the IETF Standards Process, and derivative works of it may not be created outside the IETF Standards Process, except to format it for publication as an RFC or to translate it into languages other than English.

Table of Contents

1. Introduction	4
2. Test topology	5
3. TEST PARAMETERS	6
3.1. Packing ratios	7
3.2. Test traffic	7
3.3. IGP metrics	7
3.4. Internal routers matrix	7
3.5. Number of next-hops	8
3.6. 'e' - Failure and Restoration start entropy	8
4. TEST PROCEDURES	8
4.1. Initialisation time	8
4.2. Generic data-plane failure test	9
4.3. Generic test procedure for	9
5. Failure and restoration scenarios	10
5.1. Loss of Signal on the link attached to DUT	10
5.2. Link failure without LoS	10
5.3. Non-direct link failure	11
5.4. Best route withdrawal	11
5.5. iBGP next-hop failure	12
6. Test report	12
7. Link bundling and Equal Cost Multi-Path	13
8. Graceful Restart and Non-Stop Forwarding	13
9. Security considerations	13
10. IANA Considerations	14
11. Acknowledgments	14
12. Normative References	14
Authors' Addresses	14

1. Introduction

Ability of the network to restore traffic flow when primary path fails has always been important subject for network engineers, researchers and equipment manufacturers. Time to recover from a link or node failure has often been linked to routing protocols convergence; and benchmarking of a routing protocol convergence has often been considered sufficient for quantifying recovery performance. As long as routers could obtain new best path only after relevant routing protocols perform their calculations such methodology was reasonable. However continuous improvements in hardware and software result in more and more routers being able to restore traffic flow even before routing protocols converge. Methodology described in this document takes such fact into account.

When a failure occurs on the network a router needs to:

1. select new best path so that the packets, which already arrived to the router, can be forwarded
2. let other routers know about new network state so they can find new best path from their perspective

How fast a router can perform these two functions characterise router's performance with regards to convergence. Note that in general case each of these characteristics may or may not be related to the other. For example, some platform may need to perform calculations to find new best path and only then update local FIB and send relevant protocol updates to other routers, another platform can update local FIB without waiting for calculations to complete but still needs to wait for calculations before sending routing protocol updates, third platform can use different optimisation for both FIB changes and routing protocol updates without waiting for completion of the calculations. Other variations are also possible. This document makes no assumption about whether local FIB changes and routing protocol updates dependencies on each other or on routing protocol calculations.

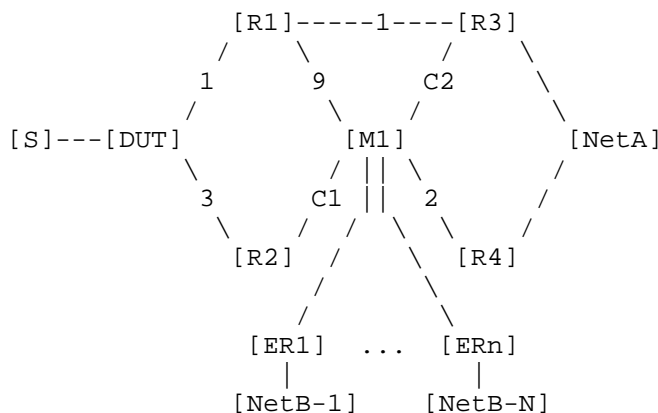
Since it is not known whether local FIB is updated before or after routing protocol calculations, forwarding-plane method is proposed to benchmark local convergence. And because it is not known whether routing protocol updates are linked to FIB modification or not the control-plane approach is used to benchmark how fast updates are propagated. However both characteristics are benchmarked using very similar test topologies and procedures. Also, an attempt is made to to minimise dependency on performance on non-DUT elements involved in the tests.

At the time of writing of this document it is not known whether existing network testers and protocol emulators are able to execute described tests out of the box. Nevertheless the authors believe that required functionality can be added with reasonable effort. Alternatively the tests can be performed with help of physical routers to create necessary test topology, which may have impact on time required to perform the test but expected to provide same degree of the test results accuracy. This also means that tests performed using a protocol simulator can be repeated using physical routers and results expected to be comparable.

This document complements draft-papneja-bgp-basic-dp-convergence.

2. Test topology

Unless specified otherwise all tests use same basic test topology outlined below:



S is source of test traffic for data-plane tests, while for control-plane tests S is an emulated or physical router with packet capturing (sniffing) capability.

Unidirectional test traffic goes from Source to NetA.

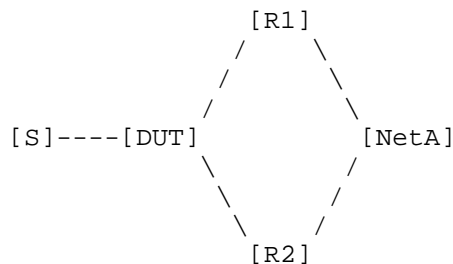
IGP between DUT and R1-R4; BGP between DUT and R3, R4; no BGP between R3 and R4 (important). If tunnelling (e.g. MPLS or GRE) is used then R1 and R2 do not need to run BGP, otherwise they MUST run BGP. Source has static default to DUT; R3 and R4 have static to NetA. NetA is in BGP but not in IGP. M1 is K*M matrix of internal routers. Metrics C1 is used to control whether R2 is LFA for DUT to NetA. Metric C2 is used to control whether R3 or R4 are best exit towards NetA. All other metrics are fixed for all tests and MUST be set to

exact values provided in the above diagram. IGP metrics from M1 to ER1 throughout ERn can be set arbitrarily, their exact values are irrelevant to this test as long as they're valid for given IGP.

Routers ER1 throughout ERn together with prefixes NetB-1 throughout NetB-N are presented to create realistic environment but not used directly in measurements. NetB-1 throughout NetB-N are distinct single-prefix sets.

Traffic restoration depends on ability of R2 and M1 to forward traffic after failure. To eliminate this dependency R2 is set to always forward traffic to R3 and NetA via M1 which in turn always forwards traffic directly via R3 or R2 depending on the test. One possibility to achieve this is to use static routes. Another alternative is to use different IGP between R2 and R3 from the one used by DUT and make routes learned via this IGP preferred on R2. E.g. DUT uses OSPF, then in addition to it R2&R3 also run ISIS and prefer ISIS routes over OSPF ones. A protocol simulator can have internal mechanism to provide required behaviour. There are no other dependencies on non-DUT devices in this tests.

For evaluating eBGP performance following topology is used:



Test topology for eBGP

In "Link failure without LoS" test direct cable between DUT and R1 is replaced with connection over an L2 switch as follow:

[DUT]---[SW1]---[R1]

3. TEST PARAMETERS

3.1. Packing ratios

Routes with different prefixes but same attributes can potentially be packed into single update message. Since both number of update messages and number of prefixes per update can affect convergence time, the tests SHOULD be performed with various prefix packing ratios. This document does not specify values of individual BGP attributes used to control packing ratio.

3.2. Test traffic

Traffic is sent from single source address located at the Source port of the tester to one address in each prefix in NetA set. Packets are sent at rate 1000 per second, which provides 1ms resolution of the convergence time as measured by tests in this document. All packets SHOULD be 64 bytes at IP layer, that is IP header plus IP payload.

3.3. IGP metrics

Basic test topology specifies fixed IGP metrics for some links. These metrics SHOULD be used verbatim. There are also two variable metrics - C1 and C2 - intended for controlling whether R2 is Loop-Free-Alternate (LFA) for DUT towards NetA, and whether R3 remains best exit towards NetA after path failure between DUT and R3. Following values SHOULD be used for C1 and C2 depending on required behaviour:

R2 is LFA?	R3 best?	C1	C2
yes	yes	1	1
yes	no	1	3
no	yes	5	1
no	no	5	3

3.4. Internal routers matrix

Basic test topology has N*K grid of internal routers denoted as M1. When N>1 or K>1 the cost of all links within grid MUST be set to 1 (one). This matrix is intended for controlling topology size, which has affect on particularly SPF run-time.

If traffic is forwarded using a tunneling mechanism, such as MPLS or GRE, the internal routers only need to have reachability information about tunnel end-points. However if traditional hop-by-hop forwarding is used, then internal routers MUST have routes to each and every prefix within NetA set.

This document does not specify how internal routers should obtain necessary reachability information. The only requirement is that after primary DUT-NetA path failure internal routers are able to forward traffic to NetA instantly. Using values of IGP metrics as described earlier addresses this requirement. Also, protocol simulator may have built-in mechanism to achieve desired behaviour.

3.5. Number of next-hops

Basic test topology has set of N edge routers ER1 throughout ERn, each advertising unique prefix. Some BGP implementations may exhibit different performance depending on number of next-hops for which IGP cost has changed after failure. By varying overall number of next-hops such dependency can be detected.

Note that prefixes NetB-1 throughout NetB-n are not used as destinations for test traffic, they're only present for creating "background environment".

3.6. 'e' - Failure and Restoration start entropy

Tests described in this document use fixed time T2 and variable offset 'e' as starting point for simulating failure or restoration event.

Fixing time T2 is necessary as reference point to which variable offset e is added for each iteration of the test. Introduction of such variable offset allows better analysis of the test results. For example, DUT may run FIB changes at certain intervals. If failure introduced close to the end of such interval, shorter outage will be observed, and if introduced close to the beginning of such interval longer outage will be observed. Running test multiple times each time using different offset will help to profile DUT better.

Test report must contain value of T2 (same for all iterations) and values of e for each iterations. This document recommends to use $T2=T1+8s$ and e from 0 to 1s in 0.01s (10ms) increments.

4. TEST PROCEDURES

This section provides generic steps that are used in all tests.

4.1. Initialisation time

The objective of this test is to measure time that must elapse between starting protocols and ability of the test topology to forward traffic. This test is not intended to reflect DUT

performance but used only as a way to find time T_1 that is used in all subsequent tests.

To execute test perform following steps:

1. Configure DUT and protocol simulator (or auxiliary nodes)
2. At T_0 start traffic and then immediately start routing protocols
3. When traffic starts arriving Sink Port 1 stop test.

The time of arrival of the first packet is T_1 .

4.2. Generic data-plane failure test

The purpose of failure test is to measure time required by DUT to resume traffic flow after best path to destination fails. Following steps are common for all failure tests:

1. Start protocols and mark time as T_0
2. At time T_1 start traffic to each prefix in set NetA
3. At $T_2 + e$ simulate failure or restoration event (see Section 5)
4. From $T_2 + e$ until T_3 packets do not arrive to NetA
5. After packets are seen again at NetA (T_3) wait until time T_4
6. Stop traffic
7. Measure total number of lost packets and calculate outage knowing packet-per-second

4.3. Generic test procedure for

1. At T_0 bring up all interfaces and protocols, and start capturing BGP packets at RS1
2. At $T_1 + e$ simulate failure/restoration event (see Section 5)
3. At $T_2 - d_1$ first UPDATE message is sent by DUT and at T_2 it will be observed at RS1
4. At $T_3 - d_2$ last UPDATE message is sent by DUT and at T_3 it will be observed at RS1

d_1 and d_2 represent serialisation and propagation delay and can be

disregarded unless DUT-RS1 link has large delay. With this in mind, T2-(T1+e) and T3-(T1+e) represent convergence time for the first and last prefix respectively.

5. Failure and restoration scenarios

This section defines set of various failure and restoration scenarios used in step 3 of the generic test procedures described in previous section. Unless otherwise specified all scenarios are applicable to both data- and control-plane test procedures.

5.1. Loss of Signal on the link attached to DUT

This scenario simulates situation where link attached to DUT fails and Loss of Signal (LoS) can be observed by DUT. In other words link fails and results in interface on the DUT going down.

To simulate LoS failure at the time defined by the test procedure shut down R1 side of the link to DUT.

To simulate LoS restoration at the time defined by the test procedure re-activate R1 side of the link to DUT.

5.2. Link failure without LoS

This scenario simulates situation where link between DUT and adjacent node fails but DUT does not observe LoS. In practice such failure can occur when, for example, link between DUT and adjacent node is implemented via carrier equipment that does not shut link down when remote side of the link fails.

DUT can use various methods to detect such failures, including but not limited to protocol HELLO or Keep-alive packets, BFD, OAM. This document does not restrict methods which DUT can use, but requires use of particular method to be recorded in the test report.

Basic network topology is modified for the purpose of this test only as follow: rather than using direct cabling between DUT and R1 the link is implemented via intermediate L2 switch that supports concept of VLAN's. Initially switch ports connected to DUT and R1 are placed into the same VLAN (same L2 broadcast domain).

To simulate failure at the time defined by the test procedure move switch port connected to R1 to a VLAN different from the one used for switch port connected to DUT.

To simulate restoration at the time defined by the test procedure

move switch port connected to R1 back to the same VLAN as the one used for switch port connected to DUT.

5.3. Non-direct link failure

This scenario simulates situation where a link not directly connected to DUT but located on the primary path to destination fails. Unmodified basic network topology is used.

Depending on technologies used in the setup different failure detection techniques can be employed by DUT. This document assumes that DUT relies exclusively on IGP information to learn about failure and that nodes adjacent to the failed link flood this information within D seconds since the event. If required exact value of D can be obtained through simple additional test, but in this document D is assumed to be 0 (zero).

It is possible, though undesirable, that some traffic and protocol simulators may continue accepting packets coming through the port that leads to simulated failed link. It is essential to assert such behaviour prior to the tests and if confirmed, exclude packets received after failure from calculations in step 7 of the test.

Failure event is triggered by simulating shutdown of R3 side of the link to R1 at the time defined by the test procedure. R1 MUST send IGP update (depending on which protocol is used) to DUT within D seconds.

Restoration event is triggered by simulating recovery of R3 side of the link to R1 at the time defined by the test procedure. R1 MUST send IGP update (depending on which protocol is used) to DUT within D seconds.

5.4. Best route withdrawal

This scenario simulates situation where best AS exit path to a destination is no longer valid and ASBR sends BGP UPDATE to its iBGP peers. Unmodified basic network topology is used.

Disconnecting R3 from NetA implies that R3 will send BGP WITHDRAW for this prefixes in its update to DUT. It is possible, though undesirable, that some protocol simulator and traffic generators will still count packets received at sink port 1 even after prefixes were withdrawn. To correctly execute this test it's mandatory that traffic received at sink port 1 after withdrawing prefixes is ignored and not counted as delivered. If traffic generator is not able to assure such functionality (should be asserted prior to the test), then packets received at the sink port 1 MUST be excluded from

calculation in step 7 of the test.

Failure event is triggered by simulating failure of the link between R3 and NetA and immediate withdrawal of all corresponding prefixes by R3.

Restoration event is triggered by simulating recovery of the link between R3 and NetA and immediate BGP UPDATE for all corresponding prefixes by R3.

5.5. iBGP next-hop failure

This scenario simulates situation where ASBR used as best exit to a destination unexpectedly fails both at control and forwarding plane. Both R1 and a router within M1 connected to R3 MUST send appropriate IGP update message to the rest of the network within D seconds. To detect failure DUT MAY rely on IGP information provided by rest of the network or it MAY employ additional techniques. This document does not restrict what detection mechanism should DUT use but requires that particular mechanism is recorded in the test report.

Failure event is triggered by simulating removal of R3 from the test topology at the time defined by the test procedure, followed by IGP update as described in previous paragraph.

Recovery event is triggered by re-introducing R3 into the test topology, followed by IGP update as described in first paragraph of this section and immediate re-activation of BGP session between R3 and DUT. Note that recovery time calculated by this method depends on DUT performance in respect to bringing up new BGP session. This is intentional. Control plane convergence benchmarking can be performed separately by a method that is outside of the scope of this document and two results can be correlated netto data-plane convergence value should that be necessary.

6. Test report

TODO: Report format is to be discussed.

Test report MUST contain following data for each test:

1. T1 and 'e'
2. Number of prefixes NetA and NetB
3. Size of M1 (recored as N*K)

4. Traffic rate, in packets per second, and packet size at IP layer in octets
5. Number of lost packets during failure, and number of lost packets during restoration

7. Link bundling and Equal Cost Multi-Path

Scenarios where DUT can balance traffic to NetA across multiple best paths is explicitly excluded from scope of this document. There are two reasons.

First, two different DUT may choose different path (out of all equal) to forward given packet, which makes it unreasonably difficult to define generic traffic that would produce comparable results when testing different platforms.

Second, mechanisms used to handle failures in ECMP (but not necessarily in link-bundling) environment are similar to those handling single-path failures. Therefore it's expected that convergence in ECMP scenario will be of the same order as in single-path scenario.

8. Graceful Restart and Non-Stop Forwarding

While Graceful Restart and Non-Stop Forwarding mechanisms are related to DUT ability to forward traffic under certain failure conditions, the test covering DUT own ability to restore or preserve traffic flow already covered in RFC6201.

9. Security considerations

The tests described in this document intended to be performed in isolated lab environment, which inherently has no security implication on the live network of the organisation or Internet as whole.

Authors foresee that some people or organisations might be interested to benchmark performance of the live networks. The tests described in this document are disruptive by their nature and will have impact at least on the network where they're executed, and depending on the role of that network effect can extend to other parts of the Internet. Such tests MUST NOT be attempted in live environment without careful consideration.

The fact of publishing this document does not increase potential negative consequences if tests are executed in live environment because information provided here is mere recording of widely known and used techniques.

10. IANA Considerations

None.

11. Acknowledgments

Authors would like to thank Gregory Cauchie, Rob Shakir, David Freedman, Anton Elita, Saku Ytti, Andrew Yourtchenko, for their valuable contribution and peer-review of this work.

12. Normative References

[RFC4760] Bates, T., Chandra, R., Katz, D., and Y. Rekhter,
"Multiprotocol Extensions for BGP-4", RFC 4760,
January 2007.

Authors' Addresses

Ilya Varlashkin
Easynet Global Services

Email: ilya.varlashkin@easynet.com

Rajiv Papneja
Huawei Technologies (USA)

Email: rajiv.papneja@huawei.com

Bhavani Parise
Cisco

Email: bhavani@cisco.com

Tara Van Unen
Ixia

Email: TVanUnen@ixiacom.com

