Network Working Group
Internet Draft

Vishwas Manral
Netplane Systems
Russ White
Cisco Systems
Aman Shaikh
University of California

Expiration Date: December 2004
File Name: draft-bmwg-ospfconv-intraarea-09.txt

June 2004

Benchmarking Basic OSPF Single Router Control Plane Convergence
draft-ietf-bmwg-ospfconv-intraarea-09.txt

Status of this Memo

This document is an Internet-Draft and is in full conformance with all provisions of Section 10 of RFC2026.

Internet Drafts are working documents of the Internet Engineering Task Force (IETF), its Areas, and its Working Groups. Note that other groups may also distribute working documents as Internet Drafts.

Internet Drafts are draft documents valid for a maximum of six months. Internet Drafts may be updated, replaced, or obsoleted by other documents at any time. It is not appropriate to use Internet Drafts as reference material or to cite them other than as a "working draft" or "work in progress".

The list of current Internet-Drafts can be accessed at http://www.ietf.org/ietf/1id-abstracts.txt

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

This document may not be modified, and derivative works of it may not be created, except to publish it as an RFC and to translate it into languages other than English.

Abstract

This draft provides suggestions for measuring OSPF single router control plane convergence. Its initial emphasis is on the control plane of single OSPF routers. We do not address forwarding plane performance.

NOTE: Within this document, the word convergence relates to single router control plane convergence only.

Table of Contents

1. Introduction.....1
2. Specification of Requirements.....2
3. Overview & Scope.....2
4. Reference Topologies.....3
5. Basic Performance Tests.....4
5.1 Time Required to Process and LSA.....4
5.2 Flooding Time.....5
5.3 Shortest Path First Computation Time.....5
6. Basic Intra-Area OSPF Tests.....7
6.1 Forming Adjacencies on Point-to-Point Links (Initialization)....8
6.2 Forming Adjacencies on Point-to-Point Links.....8
6.3 Forming Adjacencies with Information Already in the Database....9
6.4 Designated Router Election Time on a Broadcast Network.....10
6.5 Initial Convergence Time on a Broadcast Network, Test 1.....11
6.6 Initial Convergence Time on a Broadcast Network, Test 2.....11
6.7 Link Down with Layer Two Detection.....12
6.8 Link Down with Layer Three Detection.....12
7. IANA Considerations.....13
8. Security Considerations.....13
9. Acknowledgements.....13
10. Normative References.....13
11. Informative References.....14
12. Author's Addresses.....14
13. Full Copyright Statement.....15
14. Intellectual Property.....15

1. Introduction

There is a growing interest in routing protocol convergence testing, with many people looking at various tests to determine how long it takes for a network to converge after various conditions occur. The major problem with this sort of testing is that the framework of the tests has a major impact on the results; for instance, determining when a network is converged, what parts of the router's operation are considered within the testing, and other such things will have a major impact on what apparent performance routing protocols provide.

This document attempts to provide a framework within which Open Shortest Path First [OSPF] performance testing can be placed, and provide some tests with which some aspects of OSPF performance can be measured. The motivation of the draft is to provide a set of tests that can provide the user comparable data from various vendors with which to evaluate the OSPF protocol performance on the devices.

2. Specification of 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 [RFC2119]. RFC2119 keywords in this document are used to assure methodological control, which is very important in the specification of benchmarks. This document does not specify a network related protocol.

3. Overview & Scope

While this document describes a specific set of tests aimed at characterizing the single router control plane convergence performance of OSPF processes in routers or other boxes that incorporate OSPF functionality, a key objective is to propose methodologies that will produce directly comparable convergence related measurements.

Things which are outside the scope of this document include:

- The interactions of convergence and forwarding; testing is restricted to events occurring within the control plane. Forwarding performance is the primary focus in [INTERCONNECT] and it is expected to be dealt with in work that ensues from [FIB-TERM].

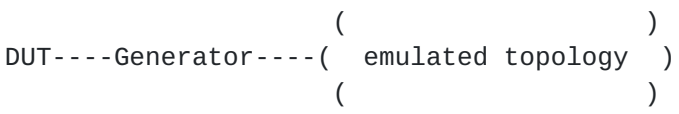
- Inter-area route generation, AS-external route generation, and simultaneous traffic on the control and data paths within the DUT. While the tests outlined in this document measure SPF time, flooding times, and other aspects of all OSPF convergence performance, it does not provide tests for measuring external or summary route generation, route translation, or other OSPF inter-area and external routing performance. These are expected to be dealt with in a later draft.

  Tests should be run more than once, since a single test run cannot be relied on to produce statistically sound results. The number of test runs and any variations between the tests should be recorded in the test results (see [TERM] for more information on what items should be recorded in the test results).

4. Reference Topologies

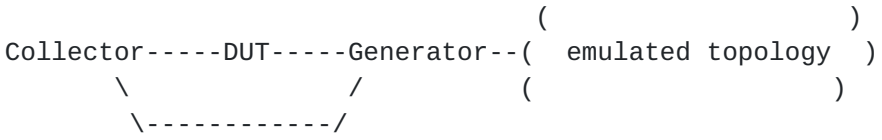
Several reference topologies will be used throughout the tests described in the remainder of this document. Rather than repeating these topologies, we've gathered them all in one section.

o Reference Topology 1 (Emulated Topology)



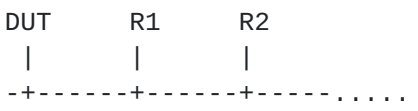
A simple back-to-back configuration. It's assumed that the link between the generator and the DUT is a point-to-point link, while the connections within the generator represent some emulated topology.

o Reference Topology 2 (Generator and Collector)



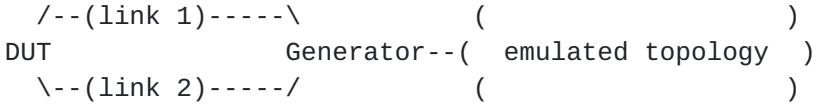
All routers are connected through point-to-point links. The cost of all links is assumed to be the same unless otherwise noted.

o Reference Topology 3 (Broadcast Network)



Any number of routers could be included on the common broadcast network.

o Reference Topology 4 (Parallel Links)



In all cases the tests and topologies are designed to allow performance measurements to be taken all on a single device, whether the DUT or some other device in the network. This eliminates the need for synchronized clocks within the test networks.

5. Basic Process Performance Tests

These tests will measure aspects of the OSPF implementation as a process on the device under test, including:

- Time required to process an LSA
- Flooding time
- Shortest Path First computation

5.1. Time required to process an LSA

- Using reference topology 1 (Emulated Topology), begin with all links up and a full adjacency established between the DUT and the generator.

  Note: The generator does not have direct knowledge of the state of the adjacency on the DUT. The fact the adjacency may be in Full on the generator does not mean that the DUT is ready. It may still (and is likely to) be requesting LSAs from the generator. This process, involving processing of requested LSAs, will affect the results of the test. The generator should either wait until it sees the DUT's router-LSA listing the adjacency with the generator or introduce a configurable delay before starting the test.

- Send an LSA that is already there in the DUT (a duplicate LSA), note the time difference between when the LSA is sent to when the ack is received. This measures the time to propagate the LSA and the ack, as well as processing time of the duplicate LSA. This is dupLSAprocTime.

- Send a new LSA from the generator to the DUT, followed immediately by a duplicate LSA (LSA that already resides in the database of DUT, but not the same as the one just sent).

- The DUT will acknowledge this second LSA immediately; note the time of this acknowledgement. This is newLSAprocTime.

  The amount of time required for an OSPF implementation to process the new LSA can be computed by subtracting dupLSAprocTime from newLSAprocTime.

  Note: The duplicate LSA cannot be the same as the one just sent

because of the MinLSInterval restriction.[RFC2328] This test is taken from [BLACKBOX].

Note: This time may or may not include the time required to perform flooding-related operations, depending on when the implementation sends the ack--before it floods the LSA further or after, or anywhere in between. In other words, this measurement may not mean the same thing in all implementations.

5.2. Flooding Time

- Using reference topology 2 (Generator and Collector), enable OSPF on all links and allow the devices to build full adjacencies. Configure the collector so it will block all flooding towards the DUT, although it continues receiving advertisements from the DUT.
- Inject a new set of LSAs from the generator towards the collector and the DUT.
- On the collector, note the time the flooding is complete across the link to the generator. Also note the time the flooding is complete across the link from the DUT.

The time between the last LSA is received on the collector from the generator and the time the last LSA is received on the collector from the DUT should be measured during this test. This time is important in link state protocols, since the loop free nature of the network is reliant on the speed at which revised topology information is flooded.

Depending on the number of LSAs flooded, the sizes of the LSAs, the number of LSUs, and the rate of flooding, these numbers could vary by some amount. The settings and variances of these numbers should be reported with the test results.

5.3. Shortest Path First Computation Time

- Use reference topology 1 (Emulated Topology), beginning with the DUT and the generator fully adjacent.
- The default SPF timer on the DUT should be set to 0, so that any new LSA that arrives, immediately results in the SPF calculation [BLACKBOX].

- The generator should inject a set of LSAs towards the DUT; the DUT should be allowed to converge and install all best paths in the local routing table, etc..

- Send an LSA that is already there in the DUT (a duplicate LSA), note the time difference between when the LSA is sent to when the ack is received. This measures the time to propagate the LSA and the ack, as well as processing time of the duplicate LSA. This is dupLSAprocTime.

- Change the link cost between the generator and the emulated network it is advertising, and transmit the new LSA to the DUT.

- Immediately inject another LSA which is a duplicate of some other LSA the generator has previously injected (preferably a stub network someplace within the emulated network).

  Note: The generator should make sure that outbound LSA packing is not performed for the duplicate LSAs and they are always sent in a separate Link-state Update packet. Otherwise, if the LSA carrying the topology change and the duplicate LSA are in the same packet, the SPF will be started the duplicate LSA is acked.

- Measure the time between transmitting the second (duplicate) LSA and the acknowledgement for that LSA; this is the totalSPFtime. The total time required to run SPF can be computed by subtracting dupLSAprocTime from totalSPFtime.

The accuracy of this test is crucially dependant on the amount of time between the transmission of the first and second LSAs. If there is too much time between them, the test is meaningless because the SPF run will complete before the second (duplicate) LSA is received. If there is too little time between the LSAs being generated, then they will both be handled before the SPF run is scheduled and started, and thus the measurement would only be for the handling of the duplicate LSA.

This test is also specified in [BLACKBOX].

Note: This test may not be accurate on systems which implement OSPF as a multithreaded process, where the flooding takes place in a separate process (or on a different processor) than shortest path first computations.

It is also possible to measure the SPF time using white box tests (using output supplied by the OSPF software implementer). For instance:

- Using reference topology 1 (Emulated Topology), establish a full adjacency between the generator and the DUT.
- Inject a set of LSAs from the generator towards the DUT. Allow the DUT to stabilize and install all best paths in the routing table, etc.
- Change the link cost between the DUT and the generator (or the link between the generator and the emulated network it is advertising), such that a full SPF is required to run, although only one piece of information is changed.
- Measure the amount of time required for the DUT to compute new shortest path tree as a result of the topology changes injected by the generator. These measurements should be taken using available show and debug information on the DUT.

Several caveats MUST be mentioned when using a white box method of measuring SPF time; for instance, such white box tests are only applicable when testing various versions or variations within a single implementation of the OSPF protocol. Further, the same set of commands MUST be used in each iteration of such a test, to ensure consistent results.

There is some interesting relationship between the SPF times reported by white box (internal) testing, and black box (external) testing; these two types of tests may be used as a "sanity check" on the other type of tests, by comparing the results of the two tests.

See [CONSIDERATIONS] for further discussion.

6. Basic Intra-Area OSPF tests

These tests measure the performance of an OSPF implementation for basic intra-area tasks, including:

- Forming Adjacencies on Point-to-Point Link (Initialization)
- Forming Adjacencies on Point-to-Point Links
- Link Up with Information Already in the Database
- Initial convergence Time on a Designated Router Electing (Broadcast) Network
- Link Down with Layer 2 Detection

- Link Down with Layer 3 Detection
- Designated Router Election Time on A Broadcast Network

6.1. Forming Adjacencies on Point-to-Point Link (Initialization)

This test measures the time required to form an OSPF adjacency from the time a layer two (data link) connection is formed between two devices running OSPF.

- Use reference topology 1 (Emulated Topology), beginning with the link between the generator and DUT disabled on the DUT. OSPF should be configured and operating on both devices.
- Inject a set of LSAs from the generator towards the DUT.
- Bring the link up at the DUT, noting the time that the link carrier is established on the generator.
- Note the time the acknowledgement for the last LSA transmitted from the DUT is received on the generator.

The time between the carrier establishment and the acknowledgement for the last LSA transmitted by the generator should be taken as the total amount of time required for the OSPF process on the DUT to react to a link up event with the set of LSAs injected, including the time required for the operating system to notify the OSPF process about the link up, etc.. The acknowledgement for the last LSA transmitted is used instead of the last acknowledgement received in order to prevent timing skews due to retransmitted acknowledgements or LSAs.

6.2. Forming Adjacencies on Point-to-Point Links

This test measures the time required to form an adjacency from the time the first communication occurs between two devices running OSPF.

- Using reference topology 1 (Emulated Topology), configure the DUT and the generator so traffic can be passed along the link between them.
- Configure the generator so OSPF is running on the point-to-point link towards the DUT, and inject a set of LSAs.

- Configure the DUT so OSPF is initialized, but not running on the point-to-point link between the DUT and the generator.
- Enable OSPF on the interface between the DUT and the generator on the DUT.
- Note the time of the first hello received from the DUT on the generator.
- Note the time of the acknowledgement from the DUT for the last LSA transmitted on the generator.

The time between the first hello received and the acknowledgement for the last LSA transmitted by the generator should be taken as the total amount of time required for the OSPF process on the DUT to build a FULL neighbor adjacency with the set of LSAs injected. The acknowledgement for the last LSA transmitted is used instead of the last acknowledgement received in order to prevent timing skews due to retransmitted acknowledgements or LSAs.

6.3. Forming adjacencies with Information Already in the Database

- Using reference topology 2 (Generator and Collector), configure all three devices to run OSPF.
- Configure the DUT so the link between the DUT and the generator is disabled .
- Inject a set of LSAs into the network from the generator; the DUT should receive these LSAs through normal flooding from the collector.
- Enable the link between the DUT and the generator.
- Note the time of the first hello received from the DUT on the generator.
- Note the time of the last DBD received on the generator.
- Note the time of the acknowledgement from the DUT for the last LSA transmitted on the generator.

The time between the hello received from the DUT by the generator and the acknowledgement for the last LSA transmitted by the generator should be taken as the total amount of time required for the OSPF process on the DUT to build a FULL neighbor adjacency with the set of

LSAs injected. In this test, the DUT is already aware of the entire network topology, so the time required should only include the processing of DBDs exchanged when in EXCHANGE state, the time to build a new router LSA containing the new connection information, and the time required to flood and acknowledge this new router LSA.

The acknowledgement for the last LSA transmitted is used instead of the last acknowledgement received in order to prevent timing skews due to retransmitted acknowledgements or LSAs.

6.4. Designated Router Election Time on A Broadcast Network

- Using reference topology 3 (Broadcast Network), configure R1 to be the designated router on the link, and the DUT to be the backup designated router.
- Enable OSPF on the common broadcast link on all the routers in the test bed.
- Disable the broadcast link on R1.
- Note the time of the last hello received from R1 on R2.
- Note the time of the first network LSA generated by the DUT as received on R2.

The time between the last hello received on R2 and the first network LSA generated by the DUT should be taken as the amount of time required for the DUT to complete a designated router election computation. Note this test includes the dead interval timer at the DUT, so this time may be factored out, or the hello and dead intervals reduced to make these timers impact the overall test times less. All changed timers, the number of routers connected to the link, and other variable factors should be noted in the test results.

Note: If R1 sends a "goodbye hello," typically a hello with its neighbor list empty, in the process of shutting down its interface, using the time this hello is received instead of the time of the last hello received would provide a more accurate measurement.

6.5. Initial Convergence Time on a Broadcast Network, Test 1

- Using reference topology 3 (Broadcast Network), begin with the DUT connected to the network with OSPF enabled. OSPF should be enabled on R1, but the broadcast link should be disabled.
- Enable the broadcast link between R1 and the DUT. Note the time of the first hello received by R1.
- Note the time the first network LSA is flooded by the DUT at R1.
- The differential between the first hello and the first network LSA is the time required by the DUT to converge on this new topology.

This test assumes that the DUT will be the designated router on the broadcast link. A similar test could be designed to test the convergence time when the DUT is not the designated router as well.

This test may be performed with varying numbers of devices attached to the broadcast network, and varying sets of LSAs being advertised to the DUT from the routers attached to the broadcast network. Variations in the LSA sets and other factors should be noted in the test results.

The time required to elect a designated router, as measured in Designated Router Election Time on A Broadcast Network, above, may be subtracted from the results of this test to provide just the convergence time across a broadcast network.

Note all the other tests in the document include route calculation time in the convergence time, as described in [TERM], this test may not include route calculation time in the resulting measured convergence time, because initial route calculation may occur after the first network LSA is flooded.

6.6. Initial Convergence Time on a Broadcast Network, Test 2

- Using reference topology 3 (Broadcast Network), begin with the DUT connected to the network with OSPF enabled. OSPF should be enabled on R1, but the broadcast link should be disabled.
- Enable the broadcast link between R1 and the DUT. Note the time of the first hello transmitted by the DUT with a designated router listed.

- Note the time the first network LSA is flooded by the DUT at R1.

- The differential between the first hello with a designated router lists and the first network LSA is the time required by the DUT to converge on this new topology.

6.7. Link Down with Layer 2 Detection

- Using reference topology 4 (Parallel Links), begin with OSPF in the full state between the generator and the DUT. Both links should be point-to-point links with the ability to notify the operating system immediately upon link failure.

- Disable link 1; this should be done in such a way that the keepalive timers at the data link layer will have no impact on the DUT recognizing the link failure (the operating system in the DUT should recognize this link failure immediately). Disconnecting the cable on the generator end would be one possibility, or shutting the link down.

- Note the time of the link failure on the generator.

- At the generator, note the time of the receipt of the new router LSA from the DUT notifying the generator of the link 2 failure.

  The difference in the time between the initial link failure and the receipt of the LSA on the generator across link 2 should be taken as the time required for an OSPF implementation to recognize and process a link failure, including the time required to generate and flood an LSA describing the link down event to an adjacent neighbor.

6.8. Link Down with Layer 3 Detection

- Using reference topology 4 (Parallel Links), begin with OSPF in the full state between the generator and the DUT.

- Disable OSPF processing on link 1 from the generator. This should be done in such a way so it does not affect link status; the DUT MUST note the failure of the adjacency through the dead interval.

- At the generator, note the time of the receipt of the new router LSA from the DUT notifying the generator of the link 2 failure.

The difference in the time between the initial link failure and the receipt of the LSA on the generator across link 2 should be taken as the time required for an OSPF implementation to recognize and process an adjacency failure.

7. IANA Considerations

This document requires no IANA considerations.

8. Security Considerations

This document does not modify the underlying security considerations in [OSPF].

9. Acknowledgements

Thanks to Howard Berkowitz, (hcb@clark.net), for his encouragement and support. Thanks also to Alex Zinin (zinin@psg.net), Gurpreet Singh (Gurpreet.Singh@SpirentCom.COM), and Yasuhiro Ohara (yasu@sfc.wide.ad.jp) for their comments as well.

10. Normative References

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

[TERM]Manral, V., "OSPF Convergence Testing Terminology and Concepts", draft-ietf-bmwg-ospfconv-term-08, June 2004

[CONSIDERATIONS]
Manral, V., "Considerations When Using Basic OSPF Convergence Benchmarks", draft-ietf-bmwg-ospfconv-applicability-06, June 2004

[RFC2119]
Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997

11. Informative References

[INTERCONNECT]
Bradner, S., McQuaid, J., "Benchmarking Methodology for Network Interconnect Devices", RFC2544, March 1999.

[MILLISEC]
Alaettinoglu C., et al., "Towards Milli-Second IGP Convergence" draft-alaettinoglu-isis-convergence

[FIB-TERM]
Trotter, G., "Terminology for Forwarding Information Base (FIB) based Router Performance", RFC3222, October 2001.

[BLACKBOX]
Shaikh, Aman, Greenberg, Albert, "Experience in Black-Box OSPF measurement"

12. Authors' Addresses

Vishwas Manral
Netplane Systems
189 Prashasan Nagar
Road number 72
Jubilee Hills
Hyderabad, India

vmanral@netplane.com

Russ White
Cisco Systems, Inc.
7025 Kit Creek Rd.
Research Triangle Park, NC 27709

riw@cisco.com

Aman Shaikh
AT&T Labs (Research)
180, Park Av
Florham Park, NJ 07932

ashaikh@research.att.com

13. Full Copyright Statement

Copyright (C) The Internet Society (2004). This document is subject to the rights, licenses and restrictions contained in BCP 78, and except as set forth therein, the authors retain all their rights.

This document and the information contained herein are provided on an "AS IS" basis and THE CONTRIBUTOR, THE ORGANIZATION HE/SHE REPRESENTS OR IS SPONSORED BY (IF ANY), THE INTERNET SOCIETY AND THE INTERNET ENGINEERING TASK FORCE DISCLAIM ALL WARRANTIES, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO ANY WARRANTY THAT THE USE OF THE INFORMATION HEREIN WILL NOT INFRINGE ANY RIGHTS OR ANY IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE.

14. Intellectual Property

By submitting this Internet-Draft, I certify that any applicable patent or other IPR claims of which I am aware have been disclosed, and any of which I become aware will be disclosed, in accordance with RFC 3668.

The IETF takes no position regarding the validity or scope of any Intellectual Property Rights or other rights that might be claimed to pertain to the implementation or use of the technology described in this document or the extent to which any license under such rights might or might not be available; nor does it represent that it has made any independent effort to identify any such rights. Information on the procedures with respect to rights in RFC documents can be found in BCP 78 and BCP 79.

Copies of IPR disclosures made to the IETF Secretariat and any assurances of licenses to be made available, or the result of an attempt made to obtain a general license or permission for the use of such proprietary rights by implementers or users of this specification can be obtained from the IETF on-line IPR repository at http://www.ietf.org/ipr.

The IETF invites any interested party to bring to its attention any copyrights, patents or patent applications, or other proprietary rights that may cover technology that may be required to implement this standard. Please address the information to the IETF at ietf-ipr@ietf.org.