MPLS Working Group
Internet Draft
Expiration Date: March 2000

Bilel Jamoussi, Editor
Nortel Networks Corp.

September 1999

Constraint-Based LSP Setup using LDP

draft-ietf-mpls-cr-ldp-03.txt

Status of this Memo

Abstract

Label Distribution Protocol (LDP) is defined in [1] for distribution of labels inside one MPLS domain. One of the most important services that may be offered using MPLS in general and LDP in particular is support for constraint-based routing of traffic across the routed network. Constraint-based routing offers the opportunity to extend the information used to setup paths beyond what is available for the routing protocol. For instance, an LSP can be setup based on explicit route constraints, QoS constraints, and other constraints. Constraint-based routing (CR) is a mechanism used to meet Traffic Engineering requirements that have been proposed by [2], [3] and [4]. These requirements may be met by extending LDP for support of constraint-based routed label switched paths (CR-LSPs). Other uses for CR-LSPs include MPLS-based VPNs.

This draft specifies mechanisms and TLVs for support of CR-LSPs using LDP.

Table of Contents

1. Introduction

The need for constraint-based routing (CR) in MPLS has been explored elsewhere [3], [2], and [4]. Explicit routing is a subset of the more general constraint-based routing function. At the MPLS WG meeting held during the Washington IETF (December 1997) there was consensus that LDP should support explicit routing of LSPs with provision for indication of associated (forwarding) priority. In the Chicago meeting (August 1998), a decision was made that support for explicit path setup in LDP will be moved to a separate document. This document provides that support and it has been accepted as a working document in the Orlando meeting (December 1998).

This specification proposes an end-to-end setup mechanism of a constraint-based routed LSP (CR-LSP) initiated by the ingress LSR. We also specify mechanisms to provide means for reservation of resources using LDP.

This document introduce TLVs and procedures that provide support for:

- Strict and Loose Explicit Routing
- Specification of Traffic Parameters
- Route Pinning
- CR-LSP Pre-emption though setup/holding priorities
- Handling Failures
- LSPID
- Resource Class

Section 2 introduces the various constraints defined in this specification. Section 3 outlines the CR-LDP solution. Section 4 defines the TLVs and procedures used to setup constraint-based routed label switched paths. Appendix A provides several examples of CR-LSP path setup. Appendix B provides Service Definition Examples.

2. Constraint-based Routing Overview

Constraint-based routing is a mechanism that supports the Traffic Engineering requirements defined in [4]. Explicit Routing is a subset of the more general constraint-based routing where the

constraint is the explicit route (ER). Other constraints are defined to provide a network operator with control over the path taken by an LSP. This section is an overview of the various constraints supported by this specification.

2.1 Strict and Loose Explicit Routes

Like any other LSP a CR-LSP is a path through an MPLS network. The difference is that while other paths are setup solely based on information in routing tables or from a management system, the constraint-based route is calculated at one point at the edge of network based on criteria, including but not limited to routing information. The intention is that this functionality shall give desired special characteristics to the LSP in order to better support the traffic sent over the LSP. The reason for setting up CR-LSPs might be that one wants to assign certain bandwidth or other Service Class characteristics to the LSP, or that one wants to make sure that alternative routes use physically separate paths through the network.

An explicit route is represented in a Label Request Message as a list of nodes or groups of nodes along the constraint-based route. When the CR-LSP is established, all or a subset of the nodes in a group may be traversed by the LSP. Certain operations to be performed along the path can also be encoded in the constraint-based route.

The capability to specify, in addition to specified nodes, groups of nodes, of which a subset will be traversed by the CR-LSP, allows the system a significant amount of local flexibility in fulfilling a request for a constraint-based route. This allows the generator of the constraint-based route to have some degree of imperfect information about the details of the path.

The constraint-based route is encoded as a series of ER-Hops contained in a constraint-based route TLV. Each ER-Hop may identify a group of nodes in the constraint-based route. A constraint-based route is then a path including all of the identified groups of nodes in the order in which they appear in the TLV.

To simplify the discussion, we call each group of nodes an abstract node. Thus, we can also say that a constraint-based route is a path including all of the abstract nodes, with the specified operations occurring along that path.

2.2 Traffic Characteristics

The traffic characteristics of a path are described in the Traffic Parameters TLV in terms of a peak rate, committed rate, and service granularity. The peak and committed rates describe the bandwidth constraints of a path while the service granularity can be used to

specify a constraint on the delay variation that the CR-LDP MPLS domain may introduce to a path's traffic.

2.3 Pre-emption

CR-LDP signals the resources required by a path on each hop of the route. If a route with sufficient resources can not be found, existing paths may be rerouted to reallocate resources to the new path. This is the process of path pre-emption. Setup and holding priorities are used to rank existing paths (holding priority) and the new path (setup priority) to determine if the new path can pre-empt an existing path.

The setupPriority of a new CR-LSP and the holdingPriority attributes of the existing CR-LSP are used to specify priorities. Signaling a higher holding priority express that the path, once it has been established, should have a lower chance of being pre-empted. Signaling a higher setup priority expresses the expectation that, in the case that resource are unavailable, the path is more likely to pre-empt other paths. The exact rules determining bumping are an aspect of network policy.

The allocation of setup and holding priority values to paths is an aspect of network policy.

The setup and holding priority values range from zero (0) to seven (7). The value zero (0) is the priority assigned to the most important path. It is referred to as the highest priority. Seven (7) is the priority for the least important path. The use of default priority values is an aspect of network policy.

The setupPriority of a CR-LSP should not be higher (numerically less) than its holdingPriority since it might bump an LSP and be bumped by the next _equivalent_ request.

2.4 Route Pinning

Route pinning is applicable to segments of an LSP that are loosely routed - i.e. those segments which are specified with a next hop with the `L' bit set or where the next hop is an _abstract node_. A CR-LSP may be setup using route pinning if it is undesirable to change the path used by an LSP even when a better next hop becomes available at some LSR along the loosely routed portion of the LSP.

2.5 Resource Class

The network operator may classify network resources in various ways. These classes are also known as _colors_ or _administrative groups_. When a CR-LSP is being established, it's necessary to indicate which resource classes the CR-LSP can draw from.

3. Solution Overview

CR-LSP over LDP Specification is designed with the following goals:

1. Meet the requirements outlined in [4] for performing traffic engineering and provide a solid foundation for performing more general constraint-based routing.

2. Build on already specified functionality that meets the requirements whenever possible. Hence, this specification is based on [1].

3. Keep the solution simple.

In this document, support for unidirectional point-to-point CR-LSPs is specified. Support for point-to-multipoint, multipoint-to-point, is for further study (FFS).

Support for constraint-based routed LSPs in this specification depends on the following minimal LDP behaviors as specified in [1]:

- Use of Basic and/or Extended Discovery Mechanisms.
- Use of the Label Request Message defined in [1] in downstream on demand label advertisement mode with ordered control.
- Use of the Label Mapping Message defined in [1] in downstream on demand mode with ordered control.
- Use of the Notification Message defined in [1].
- Use of the Withdraw and Release Messages defined in [1].

- Use of the Loop Detection (in the case of loosely routed segments of a CR-LSP) mechanisms defined in [1].

In addition, the following functionality is added to what's defined in [1]:

- The Label Request Message used to setup a CR-LSP includes one or more CR-TLVs defined in Section 4. For instance, the Label Request Message may include the ER-TLV.
- An LSR implicitly infers ordered control from the existence of one or more CR-TLVs in the Label Request Message. This means that the LSR can still be configured for independent control for LSPs established as a result of dynamic routing. However, when a Label Request Message includes one or more of the CR-TLVs, then ordered control is used to setup the CR-LSP. Note that this is also true for the loosely routed parts of a CR-LSP.
- New status codes are defined to handle error notification for failure of established paths specified in the CR-TLVs.

Optional TLVs are not required in the CR-LDP messages for the messages to be compliant with the protocol. Optional parameters MAY be required for a particular operation to work (or work correctly), however.

Examples of CR-LSP establishment are given in Appendix A to illustrate how the mechanisms described in this draft work.

3.1 Required Messages and TLVs

Any Messages, TLVs, and procedures not defined explicitly in this document are defined in the LDP Specification [1]. The state transitions, which relate to CR-LDP messages, can be found in [5]. The following subsections are meant as a cross-reference to the [1] document and indication of additional functionality beyond what's defined in [1] where necessary.

3.2 Label Request Message

The Label Request Message is as defined in 3.5.8 of [1] with the following modifications (required only if any of the CR-TLVs is included in the Label Request Message):

- Only a single FEC-TLV may be included in the Label Request Message. The CR-LSP FEC TLV should be used.

- The Optional Parameters TLV includes the definition of any of the Constraint-based TLVs specified in Section 4.

- The Procedures to handle the Label Request Message are augmented by the procedures for processing of the CR-TLVs as defined in Section 4.

The encoding for the CR-LDP Label Request Message is as follows:

```
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
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|0|   Label Request (0x0401)   |      Message Length            |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                     Message ID                                |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                     FEC TLV                                   |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                     LSPID TLV           (CR-LDP, mandatory)  |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                     ER-TLV              (CR-LDP, optional)    |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                     Traffic  TLV        (CR-LDP, optional)    |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                     Pinning TLV         (CR-LDP, optional)    |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                     Resource Class TLV (CR-LDP, optional)     |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                     Pre-emption  TLV     (CR-LDP, optional)   |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```


3.3 Label Mapping Message

The Label Mapping Message is as defined in 3.5.7 of [1] with the following modifications:

- Only a single Label-TLV may be included in the Label Mapping Message.

- The Label Mapping Message Procedures are limited to downstream on demand ordered control mode.

A Mapping message is transmitted by a downstream LSR to an upstream LSR under one of the following conditions:

1. The LSR is the egress end of the CR-LSP and an upstream mapping has been requested.

2. The LSR received a mapping from its downstream next hop LSR for an CR-LSP for which an upstream request is still

pending.

The encoding for the CR-LDP Label Mapping Message is as follows:

```
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
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|0|   Label Mapping (0x0400)   |      Message Length            |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                     Message ID                                |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                     FEC TLV                                   |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                     Label TLV                                 |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|              Label Request Message ID TLV                     |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                     LSPID TLV           (CR-LDP, optional)    |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                     Traffic  TLV        (CR-LDP, optional)    |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

3.4 Notification Message

The Notification Message is as defined in Section 3.5.1 of [1] and the Status TLV encoding is as defined in Section 3.4.6 of [1]. Establishment of an CR-LSP may fail for a variety of reasons. All such failures are considered advisory conditions and they are signaled by the Notification Message.

Notification Messages carry Status TLVs to specify events being signaled. New status codes are defined in Section 4.11 to signal error notifications associated with the establishment of a CR-LSP and the processing of the CR-TLV.

The Notification Message may carry the LSPID TLV of the corresponding CR-LSP.

Notification Messages MUST be forwarded toward the LSR originating the Label Request at each hop and at any time that procedures in this specification - or in [1] - specify sending of a Notification Message in response to a Label Request Message.

The encoding of the notification message is as follows:

```
 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
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|0|   Notification (0x0001)     |      Message Length           |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                     Message ID                                |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                     Status (TLV)                              |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                     Optional Parameters                       |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

3.5 Release , Withdraw, and Abort Messages

The Label Release , Label Withdraw, and Label Abort Request Messages are used as specified in [1]. These messages may also carry the LSPID TLV.

4. Protocol Specification

The Label Request Message defined in [1] optionally carries one or more of the optional Constraint-based Routing TLVs (CR-TLVs) defined in this section. If needed, other constraints can be supported later through the definition of new TLVs. In this specification, the following TLVs are defined:

- Explicit Route TLV
- Explicit Route Hop TLV
- Traffic Parameters TLV
- Preemption TLV
- LSPID TLV
- Route Pinning TLV
- Resource Class TLV
- CR-LSP FEC TLV

4.1 Explicit Route TLV (ER-TLV)

The ER-TLV is an object that specifies the path to be taken by the LSP being established. It is composed of one or more Explicit Route Hop TLVs (ER-Hop TLVs) defined in Section 4.2.

```
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
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|0|0|         ER-TLV  (0x0800)  |      Length                   |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                          ER-Hop TLV 1                         |
```

```
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                          ER-Hop TLV 2                         |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
~                          ............                         ~
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                          ER-Hop TLV n                         |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Type
A two-byte field carrying the value of the ER-TLV type whichis 0x800.

Length
Specifies the length of the value field in bytes.

ER-Hop TLVs
One or more ER-Hop TLVs defined in Section 4.2.

4.2 Explicit Route Hop TLV (ER-Hop TLV)

The contents of an ER-TLV are a series of variable length ER-Hop TLVs.

A node receiving a label request message including an ER-Hop type that is not supported should not progress the label request message to the downstream LSR and should send back a _No Route_ Notification Message.

Each ER-Hop TLV has the form:

```
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
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|0|0|          ER-Hop-Type      |      Length                   |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|L|                                  Content //                 |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

ER-Hop Type

A fourteen-bit field indicating the type of contents of the ER-Hop. Currently defined values are:

```
Value Type
 ----- ------------------------
0x801 IPv4 prefix
0x802 IPv6 prefix
0x803 Autonomous system number
```

0x804 LSPID

Length
Specifies the length of the value field in bytes.

L bit
The L bit in the ER-Hop is a one-bit attribute. If the L bit is set, then the value of the attribute is _loose._ Otherwise, the value of the attribute is _strict._ For brevity, we say that if the value of the ER-Hop attribute is loose then it is a _loose ER-Hop._ Otherwise, it's a _strict ER-Hop._ Further, we say that the abstract node of a strict or loose ER-Hop is a strict or a loose node, respectively. Loose and strict nodes are always interpreted relative to their prior abstract nodes. The path between a strict node and its prior node MUST include only network nodes from the strict node and its prior abstract node.

The path between a loose node and its prior node MAY include other network nodes, which are not part of the strict node or its prior abstract node.

Contents
A variable length field containing a node or abstract node which is one of the consecutive nodes that make up the explicitly routed LSP.

4.3 Traffic Parameters TLV

The following sections describe the CR-LSP Traffic Parameters. The required characteristics of a CR-LSP are expressed by the Traffic Parameter values.

A Traffic Parameters TLV, is used to signal the Traffic Parameter values. The Traffic Parameters are defined in the subsequent sections.

The Traffic Parameters TLV contains a Flags field, a Frequency, a Weight, and the five Traffic Parameters PDR, PBS, CDR, CBS, EBS. The Traffic Parameters TLV is shown below:


```
 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
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|0|0| Traf. Param. TLV  (0x0810)|      Length                   |
```

```
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|     Flags     |    Frequency  |     Reserved  |    Weight     |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                    Peak Data Rate (PDR)                       |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                    Peak Burst Size (PBS)                      |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                    Committed Data Rate (CDR)                  |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                    Committed Burst Size (CBS)                 |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                    Excess Burst Size (EBS)                    |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Type
A fourteen-bit field carrying the value of the ER-TLV type which is 0x810.

Length
Specifies the length of the value field in bytes.

Flags
The Flags field is shown below:

```
+--+--+--+--+--+--+--+--+
| Res |F6|F5|F4|F3|F2|F1|
+--+--+--+--+--+--+--+--+
```

Res - These bits are reserved.
Zero on transmission.
Ignored on receipt.
F1 - Corresponds to the PDR.
F2 - Corresponds to the PBS.
F3 - Corresponds to the CDR.
F4 - Corresponds to the CBS.
F5 - Corresponds to the EBS.
F6 - Corresponds to the Weight.

Each flag Fi is a Negotiable Flag corresponding to a Traffic Parameter. The Negotiable Flag value zero denotes NotNegotiable and value one denotes Negotiable.

Frequency

The Frequency field is coded as an 8 bit unsigned integer with the following code points defined:

0- Unspecified
1- Frequent
2- VeryFrequent
3-255 - Reserved
Reserved - Zero on transmission. Ignored on receipt.

Weight

An 8 bit unsigned integer indicating the weight of the CR-LSP. Valid weight values are from 1 to 255. The value 0 means that weight is not applicable for the CR-LSP.

Traffic Parameters

Each Traffic Parameter is encoded as a 32-bit IEEE single-precision floating-point number. A value of positive infinity is represented as an IEEE single-precision floating-point number with an exponent of all ones (255) and a sign and mantissa of all zeros. The values PDR and CDR are in units of bytes per second. The values PBS, CBS and EBS are in units of bytes.

The value of PDR MUST be greater than or equal to the value of CDR in a correctly encoded Traffic Parameters TLV.

4.3.1 Semantics

4.3.1.1 Frequency

The Frequency specifies at what granularity the CDR allocated to the CR-LSP is made available. The value VeryFrequent means that the available rate should average at least the CDR when measured over any time interval equal to or longer than the shortest packet time at the CDR. The value Frequent means that the available rate should average at least the CDR when measured over any time interval equal to or longer than a small number of shortest packet times at the CDR.

The value Unspecified means that the CDR MAY be provided at any granularity.

4.3.1.2 Peak Rate

The Peak Rate defines the maximum rate at which traffic SHOULD be sent to the CR-LSP. The Peak Rate is useful for the purpose of resource allocation. If resource allocation within the MPLS domain depends on the Peak Rate value then it should be enforced at the ingress to the MPLS domain.

The Peak Rate is defined in terms of the two Traffic Parameters PDR and PBS, see section 4.3.1.5 below.

4.3.1.3 Committed Rate

The Committed Rate defines the rate that the MPLS domain commits to be available to the CR-LSP.

The Committed Rate is defined in terms of the two Traffic Parameters CDR and CBS, see section 4.3.1.6 below.

4.3.1.4 Excess Burst Size

The Excess Burst Size may be used at the edge of an MPLS domain for the purpose of traffic conditioning. The EBS MAY be used to measure the extent by which the traffic sent on a CR-LSP exceeds the committed rate.

The possible traffic conditioning actions, such as passing, marking or dropping, are specific to the MPLS domain.

The Excess Burst Size is defined together with the Committed Rate, see section 4.3.1.6 below.

4.3.1.5 Peak Rate Token Bucket

The Peak Rate of a CR-LSP is specified in terms of a token bucket P with token rate PDR and maximum token bucket size PBS.

The token bucket P is initially (at time 0) full, i.e., the token count Tp(0) = PBS. Thereafter, the token count Tp, if less than PBS, is incremented by one PDR times per second. When a packet of size B bytes arrives at time t, the following happens:

- If Tp(t)-B >= 0, the packet is not in excess of the peak rate and Tp is decremented by B down to the minimum value of 0, else

- the packet is in excess of the peak rate and Tp is not decremented.

Note that according to the above definition, a positive infinite value of either PDR or PBS implies that arriving packets are never in excess of the peak rate.

The actual implementation of an LSR doesn't need to be modeled according to the above formal token bucket specification.

4.3.1.6 Committed Data Rate Token Bucket

The committed rate of a CR-LSP is specified in terms of a token bucket C with rate CDR. The extent by which the offered rate exceeds the committed rate MAY be measured in terms of another token

bucket E, which also operates at rate CDR. The maximum size of the token bucket C is CBS and the maximum size of the token bucket E is EBS.

The token buckets C and E are initially (at time 0) full, i.e., the token count Tc(0) = CBS and the token count Te(0) = EBS. Thereafter, the token counts Tc and Te are updated CDR times per second as follows:

- If Tc is less than CBS, Tc is incremented by one, else
- if Te is less then EBS, Te is incremented by one, else
- neither Tc nor Te is incremented.

When a packet of size B bytes arrives at time t, the following happens:

- If Tc(t)-B >= 0, the packet is not in excess of the Committed Rate and Tc is decremented by B down to the minimum value of 0, else
- if Te(t)-B >= 0, the packet is in excess of the Committed rate but is not in excess of the EBS and Te is decremented by B down to the minimum value of 0, else
- the packet is in excess of both the Committed Rate and the EBS and neither Tc nor Te is decremented.

Note that according to the above specification, a CDR value of positive infinity implies that arriving packets are never in excess of either the Committed Rate or EBS. A positive infinite value of either CBS or EBS implies that the respective limit cannot be exceeded.

The actual implementation of an LSR doesn't need to be modeled according to the above formal specification.

4.3.1.7 Weight

The weight determines the CR-LSP's relative share of the possible excess bandwidth above its committed rate. The definition of _relative share_ is MPLS domain specific.

4.3.2 Procedures

4.3.2.1 Label Request Message

If an LSR receives an incorrectly encoded Traffic Parameters TLV in which the value of PDR is less than the value of CDR then it MUST send a Notification Message including the Status code _Traffic Parameters Unavailable_ to the upstream LSR from which it received the erroneous message.

If a Traffic Parameter is indicated as Negotiable in the Label Request Message by the corresponding Negotiable Flag then an LSR MAY replace the Traffic Parameter value with a smaller value.

If the Weight is indicated as Negotiable in the Label Request Message by the corresponding Negotiable Flag then an LSR may replace the Weight value with a lower value (down to 0).

If, after possible Traffic Parameter negotiation, an LSR can support the CR-LSP Traffic Parameters then the LSR MUST reserve the corresponding resources for the CR-LSP.

If, after possible Traffic Parameter negotiation, an LSR cannot support the CR-LSP Traffic Parameters then the LSR MUST send a Notification Message that contains the _Resource Unavailable_ status code.

4.3.2.2 Label Mapping Message

If an LSR receives an incorrectly encoded Traffic Parameters TLV in which the value of PDR is less than the value of CDR then it MUST send a Label Release message containing the Status code _Traffic Parameters Unavailable_ to the LSR from which it received the erroneous message. In addition, the LSP should send a Notification Message upstream with the status code _Label Request Aborted_.

If the negotiation flag was set in the label request message, the egress LSR MUST include the (possibly negotiated) Traffic Parameters and Weight in the Label Mapping message.

The Traffic Parameters and the Weight in a Label Mapping message MUST be forwarded unchanged.

An LSR SHOULD adjust the resources that it reserved for a CR-LSP when it receives a Label Mapping Message if the Traffic Parameters differ from those in the corresponding Label Request Message.

4.3.2.3 Notification Message

If an LSR receives a Notification Message for a CR-LSP, it SHOULD release any resources that it possibly had reserved for the CR-LSP. In addition, on receiving a Notification Message from a Downstream LSR that is associated with a Label Request from an upstream LSR, the local LSR MUST propagate the Notification message using the procedures in [1].

4.4 Preemption TLV

The defualt value of the setup and holding priorities should be in the middle of the range (e.g., 4) so that this feature can be turned on gradually in an operational network by increasing or decerasing the priority starting at the middle of the range.


```
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
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|0|0| Preemption-TLV  (0x0820)  |      Length                   |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|  SetPrio      | HoldPrio      |      Reserved                 |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Type
A fourteen-bit field carrying the value of the Preemption-TLV type which is 0x820.

Length
Specifies the length of the value field in bytes.

Reserved
Zero on transmission. Ignored on receipt.

SetPrio
A SetupPriority of value zero (0) is the priority assigned to the most important path. It is referred to as the highest priority. Seven (7) is the priority for the least important path. The higher the setup priority, the more paths CR-LDP can bump to set up the path. The default value should be 4.

HoldPrio
A HoldingPriority of value zero (0) is the priority assigned to the most important path. It is referred to as the highest priority. Seven (7) is the priority for the least important path. The default value should be 4.
The higher the holding priority, the less likely it is for CR-LDP to reallocate its bandwidth to a new path.

4.5 LSPID TLV

LSPID is a unique identifier of a CR-LSP within an MPLS network.

The LSPID is composed of the ingress LSR Router ID (or any of its own Ipv4 addresses) and a Locally unique CR-LSP ID to that LSR.

The LSPID is useful in network management, in CR-LSP repair, and in using an already established CR-LSP as a hop in an ER-TLV.

An _action indicator flag_ is carried in the LSPID TLV. This _action indicator flag_ indicates explicitly the action that should be taken if the LSP already exists on the LSR receiving the message.

After a CR-LSP is set up, its bandwidth reservation may need to be changed by the network operator, due to the new requirements for the traffic carried on that CR-LSP. The _action indicator flag_ is used

indicate the need to modify the bandwidth and possibly other parameters of an established CR-LSP without service interruption. This feature has application in dynamic network resources management where traffic of different priorities and service classes is involved.

The procedure for the code point _modify_ is defined in Appendix C. The procedures for other flags are FFS.

```
 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
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|0|0|      LSPID-TLV  (0x0821)  |      Length                   |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|       Reserved        |ActFlg |      Local CR-LSP ID           |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                       Ingress LSR Router ID                   |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Type
     A fourteen-bit field carrying the value of the LSPID-TLV type which is 0x821.

Length
     Specifies the length of the value field in bytes.

ActFlg
     Action Indicator Flag: A 4-bit field that indicates explicitly the action that should be taken if the LSP already exists on the LSR receiving the message. A set of indicator code points is proposed as follows:

        0000: indicates initial LSP setup
        0001: indicates modify LSP

Reserved
     Zero on transmission. Ignored on receipt.

Local CR-LSP ID
     The Local LSP ID is an identifier of the CR-LSP locally unique within the Ingress LSR originating the CR-LSP.

Ingress LSR Router ID
     An LSR may use any of its own IPv4 addresses in this field.

4.6 Resource Class (Color) TLV

The Resource Class as defined in [4] is used to specify which links are acceptable by this CR-LSP. This information allows for the network's topology to be pruned.


```
 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
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|0|0|      ResCls-TLV  (0x0822) |      Length                   |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                             RsCls                             |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Type
     A fourteen-bit field carrying the value of the ResCls-TLV type which is 0x822.

Length
     Specifies the length of the value field in bytes.

RsCls
     The Resource Class bit mask indicating which of the 32 _administrative groups_ or _colors_ of links the CR-LSP can traverse.

4.7 ER-Hop semantics

4.7.1. ER-Hop 1: The IPv4 prefix

The abstract node represented by this ER-Hop is the set of nodes, which have an IP address, which lies within this prefix. Note that a prefix length of 32 indicates a single IPv4 node.

```
   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
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |0|0|         0x801             |      Length                   |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |L|      Reserved                               |    PreLen     |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |                    IPv4 Address (4 bytes)                     |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Type
     IPv4 Address 0x801

Length
     Specifies the length of the value field in bytes.

L Bit
     Set to indicate Loose hop.
     Cleared to indicate a strict hop.

Reserved
     Zero on transmission. Ignored on receipt.

PreLen

     Prefix Length 1-32

IP Address
     A four-byte field indicating the IP Address.

4.7.2. ER-Hop 2: The IPv6 address

```
   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
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |0|0|          0x802            |      Length                   |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |L|             Reserved                        |    PreLen     |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |                  IPV6 address                                 |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |                  IPV6 address (continued)                     |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |                  IPV6 address (continued)                     |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |                  IPV6 address (continued)                     |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Type

0x802 IPv6 address

Length
Specifies the length of the value field in bytes.

L Bit
Set to indicate Loose hop.
Cleared to indicate a strict hop.

Reserved
Zero on transmission. Ignored on receipt.

PreLen
Prefix Length 1-128

IPv6 address
A 128-bit unicast host address.

4.7.3. ER-Hop 3: The autonomous system number

The abstract node represented by this ER-Hop is the set of nodes belonging to the autonomous system.


```
 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
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|0|0|          0x803            |      Length                   |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|L|          Reserved           |                AS Number      |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Type
AS Number 0x803

Length
Specifies the length of the value field in bytes.

L Bit
Set to indicate Loose hop.
Cleared to indicate a strict hop.

Reserved
Zero on transmission. Ignored on receipt.

AS Number
    Autonomous System number

4.7.4. ER-Hop 4: LSPID

The LSPID is used to identify the tunnel ingress point as the next hop in the ER. This ER-Hop allows for stacking new CR-LSPs within an already established CR-LSP. It also allows for splicing the CR-LSP being established with an existing CR-LSP.

If an LSPID Hop is the last ER-Hop in an ER-TLV, than the LSR may splice the CR-LSP of the incoming Label Request to the CR-LSP that currently exists with this LSPID. This is useful, for example, at the point at which a Label Request used for local repair arrives at the next ER-Hop after the loosely specified CR-LSP segment. Use of the LSPID Hop in this scenario eliminates the need for ER-Hops to keep the entire remaining ER-TLV at each LSR that is at either (upstream or downstream) end of a loosely specified CR-LSP segment as part of its state information. This is due to the fact that the upstream LSR needs only to keep the next ER-Hop and the LSPID and the downstream LSR needs only to keep the LSPID in order for each end to be able to recognize that the same LSP is being identified.

If the LSPID Hop is not the last hop in an ER-TLV, the LSR must forward the remaining ER-TLV in a Label Request message, using the CR-LSP specified by the LSPID, to the LSR that is the CR-LSP's egress. That LSR will continue processing of the CR-LSP Label Request Message. The result is a tunneled, or stacked, CR-LSP.


```
 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
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|0|0|          0x804            |      Length                   |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|L|          Reserved           |               Local LSPID     |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                       Ingress LSR Router ID                   |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Type
    LSPID 0x804

Length
    Specifies the length of the value field in bytes.

L Bit
Set to indicate Loose hop.
Cleared to indicate a strict hop.

Reserved
Zero on transmission. Ignored on receipt.

Local LSPID
A 2 byte field indicating the LSPID which is unique with reference to its Ingress LSR.

Ingress LSR Router ID
An LSR may use any of its own IPv4 addresses in this field.

4.8. Processing of the Explicit Route TLV

4.8.1. Selection of the next hop

A Label Request Message containing an explicit route TLV must determine the next hop for this path. Selection of this next hop may involve a selection from a set of possible alternatives. The mechanism for making a selection from this set is implementation dependent and is outside of the scope of this specification. Selection of particular paths is also outside of the scope of this specification, but it is assumed that each node will make a best effort attempt to determine a loop-free path. Note that such best efforts may be overridden by local policy.

To determine the next hop for the path, a node performs the following steps:

1. The node receiving the Label Request Message must first evaluate the first ER-Hop. If the L bit is not set in the first ER-Hop and if the node is not part of the abstract node described by the first ER-Hop, it has received the message in

error, and should return a _Bad Initial ER-Hop_ error. If the L bit is set and the local node is not part of the abstract node described by the first ER-Hop, the node selects a next hop that is along the path to the abstract node described by the first ER-Hop. If there is no first ER-Hop, the message is also in error and the system should return a _Bad Explicit Routing TLV_ error using a Notification Message sent upstream.

2. If there is no second ER-Hop, this indicates the end of the explicit route. The explicit route TLV should be removed from the Label Request Message. This node may or may not be the end of the LSP. Processing continues with section 4.8.2,

where a new explicit route TLV may be added to the Label Request Message.

3. If the node is also a part of the abstract node described by the second ER-Hop, then the node deletes the first ER-Hop and continues processing with step 2, above. Note that this makes the second ER-Hop into the first ER-Hop of the next iteration.

4. The node determines if it is topologically adjacent to the abstract node described by the second ER-Hop. If so, the node selects a particular next hop which is a member of the abstract node. The node then deletes the first ER-Hop and continues processing with section 4.8.2.

5. Next, the node selects a next hop within the abstract node of the first ER-Hop that is along the path to the abstract node of the second ER-Hop. If no such path exists then there are two cases:

   5.a If the second ER-Hop is a strict ER-Hop, then there is an error and the node should return a _Bad Strict Node_ error.

   5.b Otherwise, if the second ER-Hop is a loose ER-Hop, then the node selects any next hop that is along the path to the next abstract node. If no path exists within the MPLS domain, then there is an error, and the node should return a _Bad loose node_ error.

6. Finally, the node replaces the first ER-Hop with any ER-Hop that denotes an abstract node containing the next hop. This is necessary so that when the explicit route is received by the next hop, it will be accepted.

7. Progress the Label Request Message to the next hop.

4.8.2. Adding ER-Hops to the explicit route TLV

After selecting a next hop, the node may alter the explicit route in the following ways.

If, as part of executing the algorithm in section 4.8.1, the explicit route TLV is removed, the node may add a new explicit route TLV.

Otherwise, if the node is a member of the abstract node for the first ER-Hop, then a series of ER-Hops may be inserted before the first ER-Hop or may replace the first ER-Hop. Each ER-Hop in this series must denote an abstract node that is a subset of the current abstract node.

Alternately, if the first ER-Hop is a loose ER-Hop, an arbitrary series of ER-Hops may be inserted prior to the first ER-Hop.

4.9 Route Pinning TLV

Section 2.4 describes the use of route pinning. The encoding of the Route Pinning TLV is as follows:

```
 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
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|0|0|          0x823            |      Length                   |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|P|                        Reserved                             |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Type
Pinning-TLV type 0x823

Length
Specifies the length of the value field in bytes.

P Bit
The P bit is set to 1 to indicate that route pinning is requested.
The P bit is set to 0 to indicate that route pinning is not requested

Reserved
Zero on transmission. Ignored on receipt.

4.10 CR-LSP FEC Element

A new FEC element is introduced in this specification to support CR-LSPs. This new FEC element does not preclude the use of other FECs elements (Type=0x01, 0x02, 0x03) defined in the LDP spec in CR-LDP messages. The CR-LDP FEC Element is an opaque FEC to be used only in Messages of CR-LSPs.

FEC Element Type name	Type	Value
CR-LSP	0x04	No value; i.e., 0 value octets;

The CR-LSP FEC TLV encoding is as follows:

```
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
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|0|0|         FEC(0x0100)      |      Length                   |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
| CR-LSP (4)    |
+-+-+-+-+-+-+-+-+
```

Type
    FEC TLV type 0x0100

Length
    Specifies the length of the value field in bytes.

CR-LSP FEC Element Type
    0x04

4.11 Error subcodes

In the processing described above, certain errors need to be reported as part of the Notification Message. This section defines the status codes for the errors described in this specification.

Status Code	Type
Bad Explicit Routing TLV Error	0x44000001
Bad Strict Node Error	0x44000002
Bad Loose Node Error	0x44000003
Bad Initial ER-Hop Error	0x44000004
Resource Unavailable	0x44000005
Traffic Parameters Unavailable	0x44000006
Setup Abort (Label Request Aborted in [1])	0x04000015
Modify Request Not Supported	0x44000008

5. Security

Pre-emption has to be controlled by the MPLS domain.

Resource reservation requires the LSRs to have an LSP admission control function.

Traffic Engineered LSPs can bypass normal routing.

6. Acknowledgments

The messages used to signal the CR-LSP setup are based on the work done by the [1] team.

The authors would also like to acknowledge the careful review and comments of Ken Hayward, Greg Wright, Geetha Brown, Brian Williams, Paul Beaubien, Matthew Yuen, Liam Casey, Ankur Anand, Adrian Farrel.

7. Intellectual Property Consideration

The IETF has been notified of intellectual property rights claimed in regard to some or all of the specification contained in this document. For more information consult the online list of claimed rights.

8. References

1 Andersson et al, "Label Distribution Protocol Specification" work in progress (draft-ietf-mpls-ldp-05), June 1999.

2 Callon et al, "Framework for Multiprotocol Label Switching", work in progress (draft-ietf-mpls-framework-05), September 1999.

3 Rosen et al, "Multiprotocol Label Switching Architecture", work in progress (draft-ietf-mpls-arch-06), September 1999.

4 Awduche et al, "Requirements for Traffic Engineering Over MPLS", RFC 2702, September 1999.

5 L. Wu, et. al., "LDP State Machine" work in progress (draft-ietf-mpls-ldp-state-00), Feb 1999.

9. Author's Addresses

Osama S. Aboul-Magd
Nortel Networks
P O Box 3511 Station C
Ottawa, ON K1Y 4H7
Canada
Phone: +1 613 763-5827
Osama@nortelnetworks.com

Loa Andersson
Nortel Networks
S:t Eriksgatan 115
PO Box 6701
113 85 Stockholm
Tel: +46 8 508 835 00
Fax: +46 8 508 835 01
Loa_andersson@nortelnetworks.com

Peter Ashwood-Smith
Nortel Networks
P O Box 3511 Station C
Ottawa, ON K1Y 4H7
Canada
Phone: +1 613 763-4534
Petera@nortelnetworks.com

Ross Callon
IronBridge Networks
55 Hayden Avenue,
Lexington, MA 02173
Phone: +1-781-402-8017
Rcallon@ironbridgenetworks.com

Ram Dantu
Alcatel USA Inc.

Paul Doolan
Ennovate Networks

IP Competence Center
1201 E. Campbell Road.,446-315
Richadson, TX USA., 75081-2206
Phone: 972 996 2938
Fax: 972 996 5902
Ram.dantu@alcatel.com

330 Codman Hill Rd
Marlborough MA 01719
Phone: 978-263-2002
Pdoolan@ennovatenetworks.com

Nancy Feldman
IBM Corp.
17 Skyline Drive
Hawthorne NY 10532
Phone: 914-784-3254
Nkf@us.ibm.com

Andre Fredette
Nortel Networks
600 Technology Park Drive
Billerica, MA 01821
978-288-8524
Fredette@nortelnetworks.com

Eric Gray
Lucent Technologies, Inc
1600 Osgood St.
North Andover, MA 01847
Phone: 603-659-3386
Ewgray@lucent.com

Joel M. Halpern
Institutional Venture Partners
650-926-5633
Joel@mcquillan.com

Juha Heinanen
Telia Finland, Inc.
Myyrmaentie 2
01600 VANTAA
Finland
Tel: +358 41 500 4808
Jh@telia.fi

Fiffi Hellstrand
Ericsson Telecom AB
S-126 25 STOCKHOLM
Sweden
Tel: +46 8 719 4933
Etxfiff@etxb.ericsson.se

Bilel Jamoussi
Nortel Networks Corp.
600 Technology Park Drive
Billerica, MA 01821
USA
Phone: +1 978 288-4506
Jamoussi@nortelnetworks.com

Timothy E. Kilty
Northchurch Communications
5 Corporate Drive,
Andover, MA 018110
phone: 978 691-4656
Tkilty@northc.com

Andrew G. Malis
Ascend Communications, Inc.
1 Robbins Road
Westford, MA 01886
Phone: 978 952-7414
fax: 978 392-2074
Malis@ascend.com

Muckai K Girish
SBC Technology Resources,
4698 Willow Road
Pleasanton, CA 94588
Phone: (925) 598-1263
Fax: (925) 598-1321
Mgirish@tri.sbc.com

Kenneth Sundell
Nortel Networks

Pasi Vaananen
Nokia Telecommunications

S:t Eriksgatan 115
PO Box 6701
113 85 Stockholm
Tel: +46 8 508 835 00
Fax: +46 8 508 835 01

3 Burlington Woods Drive,
Burlington, MA 01803
Phone: +1-781-238-4981
Pasi.vaanenen@ntc.nokia.com

Ksundell@nortelnetworks.com

Tom Worster
Nokia
3 Burlington Woods Dr.
Suite 250
Burlington MA 01803 USA
+1 617 247 2624
Tom.worster@nokia.com

Liwen Wu
Alcatel U.S.A
44983 Knoll Square
Ashburn, Va. 20147
Phone: (703) 724-2619
FAX: (703) 724-2005
Liwen.wu@and.alcatel.com

Appendix A: CR-LSP Establishment Examples

A.1 Strict Explicit Route Example

This appendix provides an example for the setup of a strictly routed CR-LSP. In this example, a specific node represents each abstract node.

The sample network used here is a four node network with two edge LSRs and two core LSRs as follows:

```
abc
LSR1------LSR2------LSR3------LSR4
```

LSR1 generates a Label Request Message as described in Section 3.1 of this draft and sends it to LSR2. This message includes the CR-TLV.

A vector of three ER-Hop TLVs <a, b, c> composes the ER-TLV. The ER-Hop TLVs used in this example are of type 0x0801 (IPv4 prefix) with a prefix length of 32. Hence, each ER-Hop TLV identifies a specific node as opposed to a group of nodes. At LSR2, the following processing of the ER-TLV per Section 4.8.1 of this draft takes place:

1) The node LSR2 is part of the abstract node described by the first hop <a>. Therefore, the first step passes the test. Go to step 2.

2) There is a second ER-Hop, <b>. Go to step 3.

3) LSR2 is not part of the abstract node described by the second ER-Hop <b>. Go to Step 4.

4) LSR2 determines that it is topologically adjacent to the abstract node described by the second ER-Hop <b>. LSR2 selects a next hop (LSR3) which is the abstract node. LSR2 deletes the first ER-Hop <a> from the ER-TLV, which now becomes <b, c>. Processing continues with Section 4.8.2.

At LSR2, the following processing of Section 4.8.2 takes place:

Executing algorithm 4.8.1 did not result in the removal of the ER-TLV.

Also, LSR2 is not a member of the abstract node described by the first ER-Hop <b>.

Finally, the first ER-Hop <b> is a strict hop.

Therefore, processing section 4.8.2 does not result in the insertion of new ER-Hops. The selection of the next hop has been already done is step 4 of Section 4.8.1 and the processing of the ER-TLV is

completed at LSR2. In this case, the Label Request Message including the ER-TLV <b, c> is progressed by LSR2 to LSR3.

At LSR3, a similar processing to the ER-TLV takes place except that the incoming ER-TLV = <b, c> and the outgoing ER-TLV is <c>.

At LSR4, the following processing of section 4.8.1 takes place:

1) The node LSR4 is part of the abstract node described by the first hop <c>. Therefore, the first step passes the test. Go to step 2.
2) There is no second ER-Hop, this indicates the end of the CR-LSP. The ER-TLV is removed from the Label Request Message. Processing continues with Section 4.8.2.

At LSR4, the following processing of Section 4.8.2 takes place: Executing algorithm 4.8.1 resulted in the removal of the ER-TLV. LSR4 does not add a new ER-TLV.

Therefore, processing section 4.8.2 does not result in the insertion of new ER-Hops. This indicates the end of the CR-LSP and the processing of the ER-TLV is completed at LSR4.

At LSR4, processing of Section 3.2 is invoked. The first condition is satisfied (LSR4 is the egress end of the CR-LSP and upstream mapping has been requested). Therefore, a Label Mapping Message is generated by LSR4 and sent to LSR3.

At LSR3, the processing of Section 3.2 is invoked. The second condition is satisfied (LSR3 received a mapping from its downstream next hop LSR4 for a CR-LSP for which an upstream request is still pending). Therefore, a Label Mapping Message is generated by LSR3 and sent to LSR2.

At LSR2, a similar processing to LSR 3 takes place and a Label Mapping Message is sent back to LSR1, which completes the end-to-end CR-LSP setup.

A.2 Node Groups and Specific Nodes Example

A request at ingress LSR to setup a CR-LSP might originate from a management system or an application, the details are implementation specific.

The ingress LSR uses information provided by the management system or the application and possibly also information from the routing database to calculate the explicit route and to create the Label Request Message.

The Label request message carries together with other necessary information an ER-TLV defining the explicitly routed path. In our example the list of hops in the ER-Hop TLV is supposed to contain an

abstract node representing a group of nodes, an abstract node representing a specific node, another abstract node representing a group of nodes, and an abstract node representing a specific egress point.

In--{Group 1}--{Specific A}--{Group 2}--{Specific Out: B}
The ER-TLV contains four ER-Hop TLVs:

1. An ER-Hop TLV that specifies a group of LSR valid for the first abstract node representing a group of nodes (Group 1).

2. An ER-Hop TLV that indicates the specific node (Node A).

3. An ER-Hop TLV that specifies a group of LSRs valid for the second abstract node representing a group of nodes (Group 2).

4. An ER-Hop TLV that indicates the specific egress point for the CR-LSP (Node B).

All the ER-Hop TLVs are strictly routed nodes.
The setup procedure for this CR-LSP works as follows:

1. The ingress node sends the Label Request Message to a node that is a member the group of nodes indicated in the first ER-Hop TLV, following normal routing for the specific node (A).

2. The node that receives the message identifies itself as part of the group indicated in the first ER-Hop TLV, and that it is not the specific node (A) in the second. Further it realizes that the specific node (A) is not one of its next

hops.

3. It keeps the ER-Hop TLVs intact and sends a Label Request Message to another node that is part of the group indicated in the first ER-Hop TLV (Group 1), following normal routing for the specific node (A).

4. The node that receives the message identifies itself as part of the group indicated in the first ER-Hop TLV, and that it is not the specific node (A) in the second ER-Hop TLV. Further it realizes that the specific node (A) is one of its next hops.

5. It removes the first ER-Hop TLVs and sends a Label Request Message to the specific node (A).

6. The specific node (A) recognizes itself in the first ER-Hop TLV. Removes the specific ER-Hop TLV.

7. It sends a Label Request Message to a node that is a member of the group (Group 2) indicated in the ER-Hop TLV.

8. The node that receives the message identifies itself as part of the group indicated in the first ER-Hop TLV, further it realizes that the specific egress node (B) is one of its next hops.

9. It sends a Label Request Message to the specific egress node (B).

10. The specific egress node (B) recognizes itself as the egress for the CR-LSP, it returns a Label Mapping Message, that will traverse the same path as the Label Request Message in the opposite direction.

Appendix B. QoS Service Examples

B.1 Service Examples

Construction of an end-to-end service is the result of the rules enforced at the edge and the treatment that packets receive at the network nodes. The rules define the traffic conditioning actions that are implemented at the edge and they include policing with pass, mark, and drop capabilities. The edge rules are expected tobe defined by the mutual agreements between the service providers and their customers and they will constitute an essential part of the SLA. Therefore edge rules are not included in the signaling protocol.

Packet treatment at a network node is usually referred to as the local behavior. Local behavior could be specified in many ways. One example for local behavior specification is the service frequency introduced in section 4.3.2.1, together with the resource reservation rules implemented at the nodes.

Edge rules and local behaviors can be viewed as the main building blocks for the end-to-end service construction. The following table illustrates the applicability of the building block approach for constructing different services including those defined for ATM.

Service Examples	PDR	PBS	CDR	CBS	EBS	Service Frequency	Conditioning Action
DS	S	S	=PDR	=PBS	0	Frequent	drop>PDR
TS	S	S	S	S	0	Unspecified	drop>PDR,PBS mark>CDR,CBS
BE	inf	inf	inf	inf	0	Unspecified	-
FRS	S	S	CIR	~B_C	~B_E	Unspecified	drop>PDR,PBS mark>CDR,CBS,EBS
ATM-CBR	PCR	CDVT	=PCR	=CDVT	0	VeryFrequent	drop>PCR
ATM-VBR.3(rt)	PCR	CDVT	SCR	MBS	0	Frequent	drop>PCR mark>SCR,MBS
ATM-VBR.3(nrt)	PCR	CDVT	SCR	MBS	0	Unspecified	drop>PCR mark>SCR,MBS
ATM-UBR	PCR	CDVT	-	-	0	Unspecified	drop>PCR
ATM-GFR.1	PCR	CDVT	MCR	MBS	0	Unspecified	drop>PCR
ATM-GFR.2	PCR	CDVT	MCR	MBS	0	Unspecified	drop>PCR mark>MCR,MFS
int-serv-CL	p	m	r	b	0	Frequent	drop>p drop>r,b

S= User specified

In the above table, the DS refers to a delay sensitive service where the network commits to deliver with high probability user datagrams at a rate of PDR with minimum delay and delay requirements. Datagrams in excess of PDR will be discarded.

The TS refers to a generic throughput sensitive service where the network commits to deliver with high probability user datagrams at a rate of at least CDR. The user may transmit at a rate higher than CDR but datagrams in excess of CDR would have a lower probability of being delivered.

The BE is the best effort service and it implies that there are no expected service guarantees from the network.

B.2 Establishing CR-LSP Supporting Real-Time Applications

In this scenario the customer needs to establish an LSP for supporting real-time applications such as voice and video. The Delay-sensitive (DS) service is requested in this case.

The first step is the specification of the traffic parameters in the signaling message. The two parameters of interest to the DS service are the PDR and the PBS and the user based on his requirements specifies their values. Since all the traffic parameters are included in the signaling message, appropriate values must be assigned to all of them. For DS service, the CDR and the CBS values are set equal to the PDR and the PBS respectively. An indication of whether the parameter values are subject to negotiation is flagged.

The transport characteristics of the DS service require Frequent frequency to be requested to reflect the real-time delay requirements of the service.

In addition to the transport characteristics, both the network provider and the customer need to agree on the actions enforced at the edge. The specification of those actions is expected to be a part of the service level agreement (SLA) negotiation and is not included in the signaling protocol. For DS service, the edge action is to drop packets that exceed the PDR and the PBS specifications. The signaling message will be sent in the direction of the ER path and the LSP is established following the normal LDP procedures. Each LSR applies its admission control rules. If sufficient resources are not available and the parameter values are subject to negotiation, then the LSR could negotiate down the PDR, the PBS, or both.

The new parameter values are echoed back in the Label Mapping Message. LSRs might need to re-adjust their resource reservations based on the new traffic parameter values.

B.3 Establishing CR-LSP Supporting Delay Insensitive Applications

In this example we assume that a throughput sensitive (TS) service is requested. For resource allocation the user assigns values for PDR, PBS, CDR, and CBS. The negotiation flag is set if the traffic parameters are subject to negotiation.
Since the service is delay insensitive by definition, the Unspecified frequency is signaled to indicate that the service frequency is not an issue.

Similar to the previous example, the edge actions are not subject for signaling and are specified in the service level agreement between the user and the network provider.

For TS service, the edge rules might include marking to indicate high discard precedence values for all packets that exceed CDR and the CBS. The edge rules will also include dropping of packets that conform to neither PDR nor PBS.

Each LSR of the LSP is expected to run its admission control rules and negotiate traffic parameters down if sufficient resources do not exist. The new parameter values are echoed back in the Label Mapping Message. LSRs might need to re-adjust their resources based on the new traffic parameter values.

Appendix C. LSP Modification Using CR-LDP

After a CR-LSP is set up, its bandwidth reservation may need to be changed by the network operator, due to the new requirements for the traffic carried on that CR-LSP. This contribution presents an approach to modify the bandwidth and possibly other parameters of an established CR-LSP using CR-LDP without service interruption. The LSP modification feature can be supported by CR-LDP with a minor extension of an _action indicator flag_. This feature has application in dynamic network resources management where traffic of different priorities and service classes is involved.

Conventions used in this Appendix:

L: LSP (Label Switched Path)
Lid: LSPID (LSP Identifier)
T: Traffic Parameters
R: LSR (Label Switching Router)
FTN: FEC To NHLFE
FEC: Forwarding Equivalence Class
NHLFE: Next Hop Label Forwarding Entity
TLV: Type Length Value

C.1 Introduction

Consider an LSP L1 that has been established with its set of traffic parameters T0. A certain amount of bandwidth is reserved along the path of L1. Consider then that some changes are required on L1. For example, the bandwidth of L1 needs to be increased to accommodate the increased traffic on L1. Or the SLA associated with L1 needs to be modified because a different service class is desired. The network operator, in these cases, would like to modify the characteristics of L1, for example, to change its traffic parameter set from T0 to T1, without releasing the LSP L1 to interrupt the service. In some other cases, network operators may want to reroute a CR-LSP to a different path for either improved performance or better network resource utilization. In all these cases, LSP modification is required. In section C.2 below, a method to modify an active LSP using CR-LDP is presented. The concept of LSPID in CR-LDP is used to achieve the LSP modification, without releasing the LSP and interrupting the service and, without double booking the bandwidth. Only a minimum extension on CR-LDP, an action indication flag of _modify_ is needed in order to explicitly specify the behavior, and allow the existing LSPID to support other networking capabilities in the future. Section 4.5 specifies the action indication flag of _modify_ for CR-LDP. An example is described to demonstrate an application of the presented method in dynamically managing network bandwidth requirements without interrupting service.

C.2 Basic Procedure

LSP modification can only be allowed when the LSP is already set up and active. That is, modification is not defined nor allowed during the LSP establishment or label release/withdraw phases. Only modification requested by the ingress LSR of the LSP is considered in this draft for CR-LSP. Ingress LSR cannot modify an LSP before a previous modification procedure is completed.

Assume that CR-LSP L1 is set up with LSPID L-id1, which is unique in the MPLS network. The ingress LSR R1 of L1 has in its FTN (FEC To NHLFE) table FEC1 -> Label A mapping where A is the outgoing label for LSP L1. To modify the characteristics of L1, R1 sends a Label Request Message. In the messages, the TLVs will have the new requested values, and the LSPID TLV is included which indicates the value of L-id1. The Traffic Parameters TLV, the ER-TLV, the Resource Class (color) TLV and the Preemption TLV can have values different from those in the original Label Request Message, which has been used to set up L1 earlier. Thus, L1 can be changed in its bandwidth request (traffic parameter TLV), its traffic service class (traffic parameter TLV), the route it traverses (ER TLV) and its setup and holding (Preemption TLV) priorities. The ingress LSR R1 now still has the entry in FTN as FEC1 -> Label A. R1 is waiting to establish another entry for FEC1.

When an LSR Ri along the path of L1 receives the Label Request message, its behavior is the same as that of receiving any Label request message. The only extension is that Ri examines the LSPID carried in the Label Request Message, L-id1 and identifies if it already has L-id1. If Ri does not have L-id1, Ri behaves the same as receiving a new Label Request message. If Ri already has L-id1, Ri takes the newly received Traffic Parameter TLV and computes the new bandwidth required and derives the new service class. Compared with the already reserved bandwidth for L-id1, Ri now reserves only the difference of the bandwidth requirements. This prevents Ri from doing bandwidth double booking. If a new service class is requested, Ri also prepares to receive the traffic on L1 in, perhaps a different type of queue, just the same as handling it for a Label Request Message. Ri assigns a new label for the Label Request Message.

When the Label Mapping message is received, two sets of labels exist for the same LSPID. Then the ingress LSR R1 will have two outgoing labels, A and B, associated with the same FEC, where B is the new outgoing label received for LSP L1. The ingress LSR R1 can now activate the new entry in FTN, FEC1 - > Label B. This means that R1 swaps traffic on L1 to the new label _B_ (_new_ path) for L1. The packets can now be sent with the new label B, with the new set of traffic parameters if any, on a new path, that is, if a new path is requested in the Label Request Message for the modification. All the other LSRs along the path will start to receive the incoming packets with the new label. For the incoming new label, the LSR has already

established its mapping to the new outgoing label. Thus, the packets will be sent out with the new outgoing label. The LSRs do not have to implement new procedures to track the new and old characteristics of the LSP.

The ingress LSR R1 then starts to release the original label A for LSP L1. The Label Release Message is sent by R1 towards the down stream LSRs. The Release message carries the LSPID of L-id1 and the Label TLV to indicate which label is to be released. The Release Message is propagated to the egress LSR to release the original labels previously used for L1. Upon receiving the Label Release Message, LSR R1 examines the LSPID, L-id1 and finds out that the L-id1 has still another set of label (incoming/outgoing) under it. Thus, the old label is released without releasing the resource in use. That is, if the bandwidth has been decreased for L1, the delta bandwidth is released. Otherwise, no bandwidth is released. This modification procedure can not only be applied to modify the traffic parameters and/or service class of an active LSP, but also to reroute an existing LSP, and/or change its setup/holding priority if desired. After the release procedure, the modification of the LSP is completed.

The method described above follows the normal behavior of Label Request / Mapping / Notification / Release /Withdraw procedure of a CR-LDP operated LSR with a specific action taken on LSPID. If Label Withdraw Message is used to withdraw a label associated with an LSPID, the Label TLV should be included to specify which label to withdraw. Since the LSPID can also be used for other feature support, an action indication flag of _modify_ assigned to the LSPID would explicitly explain the action/semantics that should be associated with the messaging procedure. The details of this flag are addressed in Section 4.5.

C.3 Priority Handling

When sending a Label Request Message for an active LSP L1 to request changes, the setup priority used in the label Request Message can be different from the one used in the previous Label Request Message, effectively indicating the priority of this _modification_ request. Network operators can use this feature to decide what priority is to be assigned to a modification request, based on their policies/algorithms and other traffic situations in the network. For example, the priority for modification can be determined by the priority of the customer/LSP. If a customer has exceeded the reserved bandwidth of its VPN LSP tunnel by too much, the modification request's priority may be given higher.
The Label Request message for the modification of an active LSP can also be sent with a holding priority different from its previous one. This effectively changes the holding priority of the LSP. Upon receiving a Label Request Message that requests a new holding priority, the LSR assigns the new holding priority to the bandwidth. That is, the new holding priority is assigned to both the existing

incoming / outgoing labels and the new labels to be established for the LSPID in question. In this way self-bumping is prevented.

C.4 Modification Failure Case Handling

A modification attempt may fail due to insufficient resource or other situations. A Notification message is sent back to the ingress LSR R1 to indicate the failure of Label Request Message that intended to modify the LSP. Retry may be attempted if desired by the network operator.

If the LSP on the original path failed when a modification attempt is in progress, the attempt should be aborted by using the Label Abort Request message as specified in LDP draft.