Network Working Group George Swallow
Internet Draft Cisco Systems, Inc.
Category: Standards Track
Expiration Date: April 2004

Kireeti Kompella
Juniper Networks, Inc.

Dan Tappan
Cisco Systems, Inc.

October 2003

Label Switching Router Self-Test

draft-ietf-mpls-lsr-self-test-00.txt

Status of this Memo

Abstract

This document defines a means of self test for a Label-Switching Router (LSR) to verify that its dataplane is functioning for certain key Multi-Protocol Label Switching (MPLS) applications including unicast forwarding based on LDP [LDP] and traffic engineering tunnels based on [RSVP-TE]. A new Loopback FEC type

is defined to allow an upstream neighbor to assist in the testing at very low cost. MPLS Echo Request and MPLS Echo Reply messages [LSP-Ping] messages are extended to do the actually probing.

Contents

0. Sub-IP ID Summary

(This section to be removed before publication.)

(See Abstract above.)

RELATED DOCUMENTS

May be found in the "references" section.

WHERE DOES IT FIT IN THE PICTURE OF THE SUB-IP WORK

Fits in the MPLS box.

WHY IS IT TARGETED AT THIS WG

MPLS WG is currently looking at MPLS-specific error detection and recovery mechanisms. The mechanisms proposed here are for packet-based MPLS LSPs, which is why the MPLS WG is targeted.

JUSTIFICATION

The WG should consider this document, as it allows network operators to detect MPLS LSP data plane failures in the network. This type of failures have occurred, and are a source of concern to operators implementing MPLS networks.

1. Introduction

This document defines a means of self test for a Label-Switching Router (LSR) to verify that its dataplane is functioning for certain key Multi-Protocol Label Switching (MPLS) applications including unicast forwarding based on LDP [LDP] and traffic engineering tunnels based on [RSVP-TE]. MPLS Echo Request and MPLS Echo Reply messages [LSP-Ping] messages are extended to do the actually probing. The pings are sent to an upstream neighbor, looped back through the LSR under test and intercepted, by means of TTL expiration by a downstream neighbor. Extensions to LSP-Ping are defined to allow the down stream neighbor to verify the test results.

In order to minimize the load on upstream LSRs a new loopback FEC is defined. Receipt of a packet labeled with a loopback label will cause the advertising LSR to pop the label off the label stack and send the packet out the advertised interface.

Note that use of a loopback allows an LSR to label entries for which the LSR is not currently its potential upstream neighbor's next hop. In this way label entries can be verified prior to the occurrence of a routing change.

Some routing protocls, most notably OSPF have no means of exchanging "Link Local Identifiers" used to identify unnumbered links and components of bundled links. These same test procedures can be used to associate the neighbor's interfaces with the probing LSRs interfaces. This is achieved by simply having the TTL of the MPLS Ping expire one hop sooner, i.e. at the testing LSR itself.

1.1. Conventions

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 [KEYWORDS].

2. Loopback FEC

The Loopback FEC type is defined to enable an upstream neighbor to assist in the LSR self-testing at very low cost. This FEC allows the loopback to occur in the dataplane without control plane involvement beyond the initial LDP exchange and dataplane setup.

An LSR uses the Loopback FEC to selectively advertise loopback labels to its neighbor LSRs. Each loopback label is bound to a particular interface. For multiaccess links, one label per neighbor is required since the link-level address is derived from the label lookup. When an MPLS packet with its top label set to a loopback label is received from an interface over which that label was advertised, the loopback label is popped and the packet is sent on the interface to which the loopback label was bound.

TTL treatment for loopback labels follows the Uniform model. I.e. the TTL carried in the loopback label is decremented and copied to the exposed label or IP header as the case may be.

2.1. Loopback FEC Element

FEC element type 130 is used. The FEC element is encoded as follows: (note: 130 is provisionally assigned, the actual value will be assigned by IANA.)

```
    0                   1                   2                   3
    0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |     130       |       Res     | Interface Type|   Id Length   |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |                     Interface Identifier                      |
   |                              "                                |
   |                              "                                |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Reserved

Must be set to zero on transmission and ignored on receipt.

Interface Type

#	Type	Interface Identifier
0	Unnumbered	A 32 bit Link Identifier as defined in [RFC3477]
1	IPv4 Numbered	IPv4 Address
2	IPv6 Numbered	IPv6 Address

Identifier Length

Length of the interface identifier in octets.

Address

An identifier encoded according to the Identifier Type field. The length is 4 bytes for Unnumbered and IPv4, 16 bytes for IPv6.

2.2. LDP Procedures

It is RECOMMENDED that loopback labels only be distributed in response to a Label Request message, irrespective of the label advertisement mode of the LDP session. However it is recognized that in certain cases such as OSPF with unnumbered links, the upstream LSR may not have sufficiently detailed information of the neighbors link identifier to form the request. In these cases, the downstream LSR will need to be configured to make unsolicited advertisements.

3. Data Plane Self Test

A self test operation involves three LSRs, the LSR doing the test, an upstream neighbor and a downstream neighbor. We refer to these as LSRs T, U, and D respectively. The packet flow is shown below. Although the figure shows LSRD adjacent to LSRT it may in some cases be an arbitrary number of hops away.

```
       +-------+       +-------+       +-------+
       |    ,-|-------|-<MERq|       |       |  MERq: MPLS Echo Request
       |    `-|-------|------|-------|->    |
       |      |       |    <-|-------|-<MERp|  MERp: MPLS Echo Reply
       +-------+       +-------+       +-------+
         LSRU            LSRT            LSRD
```

Figure 1: Self Test Message Flow

In order to perform a test on an incoming label stack, LSRT forms an MPLS Echo Request. Included in that is a Next Hop Verification Object which describes the interface and label stack that should be seen by LSRD. Optionally a FEC Stack TLV may be included to verify LSRD's labels are mapped to the expected FECs.

LSRT prepends the packet with the incoming label stack and the loopback label received from LSRU. The TTL values are set such that they will expire at LSRD. LSRT then forward the packet to LSRU.

LSRU receives the packet and performs normal MPLS forwarding, that is the loopback label is pop, the TTL is decremented and propagated (in this case) to the exposed label.

LSRT receives the packet and performs normal MPLS forwarding. If everything is functioning as expected this will cause the packet to arrive at LSRD with a TTL of 1.

LSRD receives the packet. It verifies that the packet was received on the interface and with the label stack contained in the Next Hop Verification TLV. If a FEC Stack TLV is also included, it verifies that the labels in the stack are mapped to those FECs. The results are recorded in an MPLS Echo Reply message and sent to LSRT.

3.1. Next Hop Verification Object

The Next Hop Verification is an optional TLV in an echo request. The Length is 12 + 4*N octets, where N is the number of Downstream Labels. The value field of a Next Hop Verification TLV has the following format:

```
    0                   1                   2                   3
    0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |                     Next Hop IPv4 Router ID                   |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |                               | Address Type  |               |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |                    Next Hop Interface Address                 |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |                 Outgoing Label                |    Reserved   |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   .                                                               .
   .                                                               .
   .                                                               .
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |                 Outgoing Label                |    Reserved   |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

The Next Hop Interface Address Type is one of:

Type #	Address Type
1	IPv4
2	Unnumbered

Reserved

Must be set to zero on transmission and ignored on receipt.

[Note a format for IPv6 will be added after one is defined in [LSP-Ping]]

3.2. Additional Error Codes

The following error codes are defined in addition to those defined in [LSP-Ping].

Value	Meaning
8	Incorrect interface
9	Label Stack mis-match
10	Label not valid on this interface
11	Binding for this label is not the given FEC
12	Not processed

3.3. Sending procedures

In order to perform a test on an incoming label stack, an LSR first determines the expected outgoing label stack, next hop router and next hop interface. These are recorded in a Next Hop Verification TLV.

If binding verification is also to be performed, then beginning at the top of the stack, for each label binding received or given to the Next Hop router, a corresponding entry in a FEC Stack TLV is included. If there are labels in the stack that are not to be processed by the Next Hop Router, corresponding entries MUST NOT be included in the FEC Stack TLV.

The LSR creates an MPLS Echo Request packet with itself as the source address and the destination set to an address in the range of 127/8. The IP TTL SHOULD be set to 1. The incoming label stack is prepended to the packet. The TTL of these labels SHOULD be set to appropriate values - 2 for those labels which will be process by itself when the packet is looped back; 1 for those labels which will be carried through. Finally the loopback label received for the incoming interface is prepended to the packet. The TTL is set such that it will have the value of 3 on the wire.

The packet is sent to the upstream neighbor on an interface for which the loopback label is valid.

3.4. Receiving procedures

The next hop router performs the following checks on the received packet.

1. Verify that the Next Hop IPv4(v6) Router ID matches one of its IP addresses.

   If this fails, return an error code 5, 'Replying router is not one of the Downstream Routers'

2. Verify that the Next Hop Interface Address matches the interface on which this packet arrived.

   If this fails, return an error code xx1, 'Incorrect interface'. A Next Hop Verification TLV SHOULD be included in the reply with the received interface and label stack.

3. Verify that the label stack on the packet matches the label stack in the Next Hop Verification TLV.

   If this fails, return an error code xx2, 'Label Stack mis-match'. A Next Hop Verification TLV SHOULD be included in the reply with the received interface and label stack.

4. Verify that the top label on the stack is valid on this interface.

   If this fails, return an error code xx3, 'Label not valid on this interface'. The error sub-code is set to the stack depth of the errant label.

4a. If the label operation in step 4 is a Pop and requires that the next label be inspected, repeat step 4 for that label.

5. If a FEC Stack TLV is present, then for each FEC verify that the corresponding label is the correct binding. If the label binding is incorrect, return an error code of xx4 'Binding for this label is not the given FEC'. The error sub-code is set to the stack depth of the errant label.

   If a FEC is reached that was not processed in step 4 above, return an error code of xx5, 'Not processed'. The error sub-code is set to the stack depth of the errant label.

3.5. Upstream Neighbor Verification

To verify that an upstream neighbor is properly echoing packets an LSR may send an MPLS Echo Request packet with the TTL set so that the packet will expire upon reaching reaching itself. This procedure not only tests that the neighbor is correctly processing the loopback label, it also allow the node to verify the neighbor's interface mapping.

```
    +------+       +------+       +------+
    |    ,-|-------|-<MERq|       |      |  MERq: MPLS Echo Request
    |    `-|-------|->    |       |      |
    |      |       |      |       |      |
    +------+       +------+       +------+
      LSRU           LSRT           LSRD
```

Figure 2: Upstream Neighbor Verification

No TLVs need to be included in the MPLS Echo Request. By noting the Sender's Handle and Sequence Number, as well as the loopback label, LSRT is able to detect that a) the packet was looped, and b) determine (or verify) the interface on which the packet was received. A Next Hop Verification TLV may be included to assist in verification. This may be particularly useful in a system where control is distributed over multiple processor.

4. Security Considerations

Were loopback labels widely known, they might be subject to abuse. It is therefore RECOMMENDED that loopback labels only be shared between trusted neighbors. Further, if the loopback labels are drawn from the Global Label Space, or any other label space shared across multiple LDP sessions, it is RECOMMENDED that all loopback label be filtered from a session except those labels pertaining to interfaces directly connected to the neighbor participating in that session.

5. IANA Considerations

TBD

6. Acknowledgments

The authors would like to thank Vanson Lim, Tom Nadeau, and Bob Thomas for their comments and suggestions.

7. References

7.1. Normative References

[RFC3036] Andersson, L. et al., "LDP Specification", January 2001.

[LSP-Ping] Bonica, R. et al., "Detecting MPLS Data Plane Liveness", work-in-progress.

[RFC3477] Kompella, K. & Y. Rekhter, "Signalling Unnumbered Links in Resource ReSerVation Protocol - Traffic Engineering (RSVP-TE)", January 2003.

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

7.2. Informative References

[RSVP-TE] Awduche, D., et al, "RSVP-TE: Extensions to RSVP for LSP tunnels", RFC 3209, December 2001.

8. Authors' Addresses

Kireeti Kompella
Juniper Networks, Inc.
1194 N. Mathilda Ave.
Sunnyvale, CA 94089
Email: kireeti@juniper.net

George Swallow
Cisco Systems, Inc.
1414 Massachusetts Ave
Boxborough, MA 01719

Email: swallow@cisco.com

Dan Tappan
Cisco Systems, Inc.
1414 Massachusetts Ave
Boxborough, MA 01719

Email: tappan@cisco.com