      Benchmarking Basic OSPF Single Router Control Plane Convergence

Status of This Memo

Copyright Notice

Abstract

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

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

Table of Contents

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 the apparent performance that routing protocols
   provide.

   This document attempts to provide a framework for Open Shortest Path
   First [OSPF] performance testing, and to provide some tests for
   measuring some aspects of OSPF performance.  The motivation of the
   document 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].  RFC 2119
   key words in this document are used to ensure methodological control,
   which is very important in the specification of benchmarks.  This
   document does not specify a network-related protocol.

3.  Overview and Scope

   Although 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 produce directly comparable convergence-related
   measurements.

   The following considerations are outside the scope of this document:

   o  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].

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

      The 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).
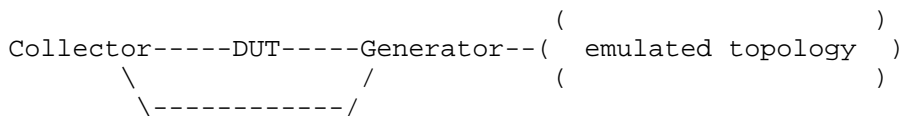
4.  Reference Topologies

   Several reference topologies that are used throughout the tests are
   described in the remaining sections of this document.  All of the
   topologies have been collectively placed in one section to avoid
   repetition.

   o  Reference Topology 1 (Emulated Topology)

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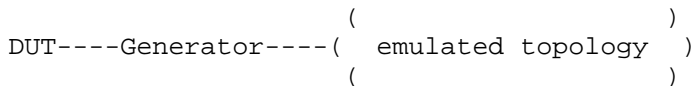
   A simple back-to-back configuration.  It's assumed that the link
   between the generator and the DUT is a point-to-point link, while
   the connections within the generator represent some emulated
   topology.

   o  Reference Topology 2 (Generator and Collector)

```
                                (                   )
      Collector-----DUT-----Generator--(  emulated topology  )
            \               /          (                   )
             \-----------/
```

   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)

```
      DUT      R1      R2
       |       |       |
      -+------+------+-----.....
```
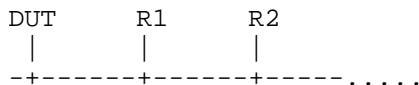
   Any number of routers could be included on the common broadcast
   network.

   o  Reference Topology 4 (Parallel Links)

```
       /--(link 1)-----\          (                   )
      DUT                Generator--(  emulated topology  )
       \--(link 2)-----/          (                   )
```
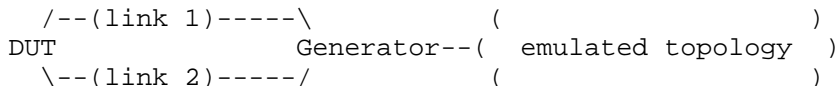
   In all cases the tests and topologies are designed to allow
   performance measurements to be taken all on a single device, whether
   this is the DUT or some other device in the network.  This eliminates
   the need for synchronized clocks within the test networks.

5.  Basic Performance Tests

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

   o  time required to process an LSA,

   o  flooding time, and

   o  Shortest Path First computation.

5.1.  Time Required to Process an LSA

   o  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 that the adjacency may be in
      Full state 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.

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

   o  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).

   o  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 [OSPF].  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 it does, or anywhere in between.  In other words, this
measurement may not mean the same thing in all implementations.

5.2.  Flooding Time

o  Using reference topology 2 (Generator and Collector), enable OSPF
   on all links and allow the devices to build full adjacencies.
   Configure the collector so that it will block all flooding toward
   the DUT (but so that it continues receiving advertisements from
   the DUT).

o  Inject a new set of LSAs from the generator toward the collector
   and the DUT.

o  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 from when the last LSA is received on the collector from the
generator to when 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

o  Use reference topology 1 (Emulated Topology), beginning with the
   DUT and the generator fully adjacent.

o  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].

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

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

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

o  Immediately inject another LSA that 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 that 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 starts after the duplicate LSA is acked.

o  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 dependent on the amount of
time between the transmissions of the first and second LSAs.  If too
much time elapsed, the test is meaningless because the SPF run will
complete before the second (duplicate) LSA is received.  If the time
elapsed is less, then both LSAs will 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 that 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), such as the
following:

o  Using reference topology 1 (Emulated Topology), establish a full
   adjacency between the generator and the DUT.

      o  Inject a set of LSAs from the generator toward the DUT.  Allow the
         DUT to stabilize and install all best paths in the routing table,
         etc.

      o  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.

      o  Measure the amount of time required for the DUT to compute a 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 a white box method of
      measuring SPF time is used.  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 an interesting relationship between the SPF times reported
      by white box (internal) testing and black box (external) testing;
      each of these two types of tests may be used as a "sanity check" on
      the other by comparing results.

      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:

      o  Forming Adjacencies on Point-to-Point Link (Initialization)

      o  Forming Adjacencies on Point-to-Point Links

      o  Link Up with Information Already in the Database

      o  Initial convergence Time on a Designated Router Electing
         (Broadcast) Network

      o  Link Down with Layer 2 Detection

      o  Link Down with Layer 3 Detection

      o  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.

   o  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.

   o  Inject a set of LSAs from the generator toward the DUT.

   o  Bring the link up at the DUT, noting the time when the link
      carrier is established on the generator.

   o  Note the time when 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.

   o  Using reference topology 1 (Emulated Topology), configure the DUT
      and the generator so that traffic can be passed along the link
      between them.

   o  Configure the generator so that OSPF is running on the point-to-
      point link toward the DUT, and inject a set of LSAs.

   o  Configure the DUT so that OSPF is initialized, but not running on
      the point-to-point link between the DUT and the generator.

   o  Enable OSPF on the interface between the DUT and the generator on
      the DUT.

   o  Note the time of the first hello received from the DUT on the
      generator.

   o  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

   o  Using reference topology 2 (Generator and Collector), configure
      all three devices to run OSPF.

   o  Configure the DUT so that the link between the DUT and the
      generator is disabled.

   o  Inject a set of LSAs into the network from the generator; the DUT
      should receive these LSAs through normal flooding from the
      collector.

   o  Enable the link between the DUT and the generator.

   o  Note the time of the first hello received from the DUT on the
      generator.

   o  Note the time of the last DBD (Database Description) received on
      the generator.

   o  Note the time of the acknowledgement from the DUT for the last LSA
      transmitted on the generator.

   The time between the hello received by the generator from the DUT 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

     o  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.

     o  Enable OSPF on the common broadcast link on all the routers in the
        test bed.

     o  Disable the broadcast link on R1.

     o  Note the time of the last hello received from R1 on R2.

     o  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 that this test includes the dead interval timer at
     the DUT, so this time may be factored out, or the hello and dead
     intervals may be reduced to lessen these timers' impact on the
     overall test times.  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 when this hello is received instead of the time when
     the last one was would provide a more accurate measurement.

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

     o  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.

     o  Enable the broadcast link between R1 and the DUT.  Note the time
        of the first hello received by R1.

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

     o  The difference 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.

This test maybe performed with a varying number of devices attached
to the broadcast network, and with 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
Section 6.4, above, may be subtracted from the results of this test
to provide just the convergence time across a broadcast network.

Note that although all the other tests in this 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

   o  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.

   o  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.

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

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

6.7.  Link Down with Layer 2 Detection

   o  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.

   o  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; shutting
the link down would be another.

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

o  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

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

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

o  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.  Security Considerations

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

8.  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.

9.  Normative References

   [OSPF]              Moy, J., "OSPF Version 2", STD 54, RFC 2328, April
                       1998.

   [TERM]             Manral, V., White, R., and A. Shaikh, "OSPF
                       Benchmarking Terminology and Concepts", RFC 4062,
                       April 2005.

   [CONSIDERATIONS] Manral, V., White, R., and A. Shaikh,
                       "Considerations When Using Basic OSPF Convergence
                       Benchmarks", RFC 4063, April 2005.

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

10.  Informative References

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

   [FIB-TERM]         Trotter, G., "Terminology for Forwarding Information
                       Base (FIB) based Router Performance", RFC 3222,
                       December 2001.

   [BLACKBOX]         Shaikh, A. and Greenberg, A., "Experience in Black-
                       box OSPF measurement", Proc. ACM SIGCOMM Internet
                       Measurement Workshop (IMW), November 2001

Authors' Addresses

    Vishwas Manral
    SiNett Corp,
    Ground Floor,
    Embassy Icon Annexe,
    2/1, Infantry Road,
    Bangalore, India

    EMail: vishwas@sinett.com


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

    EMail: riw@cisco.com


    Aman Shaikh
    AT&T Labs (Research)
    180 Park Av, PO Box 971
    Florham Park, NJ 07932

    EMail: ashaikh@research.att.com