Benchmarking Working Group
Internet-Draft
Intended status: Standards Track
Expires: December 27, 2014

R. Papneja
Huawei Technologies
B. Parise
Cisco Systems
S. Hares
Adara Networks
D. Lee
IXIA
I. Varlashkin
Easynet Global Services

June 25, 2014

Basic BGP Convergence Benchmarking Methodology for Data Plane Convergence
draft-ietf-bmwg-bgp-basic-convergence-02.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 December 27, 2014.

Table of Contents

1. Introduction

This document defines the methodology for benchmarking data plane FIB convergence performance of BGP in routers and switches using 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]. Similarly both iBGP & eBGP are covered in the tests as applicable.

The scope of this document is to provide methodology for BGP protocol FIB convergence measurements with BGP functionality limited to IPv4 & IPv6 as defined in RFC 4271 and Multi-Protocol BGP (MP-BGP) [RFC4760, RFC2545]. Other BGP extensions to support layer-2, layer-3 virtual private networks (VPN) are outside the scope of this document. Interaction with IGPs (IGP interworking) is outside the scope of this document.

1.1. Benchmarking Definitions

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

FIB (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.

This document defines methodology to test

- Data plane convergence on a single BGP device that supports the BGP functionality with scope as outlined above

- using test topology of 3 or 4 nodes which are sufficient to recreate the Convergence events used in the various tests of this draft

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 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 from different vendors.

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

These test setups 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

- AS-X: Autonomous System X
- Loopback Int: Loopback interface on the BGP enabled device
- HLP,HLP1,HLP2: Helper routers running the same version of BGP as DUT
- Enable NTP or use any external clock source to synchronize to the nodes

3.1. General Reference Topologies

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

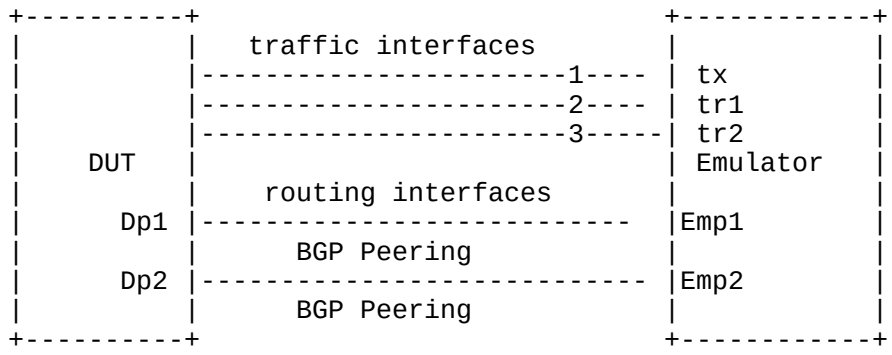

Figure 1 Basic Test Setup

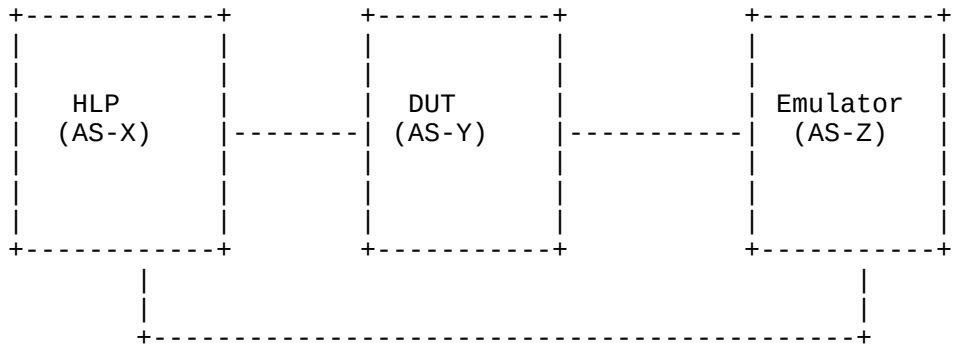

Figure 2 Three Node Setup for eBGP and iBGP Convergence

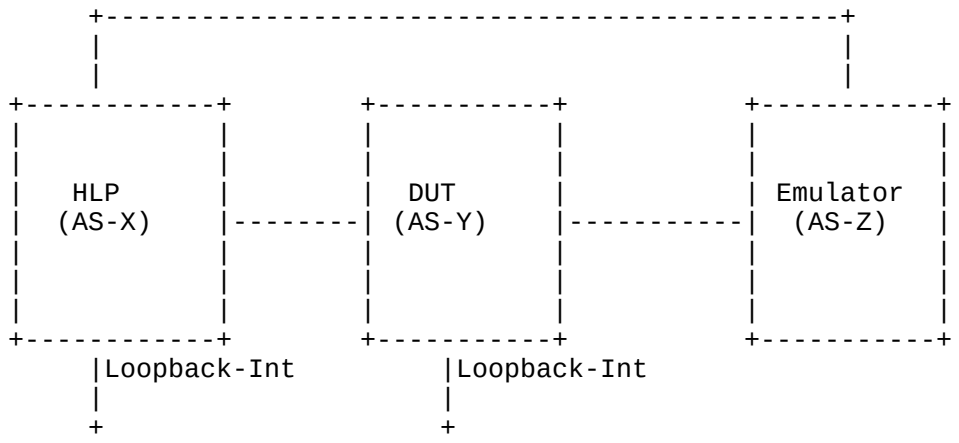

Figure 3 BGP Convergence for eBGP Multihop Scenario

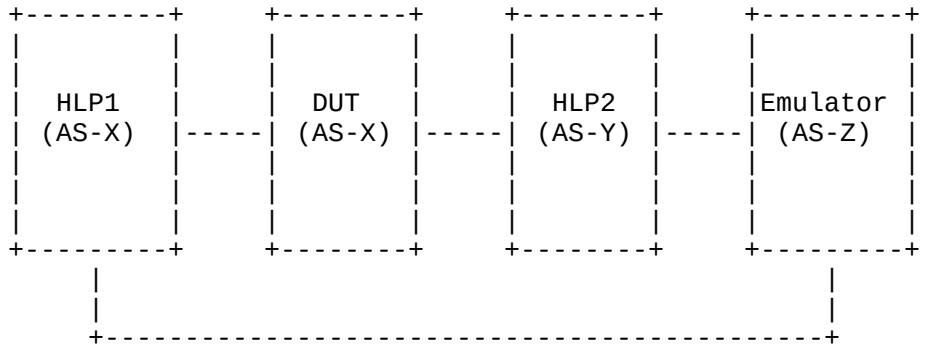

Figure 4 Four Node Setup for EBGP and IBGP Convergence

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 test case (as specified in the test).

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 advertised or learnt by the DUT per session or through a 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 consider advertising 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 MUST 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 all iterations of 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 as 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 MUST 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. For cases where the redundancy cannot be disabled, the results are no longer comparable and the level of impacts on the measurements is out of scope of this document.

5. Test Cases

All tests defined under this section assume the following:

a. BGP peers are in established state

b. BGP state should be cleared from established state to idle prior to each test. This is recommended to ensure that all tests start with the BGP peers being forced back to idle state and databases flushed.

c. Furthermore the traffic generation and routing should be verified in the topology to ensure there is no packet loss observed on any advertised routes

d. The arrival timestamp of advertised routes can be measured by installing an inline monitoring device between the emulator and DUT, or by the span port of DUT connected with an external analyzer. The time base of such inline monitor or external analyzer needs to be synchronized with the protocol and traffic emulator. Some modern emulator may have the capability to capture and timestamp every NLRI packets leaving and arriving at the emulator ports. The timestamps of these NLRI packets will be almost identical to the arrival time at DUT if the cable distance between the emulator and DUT is relatively short.

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 one peer of Emulator, Emp1.

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 tx towards the DUT targeted at a routes specified in route mixture (ex. routeA) Initially no traffic SHOULD be observed on the egress interface as the routeA is not installed in the forwarding database of the DUT.

E. Advertise routeA from the peer(Emp1) to the DUT and record the time.

   This is Tup(EMp1,Rt-A) also named 'XMT-Rt-time(Rt-A)'.

F. Record the time when the routeA from Emp1 is received at the DUT.

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

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

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

H. The difference between the Tup(DUT,RT-A) and traffic received time (TR (TDr, Rt-A) is the FIB Convergence Time for routeA in the route mixture. A full convergence for the route update is the measurement between the 1st route (Rt-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) MUST 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 (e.g. routeA). Initially no traffic SHOULD be observed on the egress interface as the routeA 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, Rt-A), also named EM-XMT-Data-Time(Rt-A)

I. Record when routeA is received by DUT.

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

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

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

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

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

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

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(first)-DUT-XMT-Data-Time(last)

RIB-convergence =DUT-RCV-Rt-Time(first)-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) MUST run the same version of BGP as DUT.

B. All devices MUST be synchronized using NTP or some local reference clock.

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 (e.g. 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(Rt-A).

L. Record the time when the route is received by the DUT. This is Tup(EMr,DUT) named Route-Rcv-time(Rt-A).

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(Rt-A)

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(Rt-A)

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

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

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, routeA over both the adjacencies and (Emp1) 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 (e.g. routeA). Initially traffic would be observed on the best egress route (Emp1) instead of Emp2.

F. Trigger the shutdown event of Best Egress Interface on DUT (Dp1).

G. Measure the Convergence Time for the event to be detected and traffic to be forwarded to Next-Best Egress Interface (Dp2)

   Time = Data-detect(Emp2) - 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 (Dp2) to Best Interface (Dp1)

   Time = Data-detect(Emp1) - 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 (e.g. 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

Convergence = TR2-traffic-on - 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 (e.g 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 MUST be synchronized using NTP or some local reference clock.

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.

Route Refresh Convergence Time = (RecTime - 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 MUST be synchronized using NTP or some local reference clock.

(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 (e.g. 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(Rt-A).

(10) Record the time when no traffic is observed on the Egress Interface. This is the RouteRemoveTime1(Rt-A).

(11) The difference between the RouteRemoveTime1 and WdrawTime1 is the WdrawConvTime1

WdrawConvTime1(Rt-A) = RouteRemoveTime1(Rt-A) - WdrawTime1(Rt-A)

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 (This assumes 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 (e.g. 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(Rt-A)

WAWtime-TRx(Rt-A) = WdrawTime2(Rt-A)

(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(Rt-A)

(7) Total withdraw processing time is

TotalWdrawTime(Rt-A) = ((RouteRemoveTime2(Rt-A) - WdrawTime2(Rt-A)) - WdrawConvTime1(Rt-A))

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 MUST be synchronized using NTP or some local reference clock.

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 (e.g. 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(Path-Change(Rt-A)) = Path Change Event Time(Rt-A)

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

DUT(Path-Change, Rt-A) = Path-switch time(Rt-A)

Convergence = Path-switch time(Rt-A) - Path Change Event Time(Rt-A)

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 as detailed in the Terminology document [RFC4098].

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 MUST be synchronized using NTP or some local reference clock.

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 (e.g. 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(Rt-A)

R. The following equation represents the Graceful Restart Convergence Time

   Graceful Restart Convergence Time(Rt-A) = ((RecTime(Rt-A) - GREventTime) - 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. Acknowledgements

We would like to thank Anil Tandon, Arvind Pandey, Mohan Nanduri, Jay Karthik, Eric Brendel for their input and discussions on various sections in the document. We also like to acknowledge Will Liu, Semion Lisyansky, Faisal Shah for their review and feedback to the document.

10. References

10.1. Normative References

[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.

[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.

[RFC4271] Rekhter, Y., Li, T., and S. Hares, "A Border Gateway Protocol 4 (BGP-4)", RFC 4271, January 2006.

[RFC6412] Poretsky, S., Imhoff, B., and K. Michielsen, "Terminology for Benchmarking Link-State IGP Data-Plane Route Convergence", RFC 6412, November 2011.

10.2. 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.

[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
Adara Networks

Email: shares@ndzh.com

Dean Lee
IXIA

Email: dlee@ixiacom.com

Ilya Varlashkin
Easynet Global Services

Email: ilya.varlashkin@easynet.com