BGP Data-Plane Benchmarking Applicable to Modern Routers

Rajiv Papneja

Ilya Varlashkin

Bhavani Parise

Dean Lee

Sue Hares

Background/ Current Status

- Terminology for Benchmarking BGP Device Convergence in the (RFC 4098) / June 2005
- Two complementing BGP Methodology drafts in progress
- Draft-1 focusing on the BGP Data Plane convergence was posted prior IETF 80
  - draft-papneja-bgp-basic-dp-convergence-03
    - Author team: Rajiv Papneja, Bhavani Parise, Sue Hares, Ilya Varlashkin
    - Contributors: Deal Lee, Eric Brendal, Mohan Nanduri, Jay Karthik
- Draft-2 focuses on Data-plane convergence on modern routers, which have local restoration capability was posted prior to IETF 82
  - draft-varlashkin-router-conv-bench-00
    - Author Team: Ilya Varlashkin, Rajiv Papneja, Bhavani Parise, Dean Lee
    - Reviewer: Tara Van Unen

Current Status & Action Items from IETF82

- Obtained feedback from multiple Service Providers. Various tests from the draft termed as ‘Valuable’ and ‘Important’ by the Providers
- Performed Proof of concept benchmarking tests based on the methodology from the draft. Results in later slides
  - Results very encouraging
- Posted revised draft addressing comments from IETF82 and comments from WG list
  - http://tools.ietf.org/html/draft-papneja-bgp-basic-dp-convergence-03

Current Status & Action Items from IETF82

- Soliciting feedback from IDR
- Current draft includes basic convergence for IPv4 & IPv6.
- Other Address Families, RRs need to be covered in same or new draft?

Convergence Tests Results Using Methodologies Proposed

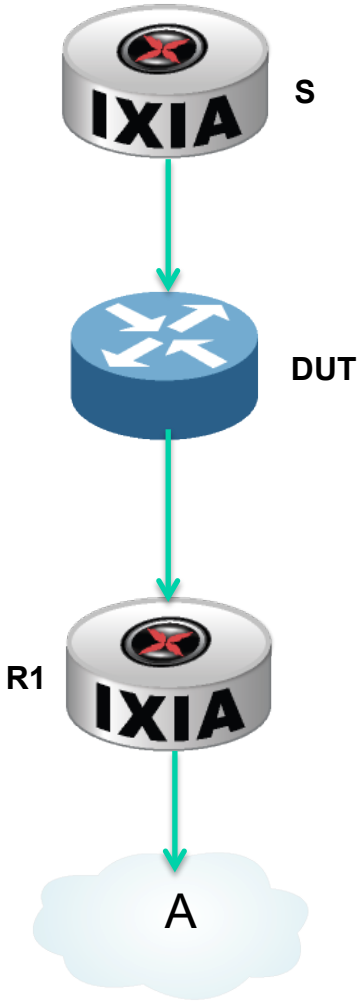
(draft-papneja-bgp-basic-dp-convergence-03)

5.1 RIB-IN Convergence

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

Test steps –

- S sends traffic toward route A
- R1 advertises route A at RCV-Rt-time
- R1 receives traffic for route A at DUT-XMT-Data-Time
- Route Update Convergence time = DUT-XMT-Data-Time - RCV-Rt-time



5.1 Test Config

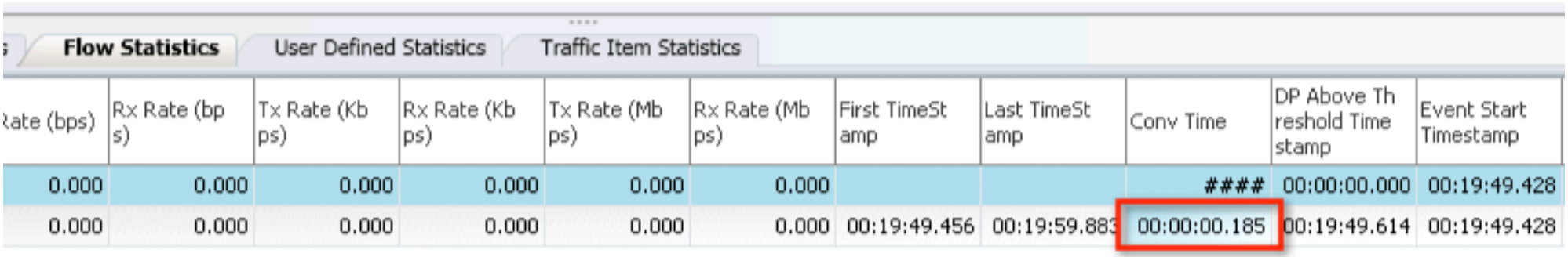
DUT BGP config

- AS 100
- Hold timer 180 sec
- Keep alive 60 sec
- Connect retry 1 sec
- No MAOI
- No MRAI
- Damping off
- MD5 off

- Tester config
  - 2 ports emulating S, R1
  - R1 with AS 202, IP 50.0.1.2
  - 1000 /32 routes, unique next hop per route

5.1 Test Result

Iteration	Convergence Time (ms)
1	179
2	183
3	183
4	185
Average	182.5
Deviation	1.75



Flow Statistics | User Defined Statistics | Traffic Item Statistics

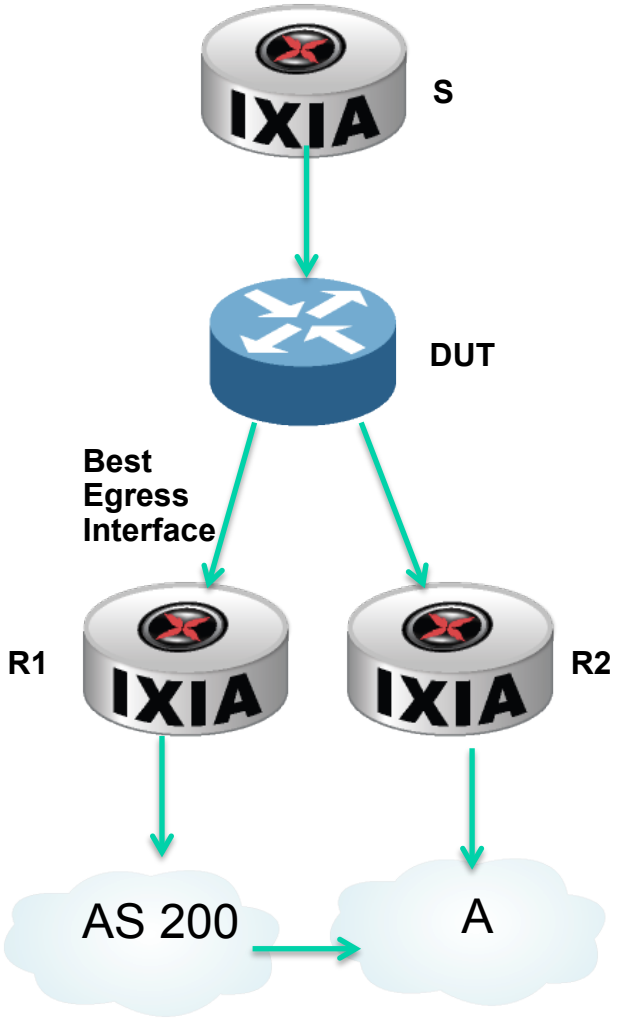
Rate (bps)	Rx Rate (bps)	Tx Rate (Kbps)	Rx Rate (Kbps)	Tx Rate (Mbps)	Rx Rate (Mbps)	First TimeStamp	Last TimeStamp	Conv Time	DP Above Threshold Time stamp	Event Start Timestamp
0.000	0.000	0.000	0.000	0.000	0.000			####	00:00:00.000	00:19:49.428
0.000	0.000	0.000	0.000	0.000	0.000	00:19:49.456	00:19:59.883	00:00:00.185	00:19:49.614	00:19:49.428

BGP Failure - 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

Test steps –

- R1 advertises the Route A. R1 is the optimal path
- S sends traffic toward route A
- R2 advertises the same Route A with optimal AS-Path at “Path-Change-Event-Time”
- R2 receives traffic at “Path-Switch-Time”
- BGP convergence time = “Path-Switch-Time” - “Path-Change-Event-Time”



5.7 Test Config

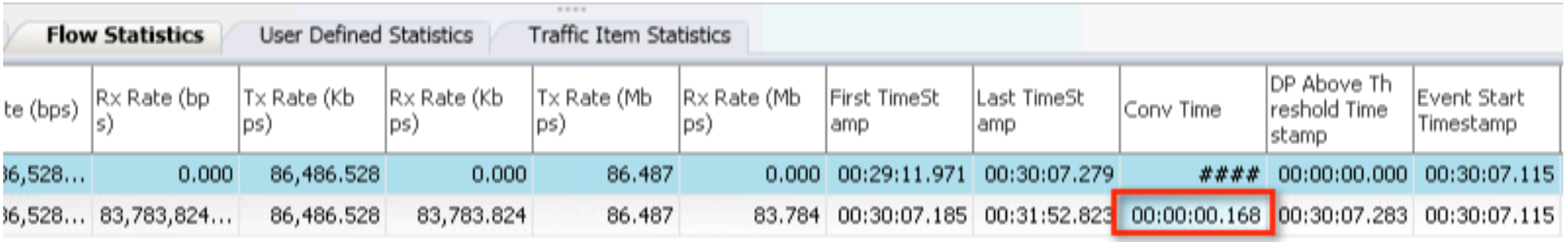
DUT BGP config

- AS 100
- Hold timer 180 sec
- Keep alive 60 sec
- Connect retry 1 sec
- No MAOI
- No MRAI
- Damping off
- MD5 off

- Tester config
  - 3 ports emulating S, R1 and R2
  - R1 with AS 202, IP 50.0.1.2, 1000 /32 routes with the same next hop and AS200
  - R2 with AS 102, IP 40.0.1.2, 1000 /32 routes with unique next hop per route

5.7 Test Result

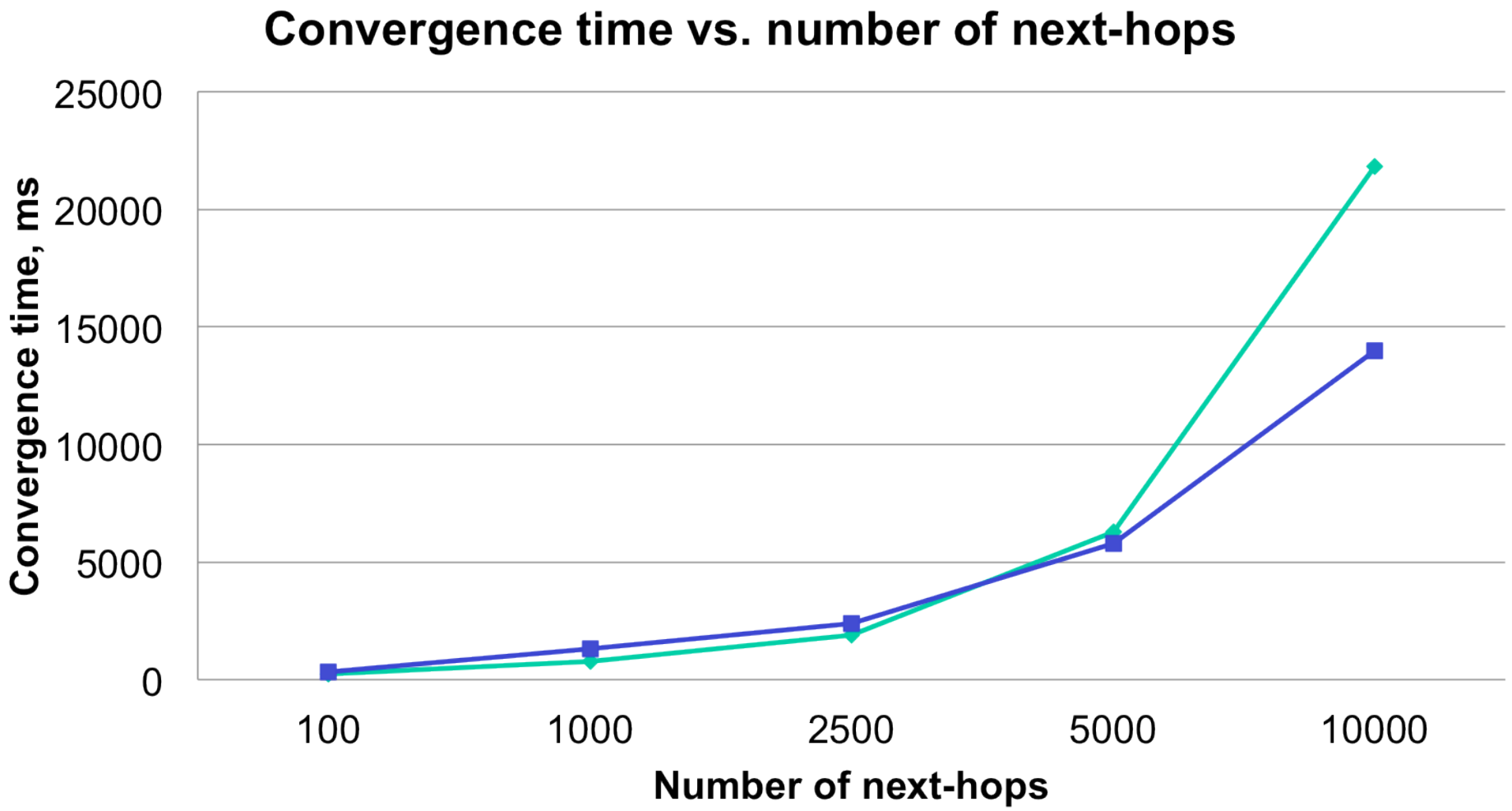
Iteration	Convergence Time (ms)
1	169
2	168
3	161
4	167
Average	167
Deviation	2.63



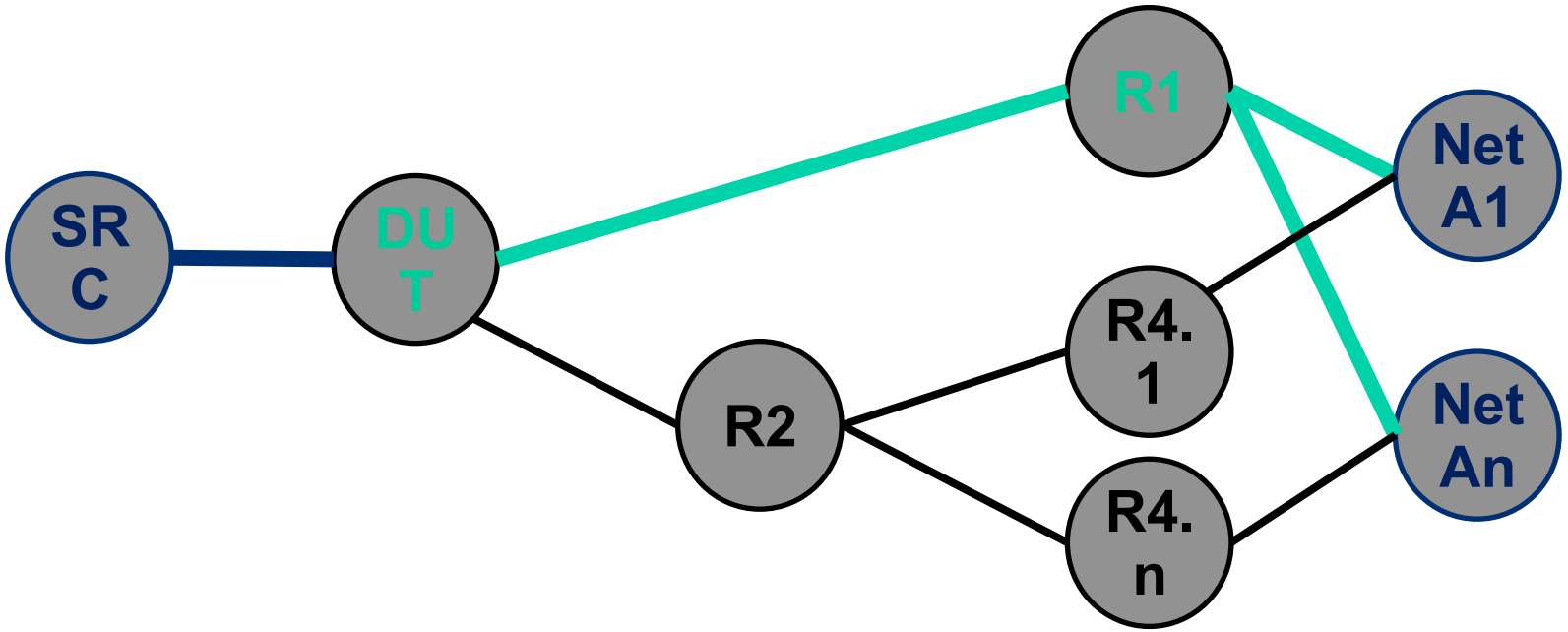
Flow Statistics | User Defined Statistics | Traffic Item Statistics

te (bps)	Rx Rate (bps)	Tx Rate (Kbps)	Rx Rate (Kbps)	Tx Rate (Mbps)	Rx Rate (Mbps)	First TimeStamp	Last TimeStamp	Conv Time	DP Above Threshold Timestamp	Event Start Timestamp
86,528...	0.000	86,486.528	0.000	86.487	0.000	00:29:11.971	00:30:07.279	####	00:00:00.000	00:30:07.115
86,528...	83,783,824...	86,486.528	83,783.824	86.487	83.784	00:30:07.185	00:31:52.823	00:00:00.168	00:30:07.283	00:30:07.115

Sample results for test 5.1



Modified test topology



Action Items

- Posted revised:
  - http://tools.ietf.org/html/draft-papneja-bgp-basic-dp-convergence-03
- Posted new draft:
  - http://tools.ietf.org/html/draft-varlashkin-router-conv-bench-01
- Can we agree on approach?
- Agree on WG-item readiness
- Inputs welcome!

Backup

Summary

(draft-papneja-bgp-basic-dp-convergence-02)

- BGP Data plane FIB convergence for both IPv4 and IPv6
- Limited to Basic BGP convergence (RFC 4271 functionality with Multi-Protocol BGP (MP-BGP) for IPv6)
- BGP Failure/Convergence Events
- Considers dependencies on factors impacting convergence:
  - Number of peers,
  - Number of routes/peers
  - Policy Processing/Reconfiguration
- Data Traffic characterization – offered load
- Various test cases that covers iBGP, eBGP and failure convergence events
- Topologies – Several 3 node, and 4 node setups

Summary
(draft-varlashkin-router-conv-bench-00)

- Metrics derived from packets, not from clock
- Single test procedure, 5 failure scenarios
  - Loss of Signal on the link attached to DUT
  - Attached link failure without LoS
  - Non-direct link failure
  - Best route withdrawal
  - BGP next-hop failure
- Background topological noise for realistic result
- Large topology scenario

Complimenting Drafts at a Glance

draft-papneja-bgp-basic-dp-convergence-02	draft-varlashkin-router-conv-bench-00
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.	This document specifies methodology for benchmarking convergence of routers without making assumptions about relation and dependencies between data- and control-planes.
Methodologies applicable in a BGP enabled setup. For IGP setup readers are encouraged to refer IGP convergence work	Provided methodology is primary intended for testing routers running BGP and some form of link-state IGP with or without MPLS
Tests discussed: RIB-IN Convergence, RIB-OUT Convergence,. eBGP Convergence, iBGP Convergence, eBGP multihop convergence	Initialization time, generic data-plane failure test
Convergence Events: Physical link failure on DUT/Remote end, ECMP Link failure on DUT end, BGP session failure, BGP hard reset, BGP soft reset, BGP route withdrawal, BGP Path attribute change, BGP Graceful restart	Convergence Events: LOS of signal, link failure without LOS, Non-direct link failure, Best route withdrawal and iBGP next hop failure