Internet Engineering Task Force (IETF)
Request for Comments: 6424
Updates: 4379
Category: Standards Track
ISSN: 2070-1721

N. Bahadur
K. Kompella
Juniper Networks, Inc.
G. Swallow
Cisco Systems

November 2011

Mechanism for Performing Label Switched Path Ping (LSP Ping) over MPLS Tunnels

Abstract

This document describes methods for performing LSP ping (specified in RFC 4379) traceroute over MPLS tunnels and for traceroute of stitched MPLS Label Switched Paths (LSPs). The techniques outlined in RFC 4379 are insufficient to perform traceroute Forwarding Equivalency Class (FEC) validation and path discovery for an LSP that goes over other MPLS tunnels or for a stitched LSP. This document deprecates the Downstream Mapping TLV (defined in RFC 4379) in favor of a new TLV that, along with other procedures outlined in this document, can be used to trace such LSPs.

Status of This Memo

This is an Internet Standards Track document.

This document is a product of the Internet Engineering Task Force (IETF). It represents the consensus of the IETF community. It has received public review and has been approved for publication by the Internet Engineering Steering Group (IESG). Further information on Internet Standards is available in Section 2 of RFC 5741.

Information about the current status of this document, any errata, and how to provide feedback on it may be obtained at http://www.rfc-editor.org/info/rfc6424.

Copyright Notice

Table of Contents

1. Introduction 4
1.1. Conventions Used in This Document 4
2. Motivation 4
3. Packet Format 5
3.1. Summary of Changes 5
3.2. New Return Codes 6
3.2.1. Return Code per Downstream 6
3.2.2. Return Code for Stitched LSPs 6
3.3. Downstream Detailed Mapping TLV 7
3.3.1. Sub-TLVs 9
3.3.1.1. Multipath Data Sub-TLV 9
3.4. Deprecation of Downstream Mapping TLV 13
4. Performing MPLS Traceroute on Tunnels 13
4.1. Transit Node Procedure 13
4.1.1. Addition of a New Tunnel 13
4.1.2. Transition between Tunnels 14
4.1.3. Modification to FEC Validation Procedure on Transit . 16
4.2. Modification to FEC Validation Procedure on Egress 16
4.3. Ingress Node Procedure 17
4.3.1. Processing Downstream Detailed Mapping TLV 17
4.3.1.1. Stack Change Sub-TLV Not Present 17
4.3.1.2. Stack Change Sub-TLV(s) Present 17
4.3.2. Modifications to Handling a Return Code 3 Reply. . . . 19
4.3.3. Handling of New Return Codes 19
4.4. Handling Deprecated Downstream Mapping TLV 20
5. Security Considerations 20
6. IANA Considerations 21
7. Acknowledgements 22
8. References 22
8.1. Normative References 22
8.2. Informative References 22

1. Introduction

This documents describes methods for performing LSP ping (specified in [RFC4379]) traceroute over MPLS tunnels. The techniques in [RFC4379] outline a traceroute mechanism that includes Forwarding Equivalency Class (FEC) validation and Equal Cost Multi-Path (ECMP) path discovery. Those mechanisms are insufficient and do not provide details when the FEC being traced traverses one or more MPLS tunnels and when Label Switched Path (LSP) stitching [RFC5150] is in use. This document deprecates the Downstream Mapping TLV [RFC4379], introducing instead a new TLV that is more extensible and that enables retrieval of detailed information. Using the new TLV format along with the existing definitions of [RFC4379], this document describes procedures by which a traceroute request can correctly traverse MPLS tunnels with proper FEC and label validations.

1.1. Conventions Used in This Document

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].

2. Motivation

An LSP ping traceroute may cross multiple MPLS tunnels en route to the destination. Let us consider a simple case.

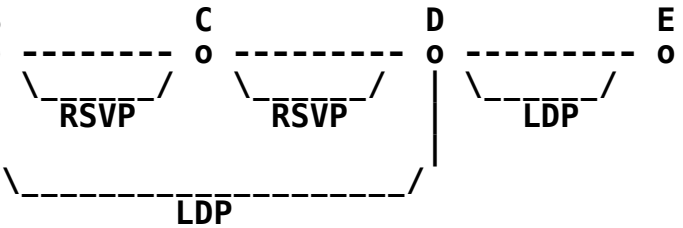

Figure 1: LDP over RSVP Tunnel

When a traceroute is initiated from router A, router B returns downstream mapping information for node C in the MPLS echo reply. The next MPLS echo request reaches router C with an LDP FEC. Node C is a pure RSVP node and does not run LDP. Node C will receive the MPLS echo request with two labels but only one FEC in the Target FEC stack. Consequently, node C will be unable to perform a complete FEC validation. It will let the trace continue by just providing next-hop information based on the incoming label, and by looking up the forwarding state associated with that label. However, ignoring FEC validation defeats the purpose of control-plane validations. The

MPLS echo request should contain sufficient information to allow node C to perform FEC validations to catch any misrouted echo requests.

The above problem can be extended for a generic case of hierarchical tunnels or stitched tunnels (e.g., B-C can be a separate RSVP tunnel and C-D can be a separate RSVP tunnel). The problem of FEC validation for tunnels can be solved if the transit routers (router B in the above example) provide some information to the ingress regarding the start of a new tunnel.

Stitched LSPs involve two or more LSP segments stitched together. The LSP segments can be signaled using the same or different signaling protocols. In order to perform an end-to-end trace of a stitched LSP, the ingress needs to know FEC information regarding each of the stitched LSP segments. For example, consider the figure below.

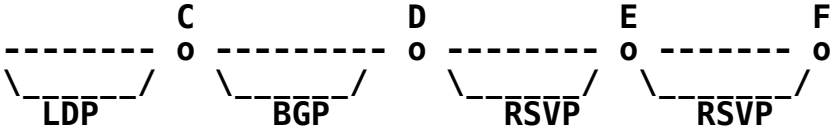

Figure 2: Stitched LSP

Consider ingress (A) tracing end-to-end stitched LSP A--F. When an MPLS echo request reaches router C, there is a FEC stack change happening at router C. With current LSP ping [RFC4379] mechanisms, there is no way to convey this information to A. Consequently, when the next echo request reaches router D, router D will know nothing about the LDP FEC that A is trying to trace.

Thus, the procedures defined in [RFC4379] do not make it possible for the ingress node to:

1. Know that tunneling has occurred.

2. Trace the path of the tunnel.

3. Trace the path of stitched LSPs.

3. Packet Format

3.1. Summary of Changes

In many cases, there is a need to associate additional data in the MPLS echo reply. In most cases, the additional data needs to be associated on a per-downstream-neighbor basis. Currently, the MPLS echo reply contains one Downstream Mapping TLV (DSMAP) per downstream

neighbor. However, the DSMAP format is not extensible; hence, it is not possible to associate more information with a downstream neighbor. This document defines a new extensible format for the DSMAP and provides mechanisms for solving the tunneled LSP ping problem using the new format. In summary, this document makes the following TLV changes:

- Addition of new Downstream Detailed Mapping TLV (DDMAP).
- Deprecation of existing Downstream Mapping TLV (DSMAP).
- Addition of Downstream FEC stack change sub-TLV to DDMAP.

3.2. New Return Codes

3.2.1. Return Code per Downstream

A new Return Code is being defined "See DDM TLV for Return Code and Return Subcode" (Section 6.3) to indicate that the Return Code is per Downstream Detailed Mapping TLV (Section 3.3). This Return Code MUST be used only in the message header and MUST be set only in the MPLS echo reply message. If the Return Code is set in the MPLS echo request message, then it MUST be ignored. When this Return Code is set, each Downstream Detailed Mapping TLV MUST have an appropriate Return Code and Return Subcode. This Return Code MUST be used when there are multiple downstreams for a given node (such as Point to Multipoint (P2MP) or Equal Cost Multi-Path (ECMP)), and the node needs to return a Return Code/Return Subcode for each downstream. This Return Code MAY be used even when there is only one downstream for a given node.

3.2.2. Return Code for Stitched LSPs

When a traceroute is being performed on stitched LSPs (Section 4.1.2), the stitching point SHOULD indicate the stitching action to the node performing the trace. This is done by setting the Return Code to "Label switched with FEC change" (Section 6.3). If a node is performing FEC hiding, then it MAY choose to set the Return Code to a value (specified in [RFC4379]) other than "Label switched with FEC change". The Return Code "Label switched with FEC change" MUST NOT be used if no FEC stack sub-TLV (Section 3.3.1.3) is present in the Downstream Detailed Mapping TLV(s). This new Return Code MAY be used for hierarchical LSPs (for indicating the start or end of an outer LSP).

3.3. Downstream Detailed Mapping TLV

Type #	Value Field
20	Downstream Detailed Mapping

The Downstream Detailed Mapping object is a TLV that MAY be included in an MPLS echo request message. Only one Downstream Detailed Mapping object may appear in an echo request. The presence of a Downstream Detailed Mapping object is a request that Downstream Detailed Mapping objects be included in the MPLS echo reply. If the replying router is the destination (Label Edge Router) of the FEC, then a Downstream Detailed Mapping TLV SHOULD NOT be included in the MPLS echo reply. Otherwise, the replying router SHOULD include a Downstream Detailed Mapping object for each interface over which this FEC could be forwarded.

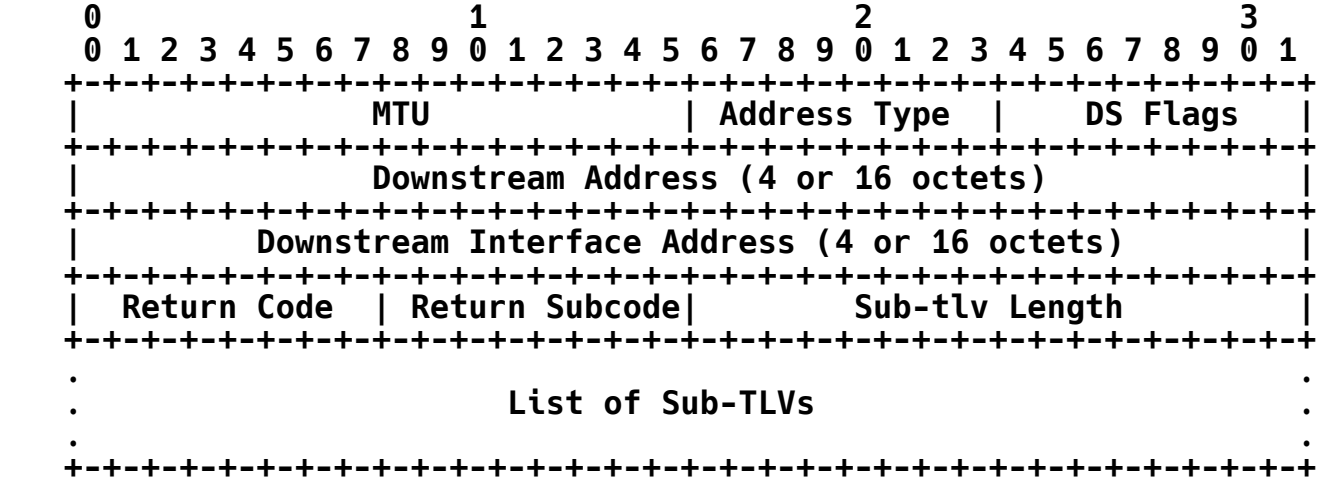

Figure 3: Downstream Detailed Mapping TLV

The Downstream Detailed Mapping TLV format is derived from the Downstream Mapping TLV format. The key change is that variable length and optional fields have been converted into sub-TLVs. The fields have the same use and meaning as in [RFC4379]. A summary of the fields taken from the Downstream Mapping TLV is as below:

Maximum Transmission Unit (MTU)

The MTU is the size in octets of the largest MPLS frame (including label stack) that fits on the interface to the Downstream Label Switching Router (LSR).

Address Type

The Address Type indicates if the interface is numbered or unnumbered. It also determines the length of the Downstream IP Address and Downstream Interface fields.

DS Flags

The DS Flags field is a bit vector of various flags.

Downstream Address and Downstream Interface Address

IPv4 addresses and interface indices are encoded in 4 octets; IPv6 addresses are encoded in 16 octets. For details regarding setting the address value, refer to [RFC4379].

The newly added sub-TLVs and their fields are as described below.

Return Code

The Return Code is set to zero by the sender. The receiver can set it to one of the values specified in the "Multi-Protocol Label Switching (MPLS) Label Switched Paths (LSPs) Ping Parameters" registry, "Return Codes" sub-registry.

If the receiver sets a non-zero value of the Return Code field in the Downstream Detailed Mapping TLV, then the receiver MUST also set the Return Code field in the echo reply header to "See DDM TLV for Return Code and Return Subcode" (Section 6.3). An exception to this is if the receiver is a bud node [RFC4461] and is replying as both an egress and a transit node with a Return Code of 3 ("Replying router is an egress for the FEC at stack-depth <RSC>") in the echo reply header.

If the Return Code of the echo reply message is not set to either "See DDM TLV for Return Code and Return Subcode" (Section 6.3) or "Replying router is an egress for the FEC at stack-depth <RSC>", then the Return Code specified in the Downstream Detailed Mapping TLV MUST be ignored.

Return Subcode

The Return Subcode is set to zero by the sender. The receiver can set it to one of the values specified in the "Multi-Protocol Label Switching (MPLS) Label Switched Paths (LSPs) Ping Parameters" registry, "Return Codes" sub-registry. This field is filled in with the stack-depth for those codes that specify the stack-depth. For all other codes, the Return Subcode MUST be set to zero.

If the Return Code of the echo reply message is not set to either "See DDM TLV for Return Code and Return Subcode" (Section 6.3) or "Replying router is an egress for the FEC at stack-depth <RSC>", then the Return Subcode specified in the Downstream Detailed Mapping TLV MUST be ignored.

Sub-tlv Length

Total length in bytes of the sub-TLVs associated with this TLV.

3.3.1. Sub-TLVs

This section defines the sub-TLVs that MAY be included as part of the Downstream Detailed Mapping TLV.

Sub-Type	Value Field
1	Multipath data
2	Label stack
3	FEC stack change

3.3.1.1. Multipath Data Sub-TLV

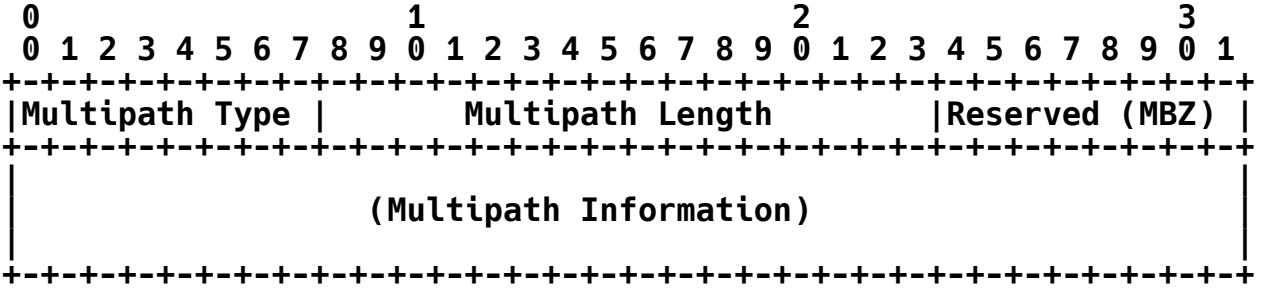

Figure 4: Multipath Sub-TLV

The multipath data sub-TLV includes Multipath Information. The sub-TLV fields and their usage is as defined in [RFC4379]. A brief summary of the fields is as below:

Multipath Type

The type of the encoding for the Multipath Information.

Multipath Length

The length in octets of the Multipath Information.

MBZ

MUST be set to zero when sending; MUST be ignored on receipt.

Multipath Information

Encoded multipath data, according to the Multipath Type.

3.3.1.2. Label Stack Sub-TLV

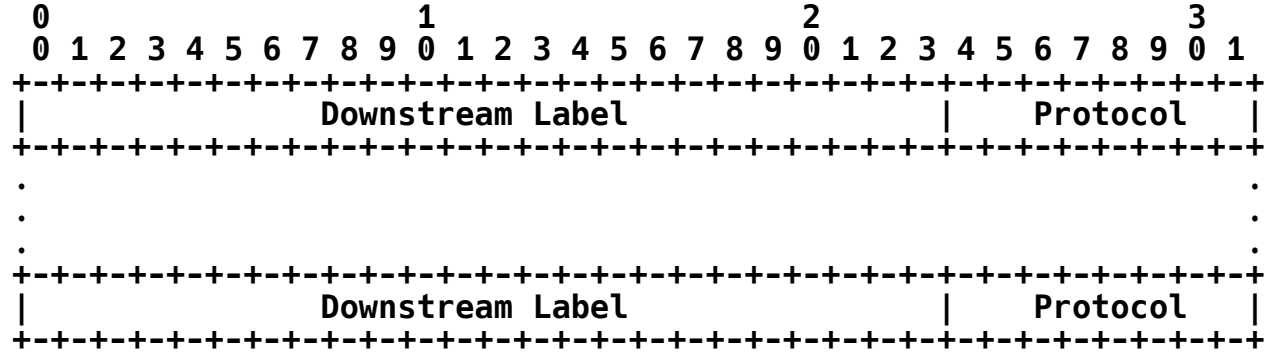

Figure 5: Label Stack Sub-TLV

The Label stack sub-TLV contains the set of labels in the label stack as it would have appeared if this router were forwarding the packet through this interface. Any Implicit Null labels are explicitly included. The number of label/protocol pairs present in the sub-TLV is determined based on the sub-TLV data length. The label format and protocol type are as defined in [RFC4379]. When the Downstream Detailed Mapping TLV is sent in the echo reply, this sub-TLV MUST be included.

Downstream Label

A Downstream label is 24 bits, in the same format as an MPLS label minus the Time to Live (TTL) field, i.e., the MSBit of the label is bit 0, the LSBit is bit 19, the Traffic Class (TC) field [RFC5462] is bits 20-22, and S is bit 23. The replying router SHOULD fill in the TC field and S bit; the LSR receiving the echo reply MAY choose to ignore these.

Protocol

This specifies the label distribution protocol for the Downstream label.

3.3.1.3. FEC Stack Change Sub-TLV

A router MUST include the FEC stack change sub-TLV when the downstream node in the echo reply has a different FEC Stack than the FEC Stack received in the echo request. One or more FEC stack change sub-TLVs MAY be present in the Downstream Detailed Mapping TLV. The format is as below.

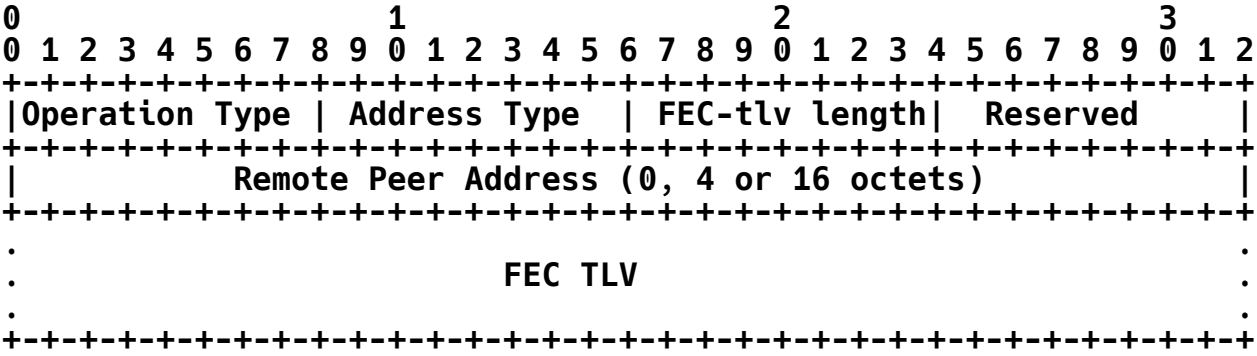

Figure 6: FEC Stack Change Sub-TLV

Operation Type

The operation type specifies the action associated with the FEC stack change. The following operation types are defined:

Type #	Operation
1	Push
2	Pop

Address Type

The Address Type indicates the remote peer's address type. The Address Type is set to one of the following values. The length of the peer address is determined based on the address type. The address type MAY be different from the address type included in the Downstream Detailed Mapping TLV. This can happen when the LSP goes over a tunnel of a different address family. The address type MAY be set to Unspecified if the peer address is either unavailable or the transit router does not wish to provide it for security or administrative reasons.

Type #	Address Type	Address length
0	Unspecified	0
1	IPv4	4
2	IPv6	16

FEC TLV Length

Length in bytes of the FEC TLV.

Reserved

This field is reserved for future use and MUST be set to zero.

Remote Peer Address

The remote peer address specifies the remote peer that is the next-hop for the FEC being currently traced. For example, in the LDP over RSVP case in Figure 1, router B would respond back with the address of router D as the remote peer address for the LDP FEC being traced. This allows the ingress node to provide information regarding FEC peers. If the operation type is PUSH, the remote peer address is the address of the peer from which the FEC being pushed was learned. If the operation type is POP, the remote peer address MAY be set to Unspecified.

For upstream-assigned labels [RFC5331], an operation type of POP will have a remote peer address (the upstream node that assigned the label) and this SHOULD be included in the FEC stack change sub-TLV. The remote peer address MAY be set to Unspecified if the address needs to be hidden.

FEC TLV

The FEC TLV is present only when the FEC-tlv length field is non-zero. The FEC TLV specifies the FEC associated with the FEC stack change operation. This TLV MAY be included when the operation type is POP. It MUST be included when the operation type is PUSH. The FEC TLV contains exactly one FEC from the list of FECs specified in [RFC4379]. A Nil FEC MAY be associated with a PUSH operation if the responding router wishes to hide the details of the FEC being pushed.

FEC stack change sub-TLV operation rules are as follows:

a. A FEC stack change sub-TLV containing a PUSH operation MUST NOT be followed by a FEC stack change sub-TLV containing a POP operation.

b. One or more POP operations MAY be followed by one or more PUSH operations.

c. One FEC stack change sub-TLV MUST be included per FEC stack change. For example, if 2 labels are going to be pushed, then one FEC stack change sub-TLV MUST be included for each FEC.

d. A FEC splice operation (an operation where one FEC ends and another FEC starts, see Figure 7) MUST be performed by including a POP type FEC stack change sub-TLV followed by a PUSH type FEC stack change sub-TLV.

e. A Downstream detailed mapping TLV containing only one FEC stack change sub-TLV with Pop operation is equivalent to IS_EGRESS (Return Code 3, [RFC4379]) for the outermost FEC in the FEC stack. The ingress router performing the MPLS traceroute MUST treat such a case as an IS_EGRESS for the outermost FEC.

3.4. Deprecation of Downstream Mapping TLV

This document deprecates the Downstream Mapping TLV. LSP ping procedures should now use the Downstream Detailed Mapping TLV. Detailed procedures regarding interoperability between the deprecated TLV and the new TLV are specified in Section 4.4.

4. Performing MPLS Traceroute on Tunnels

This section describes the procedures to be followed by an LSP ingress node and LSP transit nodes when performing MPLS traceroute over MPLS tunnels.

4.1. Transit Node Procedure

4.1.1. Addition of a New Tunnel

A transit node (Figure 1) knows when the FEC being traced is going to enter a tunnel at that node. Thus, it knows about the new outer FEC. All transit nodes that are the origination point of a new tunnel SHOULD add the FEC stack change sub-TLV (Section 3.3.1.3) to the Downstream Detailed Mapping TLV (Figure 3) in the echo reply. The transit node SHOULD add one FEC stack change sub-TLV of operation type PUSH, per new tunnel being originated at the transit node.

A transit node that sends a Downstream FEC stack change sub-TLV in the echo reply SHOULD fill the address of the remote peer; which is the peer of the current LSP being traced. If the transit node does not know the address of the remote peer, it MUST set the address type to Unspecified.

The Label stack sub-TLV MUST contain one additional label per FEC being PUSHed. The label MUST be encoded as per Figure 5. The label value MUST be the value used to switch the data traffic. If the tunnel is a transparent pipe to the node, i.e. the data-plane trace will not expire in the middle of the new tunnel, then a FEC stack change sub-TLV SHOULD NOT be added and the Label stack sub-TLV SHOULD NOT contain a label corresponding to the hidden tunnel.

If the transit node wishes to hide the nature of the tunnel from the ingress of the echo request, then it MAY not want to send details about the new tunnel FEC to the ingress. In such a case, the transit node SHOULD use the Nil FEC. The echo reply would then contain a FEC stack change sub-TLV with operation type PUSH and a Nil FEC. The value of the label in the Nil FEC MUST be set to zero. The remote peer address type MUST be set to Unspecified. The transit node SHOULD add one FEC stack change sub-TLV of operation type PUSH, per new tunnel being originated at the transit node. The Label stack sub-TLV MUST contain one additional label per FEC being PUSHed. The label value MUST be the value used to switch the data traffic.

4.1.2. Transition between Tunnels

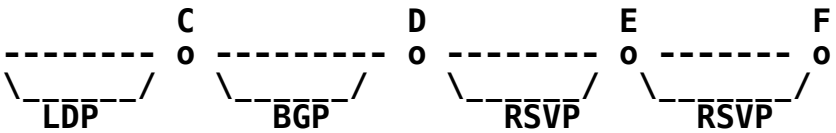

Figure 7: Stitched LSPs

In the above figure, we have three separate LSP segments stitched at C and D. Node C SHOULD include two FEC stack change sub-TLVs. One with a POP operation for the LDP FEC and one with the PUSH operation for the BGP FEC. Similarly, node D SHOULD include two FEC stack change sub-TLVs, one with a POP operation for the BGP FEC and one with a PUSH operation for the RSVP FEC. Nodes C and D SHOULD set the Return Code to "Label switched with FEC change" (Section 6.3) to indicate change in FEC being traced.

If node C wishes to perform FEC hiding, it SHOULD respond back with two FEC stack change sub-TLVs, one POP followed by one PUSH. The POP operation MAY either exclude the FEC TLV (by setting the FEC TLV length to 0) or set the FEC TLV to contain the LDP FEC. The PUSH

operation SHOULD have the FEC TLV containing the Nil FEC. The Return Code SHOULD be set to "Label switched with FEC change".

If node C performs FEC hiding and node D also performs FEC hiding, then node D MAY choose to not send any FEC stack change sub-TLVs in the echo reply since the number of labels has not changed (for the downstream of node D) and the FEC type also has not changed (Nil FEC). In such a case, node D MUST NOT set the Return Code to "Label switched with FEC change". If node D performs FEC hiding, then node F will respond as IS_EGRESS for the Nil FEC. The ingress (node A) will know that IS_EGRESS corresponds to the end-to-end LSP.

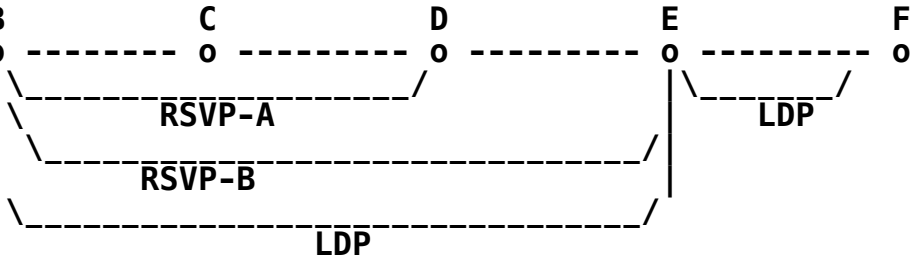

Figure 8: Hierarchical LSPs

In the above figure, we have an end-to-end LDP LSP between nodes A and F. The LDP LSP goes over RSVP LSP RSVP-B. LSP RSVP-B itself goes over another RSVP LSP RSVP-A. When node A initiates a traceroute for the end-to-end LDP LSP, then following sequence of FEC stack change sub-TLVs will be performed

Node B:

Respond with two FEC stack change sub-TLVs: PUSH RSVP-B, PUSH RSVP-A.

Node D:

Respond with Return Code 3 when RSVP-A is the top of FEC stack. When the echo request contains RSVP-B as top of stack, respond with Downstream information for node E and an appropriate Return Code.

If node B is performing tunnel hiding, then:

Node B:

Respond with two FEC stack change sub-TLVs: PUSH Nil FEC, PUSH Nil FEC.

Node D:

If D determines that the Nil FEC corresponds to RSVP-A, which terminates at D, then it SHOULD respond with Return Code 3. D can also respond with FEC stack change sub-TLV: POP (since D knows that number of labels towards next-hop is decreasing). Both responses would be valid.

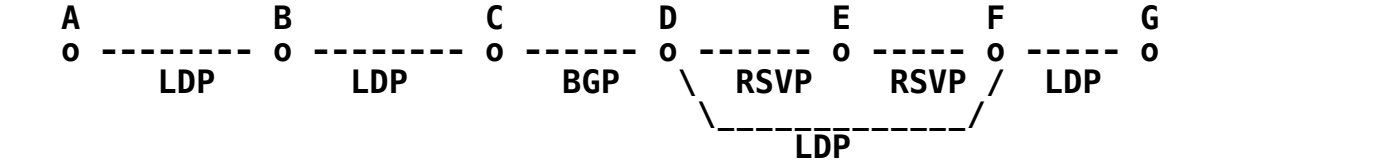

Figure 9: Stitched Hierarchical LSPs

In the above case, node D will send three FEC stack change sub-TLVs. One POP (for the BGP FEC) followed by two PUSHes (one for LDP and one for RSVP). Nodes C and D SHOULD set the Return Code to "Label switched with FEC change" (Section 6.3) to indicate change in FEC being traced.

4.1.3. Modification to FEC Validation Procedure on Transit

Section 4.4 of [RFC4379] specifies Target FEC stack validation procedures. This document enhances the FEC validation procedures as follows. If the outermost FEC of the target FEC stack is the Nil FEC, then the node MUST skip the target FEC validation completely. This is to support FEC hiding, in which the outer hidden FEC can be the Nil FEC.

4.2. Modification to FEC Validation Procedure on Egress

Section 4.4 of [RFC4379] specifies Target FEC stack validation procedures. This document enhances the FEC validation procedures as follows. If the outermost FEC of the Target FEC stack is the Nil FEC, then the node MUST skip the target FEC validation completely. This is to support FEC hiding, in which the outer hidden FEC can be the Nil FEC.

4.3. Ingress Node Procedure

It is the responsibility of an ingress node to understand tunnel within tunnel semantics and LSP stitching semantics when performing a MPLS traceroute. This section describes the ingress node procedure based on the kind of reply an ingress node receives from a transit node.

4.3.1. Processing Downstream Detailed Mapping TLV

Downstream Detailed Mapping TLV should be processed in the same way as the Downstream Mapping TLV, defined in Section 4.4 of [RFC4379]. This section describes the procedures for processing the new elements introduced in this document.

4.3.1.1. Stack Change Sub-TLV Not Present

This would be the default behavior as described in [RFC4379]. The ingress node MUST perform MPLS echo reply processing as per the procedures in [RFC4379].

4.3.1.2. Stack Change Sub-TLV(s) Present

If one or more FEC stack change sub-TLVs (Section 3.3.1.3) are received in the MPLS echo reply, the ingress node SHOULD process them and perform some validation.

The FEC stack changes are associated with a downstream neighbor and along a particular path of the LSP. Consequently, the ingress will need to maintain a FEC stack per path being traced (in case of multipath). All changes to the FEC stack resulting from the processing of FEC stack change sub-TLV(s) should be applied only for the path along a given downstream neighbor. The following algorithm should be followed for processing FEC stack change sub-TLVs.

```
    push_seen = FALSE
    fec_stack_depth = current-depth-of-fec-stack-being-traced
    saved_fec_stack = current_fec_stack

    while (sub-tlv = get_next_sub_tlv(downstream_detailed_map_tlv))

        if (sub-tlv == NULL) break

        if (sub-tlv.type == FEC-Stack-Change) {

            if (sub-tlv.operation == POP) {
                if (push_seen) {
                    Drop the echo reply
                    current_fec_stack = saved_fec_stack
                    return
                }

                if (fec_stack_depth == 0) {
                    Drop the echo reply
                    current_fec_stack = saved_fec_stack
                    return
                }

                Pop FEC from FEC stack being traced
                fec_stack_depth--;
            }

            if (sub-tlv.operation == PUSH) {
                push_seen = 1
                Push FEC on FEC stack being traced
                fec_stack_depth++;
            }
        }
     }


     if (fec_stack_depth == 0) {
         Drop the echo reply
         current_fec_stack = saved_fec_stack
         return
     }
```

Figure 10: FEC Stack Change Sub-TLV Processing Guideline

The next MPLS echo request along the same path should use the modified FEC stack obtained after processing the FEC stack change sub-TLVs. A non-Nil FEC guarantees that the next echo request along the same path will have the Downstream Detailed Mapping TLV validated for IP address, Interface address, and label stack mismatches.

If the top of the FEC stack is a Nil FEC and the MPLS echo reply does not contain any FEC stack change sub-TLVs, then it does not necessarily mean that the LSP has not started traversing a different tunnel. It could be that the LSP associated with the Nil FEC terminated at a transit node and at the same time a new LSP started at the same transit node. The Nil FEC would now be associated with the new LSP (and the ingress has no way of knowing this). Thus, it is not possible to build an accurate hierarchical LSP topology if a traceroute contains Nil FECs.

4.3.2. Modifications to Handling a Return Code 3 Reply.

The procedures above allow the addition of new FECs to the original FEC being traced. Consequently, a reply from a downstream node with Return Code 3 (IS_EGRESS) may not necessarily be for the FEC being traced. It could be for one of the new FECs that was added. On receipt of an IS_EGRESS reply, the LSP ingress should check if the depth of Target FEC sent to the node that just responded, was the same as the depth of the FEC that was being traced. If it was not, then it should pop an entry from the Target FEC stack and resend the request with the same TTL (as previously sent). The process of popping a FEC is to be repeated until either the LSP ingress receives a non-IS_EGRESS reply or until all the additional FECs added to the FEC stack have already been popped. Using an IS_EGRESS reply, an ingress can build a map of the hierarchical LSP structure traversed by a given FEC.

4.3.3. Handling of New Return Codes

When the MPLS echo reply Return Code is "Label switched with FEC change" (Section 3.2.2), the ingress node SHOULD manipulate the FEC stack as per the FEC stack change sub-TLVs contained in the downstream detailed mapping TLV. A transit node can use this Return Code for stitched LSPs and for hierarchical LSPs. In case of ECMP or P2MP, there could be multiple paths and Downstream Detailed Mapping TLVs with different Return Codes (Section 3.2.1). The ingress node should build the topology based on the Return Code per ECMP path/P2MP branch.

4.4. Handling Deprecated Downstream Mapping TLV

The Downstream Mapping TLV has been deprecated. Applications should now use the Downstream Detailed Mapping TLV. The following procedures SHOULD be used for backward compatibility with routers that do not support the Downstream Detailed Mapping TLV.

- The Downstream Mapping TLV and the Downstream Detailed Mapping TLV MUST never be sent together in the same MPLS echo request or in the same MPLS echo reply.

- If the echo request contains a Downstream Detailed Mapping TLV and the corresponding echo reply contains a Return Code 2 ("One or more of the TLVs was not understood"), then the sender of the echo request MAY resend the echo request with the Downstream Mapping TLV (instead of the Downstream Detailed Mapping TLV). In cases where a detailed reply is needed, the sender can choose to ignore the router that does not support the Downstream Detailed Mapping TLV.

- If the echo request contains a Downstream Mapping TLV, then a Downstream Detailed Mapping TLV MUST NOT be sent in the echo reply. This is to handle the case that the sender of the echo request does not support the new TLV. The echo reply MAY contain Downstream Mapping TLV(s).

- If echo request forwarding is in use (such that the echo request is processed at an intermediate LSR and then forwarded on), then the intermediate router is responsible for making sure that the TLVs being used among the ingress, intermediate and destination are consistent. The intermediate router MUST NOT forward an echo request or an echo reply containing a Downstream Detailed Mapping TLV if it itself does not support that TLV.

5. Security Considerations

1. If a network operator wants to prevent tracing inside a tunnel, one can use the Pipe Model [RFC3443], i.e., hide the outer MPLS tunnel by not propagating the MPLS TTL into the outer tunnel (at the start of the outer tunnel). By doing this, MPLS traceroute packets will not expire in the outer tunnel and the outer tunnel will not get traced.

2. If one doesn't wish to expose the details of the new outer LSP, then the Nil FEC can be used to hide those details. Using the Nil FEC ensures that the trace progresses without false negatives and all transit nodes (of the new outer tunnel) perform some minimal validations on the received MPLS echo requests.

Other security considerations, as discussed in [RFC4379], are also applicable to this document.

6. IANA Considerations

6.1. New TLV

IANA has assigned a TLV type value to the following TLV from the "Multiprotocol Label Switching Architecture (MPLS) Label Switched Paths (LSPs) Ping Parameters" registry, "TLVs and sub-TLVs" sub-registry.

Downstream Detailed Mapping TLV (see Section 3.3): 20.

6.2. New Sub-TLV Types and Associated Registry

IANA has registered the Sub-Type field of Downstream Detailed Mapping TLV. The valid range for this is 0-65535. Assignments in the range 0-16383 and 32768-49161 are made via Standards Action as defined in [RFC3692]; assignments in the range 16384-31743 and 49162-64511 are made via Specification Required [RFC4379]; values in the range 31744-32767 and 64512-65535 are for Vendor Private Use, and MUST NOT be allocated. If a sub-TLV has a Type that falls in the range for Vendor Private Use, the Length MUST be at least 4, and the first four octets MUST be that vendor's SMI Enterprise Code, in network octet order. The rest of the Value field is private to the vendor.

IANA has assigned the following sub-TLV types (see Section 3.3.1):

Multipath data: 1

Label stack: 2

FEC stack change: 3

6.3. New Return Codes

IANA has assigned new Return Code values from the "Multi-Protocol Label Switching (MPLS) Label Switched Paths (LSPs) Ping Parameters" registry, "Return Codes" sub-registry, as follows using a Standards Action value.

Value	Meaning
14	See DDM TLV for Return Code and Return Subcode
15	Label switched with FEC change

7. Acknowledgements

The authors would like to thank Yakov Rekhter and Adrian Farrel for their suggestions on the document.

8. References

8.1. Normative References

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

[RFC3692] Narten, T., "Assigning Experimental and Testing Numbers Considered Useful", BCP 82, RFC 3692, January 2004.

[RFC4379] Kompella, K. and G. Swallow, "Detecting Multi-Protocol Label Switched (MPLS) Data Plane Failures", RFC 4379, February 2006.

8.2. Informative References

[RFC3443] Agarwal, P. and B. Akyol, "Time To Live (TTL) Processing in Multi-Protocol Label Switching (MPLS) Networks", RFC 3443, January 2003.

[RFC4461] Yasukawa, S., "Signaling Requirements for Point-to-Multipoint Traffic-Engineered MPLS Label Switched Paths (LSPs)", RFC 4461, April 2006.

[RFC5150] Ayyangar, A., Kompella, K., Vasseur, JP., and A. Farrel, "Label Switched Path Stitching with Generalized Multiprotocol Label Switching Traffic Engineering (GMPLS TE)", RFC 5150, February 2008.

[RFC5331] Aggarwal, R., Rekhter, Y., and E. Rosen, "MPLS Upstream Label Assignment and Context-Specific Label Space", RFC 5331, August 2008.

[RFC5462] Andersson, L. and R. Asati, "Multiprotocol Label Switching (MPLS) Label Stack Entry: "EXP" Field Renamed to "Traffic Class" Field", RFC 5462, February 2009.

Authors' Addresses

Nitin Bahadur
Juniper Networks, Inc.
1194 N. Mathilda Avenue
Sunnyvale, CA 94089
US

Phone: +1 408 745 2000
EMail: nitinb@juniper.net
URI: www.juniper.net

Kireeti Kompella
Juniper Networks, Inc.
1194 N. Mathilda Avenue
Sunnyvale, CA 94089
US

Phone: +1 408 745 2000
EMail: kireeti@juniper.net
URI: www.juniper.net

George Swallow
Cisco Systems
1414 Massachusetts Ave
Boxborough, MA 01719
US

EMail: swallow@cisco.com
URI: www.cisco.com