Network Working Group   F. Le Faucheur, Ed.
Request for Comments: 4124   Cisco Systems, Inc.
Category: Standards Track   June 2005

Protocol Extensions for Support of Diffserv-aware MPLS Traffic Engineering

Status of This Memo

This document specifies an Internet standards track protocol for the Internet community, and requests discussion and suggestions for improvements. Please refer to the current edition of the "Internet Official Protocol Standards" (STD 1) for the standardization state and status of this protocol. Distribution of this memo is unlimited.

Copyright Notice

Abstract

This document specifies the protocol extensions for support of Diffserv-aware MPLS Traffic Engineering (DS-TE). This includes generalization of the semantics of a number of Interior Gateway Protocol (IGP) extensions already defined for existing MPLS Traffic Engineering in RFC 3630, RFC 3784, and additional IGP extensions beyond those. This also includes extensions to RSVP-TE signaling beyond those already specified in RFC 3209 for existing MPLS Traffic Engineering. These extensions address the requirements for DS-TE spelled out in RFC 3564.

Table of Contents

1. Introduction

[DSTE-REQ] presents the Service Provider requirements for support of Differentiated-Service (Diffserv)-aware MPLS Traffic Engineering (DS-TE). This includes the fundamental requirement to be able to enforce different bandwidth constraints for different classes of traffic.

This document specifies the IGP and RSVP-TE signaling extensions (beyond those already specified for existing MPLS Traffic Engineering [OSPF-TE][ISIS-TE][RSVP-TE]) for support of the DS-TE requirements spelled out in [DSTE-REQ] including environments relying on distributed Constraint-Based Routing (e.g., path computation involving head-end Label Switching Routers).

[DSTE-REQ] provides a definition and examples of Bandwidth Constraints models. The present document does not specify nor assume a particular Bandwidth Constraints model. Specific Bandwidth Constraints models are outside the scope of this document. Although the extensions for DS-TE specified in this document may not be sufficient to support all the conceivable Bandwidth Constraints models, they do support the Russian Dolls Model specified in [DSTE-RDM], the Maximum Allocation Model specified in [DSTE-MAM], and the Maximum Allocation with Reservation Model specified in [DSTE-MAR].

There may be differences between the quality of service expressed and obtained with Diffserv without DS-TE and with DS-TE. Because DS-TE uses Constraint-Based Routing, and because of the type of admission control capabilities it adds to Diffserv, DS-TE has capabilities for traffic that Diffserv does not: Diffserv does not indicate preemption, by intent, whereas DS-TE describes multiple levels of preemption for its Class-Types. Also, Diffserv does not support any means of explicitly controlling overbooking, while DS-TE allows this. When considering a complete quality of service environment, with Diffserv routers and DS-TE, it is important to consider these differences carefully.

1.1. Specification of Requirements

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in [RFC2119].

2. Contributing Authors

This document was the collective work of several authors. The text and content were contributed by the editor and the co-authors listed below. (The contact information for the editor appears in the Editor's Address section.)

Jim Boyle
Protocol Driven Networks, Inc.
1381 Kildaire Farm Road #288
Cary, NC 27511, USA

Phone: (919) 852-5160
EMail: jboyle@pdnets.com

Kireeti Kompella
Juniper Networks, Inc.
1194 N. Mathilda Ave.
Sunnyvale, CA 94099

EMail: kireeti@juniper.net

William Townsend
Tenor Networks
100 Nagog Park
Acton, MA 01720

Phone: +1-978-264-4900
EMail: btownsend@tenornetworks.com

Thomas D. Nadeau
Cisco Systems, Inc.
250 Apollo Drive
Chelmsford, MA 01824

Phone: +1-978-244-3051
EMail: tnadeau@cisco.com

Darek Skalecki
Nortel Networks
3500 Carling Ave,
Nepean K2H 8E9

Phone: +1-613-765-2252
EMail: dareks@nortelnetworks.com

3. Definitions

For readability, a number of definitions from [DSTE-REQ] are repeated here:

Traffic Trunk: an aggregation of traffic flows of the same class (i.e., treated equivalently from the DS-TE perspective), which is placed inside a Label Switched Path (LSP).

Class-Type (CT): the set of Traffic Trunks crossing a link that is governed by a specific set of bandwidth constraints. CT is used for the purposes of link bandwidth allocation, constraint-based routing and admission control. A given Traffic Trunk belongs to the same CT on all links.

TE-Class: A pair of:
i. a Class-Type
ii. a preemption priority allowed for that Class-Type. This means that an LSP transporting a Traffic Trunk from that Class-Type can use that preemption priority as the setup priority, the holding priority, or both.

Definitions for a number of MPLS terms are not repeated here. They can be found in [MPLS-ARCH].

4. Configurable Parameters

This section only discusses the differences with the configurable parameters supported for MPLS Traffic Engineering as per [TE-REQ], [ISIS-TE], [OSPF-TE], and [RSVP-TE]. All other parameters are unchanged.

4.1. Link Parameters

4.1.1. Bandwidth Constraints (BCs)

[DSTE-REQ] states that "Regardless of the Bandwidth Constraints Model, the DS-TE solution MUST allow support for up to 8 BCs."

For DS-TE, the existing "Maximum Reservable link bandwidth" parameter is retained, but its semantics is generalized and interpreted as the aggregate bandwidth constraint across all Class-Types, so that, independently of the Bandwidth Constraints Model in use:

SUM (Reserved (CTc)) <= Max Reservable Bandwidth,

where the SUM is across all values of "c" in the range 0 <= c <= 7.

Additionally, on every link, a DS-TE implementation MUST provide for configuration of up to 8 additional link parameters which are the eight potential BCs, i.e., BC0, BC1, ... BC7. The LSR MUST interpret these BCs in accordance with the supported Bandwidth Constraints Model (i.e., what BC applies to what Class-Type, and how).

Where the Bandwidth Constraints Model imposes some relationship among the values to be configured for these BCs, the LSR MUST enforce those at configuration time. For example, when the Russian Dolls Bandwidth Constraints Model ([DSTE-RDM]) is used, the LSR MUST ensure that BCi is configured smaller than or equal to BCj, where i is greater than j, and ensure that BC0 is equal to the Maximum Reservable Bandwidth. As another example, when the Maximum Allocation Model ([DSTE-MAM]) is used, the LSR MUST ensure that all BCi are configured smaller or equal to the Maximum Reservable Bandwidth.

4.1.2. Overbooking

DS-TE enables a network administrator to apply different overbooking (or underbooking) ratios for different CTs.

The principal methods to achieve this are the same as those historically used in existing TE deployment:

(i) To take into account the overbooking/underbooking ratio appropriate for the Ordered Aggregate (OA) or CT associated with the considered LSP at the time of establishing the bandwidth size of a given LSP. We refer to this method as the "LSP Size Overbooking" method. AND/OR

(ii) To take into account the overbooking/underbooking ratio at the time of configuring the Maximum Reservable Bandwidth/BCs and use values that are larger (overbooking) or smaller (underbooking) than those actually supported by the link. We refer to this method as the "Link Size Overbooking" method.

The "LSP Size Overbooking" and "Link Size Overbooking" methods are expected to be sufficient in many DS-TE environments and require no additional configurable parameters. Other overbooking methods may involve such additional configurable parameters, but are beyond the scope of this document.

4.2. LSR Parameters

4.2.1. TE-Class Mapping

In line with [DSTE-REQ], the preemption attributes defined in [TE-REQ] are retained with DS-TE and applicable within, and across, all CTs. The preemption attributes of setup priority and holding priority retain existing semantics, and in particular these semantics are not affected by the LSP CT. This means that if LSP1 contends with LSP2 for resources, LSP1 may preempt LSP2 if LSP1 has a higher setup preemption priority (i.e., lower numerical priority value) than LSP2 holding preemption priority, regardless of LSP1 CT and LSP2 CT.

DS-TE LSRs MUST allow configuration of a TE-Class mapping whereby the Class-Type and preemption level are configured for each of (up to) 8 TE-Classes.

This mapping is referred to as :

```
TE-Class[i]  <-->  < CTc , preemption p >
```

where 0 <= i <= 7, 0 <= c <= 7, 0 <= p <= 7

Two TE-Classes MUST NOT be identical (i.e., have both the same Class-Type and the same preemption priority).

There are no other restrictions on how any of the 8 Class-Types can be paired up with any of the 8 preemption priorities to form a TE-Class. In particular, one given preemption priority can be paired up with two (or more) different Class-Types to form two (or more) TE-Classes. Similarly, one Class-Type can be paired up with two (or more) different preemption priorities to form two (or more) TE-Classes. Also, there is no mandatory ordering relationship between the TE-Class index (i.e., "i" above) and the Class-Type (i.e., "c" above) or the preemption priority (i.e., "p" above) of the TE-Class.

Where the network administrator uses less than 8 TE-Classes, the DS-TE LSR MUST allow remaining ones to be configured as "Unused". Note that configuring all the 8 TE-Classes as "Unused" effectively results in disabling TE/DS-TE since no TE/DS-TE LSP can be established (nor even configured, since as described in Section 4.3.3 below, the CT and preemption priorities configured for an LSP MUST form one of the configured TE-Classes).

To ensure coherent DS-TE operation, the network administrator MUST configure exactly the same TE-Class mapping on all LSRs of the DS-TE domain.

When the TE-Class mapping needs to be modified in the DS-TE domain, care ought to be exercised during the transient period of reconfiguration during which some DS-TE LSRs may be configured with the new TE-Class mapping while others are still configured with the old TE-Class mapping. It is recommended that active tunnels do not use any of the TE-Classes that are being modified during such a transient reconfiguration period.

4.3. LSP Parameters

4.3.1. Class-Type

With DS-TE, LSRs MUST support, for every LSP, an additional configurable parameter that indicates the Class-Type of the Traffic Trunk transported by the LSP.

There is one and only one Class-Type configured per LSP.

The configured Class-Type indicates, in accordance with the supported Bandwidth Constraints Model, the BCs that MUST be enforced for that LSP.

4.3.2. Setup and Holding Preemption Priorities

As per existing TE, DS-TE LSRs MUST allow every DS-TE LSP to be configured with a setup and holding priority, each with a value between 0 and 7.

4.3.3. Class-Type/Preemption Relationship

With DS-TE, the preemption priority configured for the setup priority of a given LSP and the Class-Type configured for that LSP MUST be such that, together, they form one of the (up to) 8 TE-Classes configured in the TE-Class mapping specified in Section 4.2.1 above.

The preemption priority configured for the holding priority of a given LSP and the Class-Type configured for that LSP MUST also be such that, together, they form one of the (up to) 8 TE-Classes configured in the TE-Class mapping specified in Section 4.2.1 above.

The LSR MUST enforce these two rules at configuration time.

4.4. Examples of Parameters Configuration

For illustration purposes, we now present a few examples of how these configurable parameters may be used. All these examples assume that different BCs need to be enforced for different sets of Traffic Trunks (e.g., for Voice and for Data) so that two or more Class-Types need to be used.

4.4.1. Example 1

The network administrator of a first network using two CTs (CT1 for Voice and CT0 for Data) may elect to configure the following TE-Class mapping to ensure that Voice LSPs are never driven away from their shortest path because of Data LSPs:

```
TE-Class[0]  <-->  < CT1 , preemption 0 >
TE-Class[1]  <-->  < CT0 , preemption 1 >
TE-Class[i]  <-->  unused, for 2 <= i <= 7
```

Voice LSPs would then be configured with:

```
CT = CT1, setup priority = 0, holding priority = 0
```

Data LSPs would then be configured with:

```
CT = CT0, setup priority = 1, holding priority = 1
```

A new Voice LSP would then be able to preempt an existing Data LSP in case they contend for resources. A Data LSP would never preempt a Voice LSP. A Voice LSP would never preempt another Voice LSP. A Data LSP would never preempt another Data LSP.

4.4.2. Example 2

The network administrator of another network may elect to configure the following TE-Class mapping in order to optimize global network resource utilization by favoring placement of large LSPs closer to their shortest path:

```
TE-Class[0]  <-->  < CT1 , preemption 0 >
TE-Class[1]  <-->  < CT0 , preemption 1 >
TE-Class[2]  <-->  < CT1 , preemption 2 >
TE-Class[3]  <-->  < CT0 , preemption 3 >
TE-Class[i]  <-->  unused, for 4 <= i <= 7
```

Large-size Voice LSPs could be configured with:

```
CT = CT1, setup priority = 0, holding priority = 0
```

Large-size Data LSPs could be configured with:

```
CT = CT0, setup priority = 1, holding priority = 1
```

Small-size Voice LSPs could be configured with:
CT = CT1, setup priority = 2, holding priority = 2

Small-size Data LSPs could be configured with:
CT = CT0, setup priority = 3, holding priority = 3

A new large-size Voice LSP would then be able to preempt a small-size Voice LSP or any Data LSP in case they contend for resources. A new large-size Data LSP would then be able to preempt a small-size Data LSP or a small-size Voice LSP in case they contend for resources, but it would not be able to preempt a large-size Voice LSP.

4.4.3. Example 3

The network administrator of another network may elect to configure the following TE-Class mapping in order to ensure that Voice LSPs are never driven away from their shortest path because of Data LSPs. This also achieves some optimization of global network resource utilization by favoring placement of large LSPs closer to their shortest path:

```
TE-Class[0]  <-->  < CT1 , preemption 0 >
TE-Class[1]  <-->  < CT1 , preemption 1 >
TE-Class[2]  <-->  < CT0 , preemption 2 >
TE-Class[3]  <-->  < CT0 , preemption 3 >
TE-Class[i]  <-->  unused, for 4 <= i <= 7
```

Large-size Voice LSPs could be configured with:
CT = CT1, setup priority = 0, holding priority = 0.

Small-size Voice LSPs could be configured with:
CT = CT1, setup priority = 1, holding priority = 1.

Large-size Data LSPs could be configured with:
CT = CT0, setup priority = 2, holding priority = 2.

Small-size Data LSPs could be configured with:
CT=CT0, setup priority = 3, holding priority = 3.

A Voice LSP could preempt a Data LSP if they contend for resources. A Data LSP would never preempt a Voice LSP. A large-size Voice LSP could preempt a small-size Voice LSP if they contend for resources. A large-size Data LSP could preempt a small-size Data LSP if they contend for resources.

4.4.4. Example 4

The network administrator of another network may elect to configure the following TE-Class mapping in order to ensure that no preemption occurs in the DS-TE domain:

```
TE-Class[0]  <-->  < CT1 , preemption 0 >
TE-Class[1]  <-->  < CT0 , preemption 0 >
TE-Class[i]  <-->  unused,   for 2 <= i <= 7
```

Voice LSPs would then be configured with:
CT = CT1, setup priority =0, holding priority = 0

Data LSPs would then be configured with:
CT = CT0, setup priority = 0, holding priority = 0

No LSP would then be able to preempt any other LSP.

4.4.5. Example 5

The network administrator of another network may elect to configure the following TE-Class mapping in view of increased network stability through a more limited use of preemption:

```
TE-Class[0]  <-->  < CT1 , preemption 0 >
TE-Class[1]  <-->  < CT1 , preemption 1 >
TE-Class[2]  <-->  < CT0 , preemption 1 >
TE-Class[3]  <-->  < CT0 , preemption 2 >
TE-Class[i]  <-->  unused, for 4 <= i <= 7
```

Large-size Voice LSPs could be configured with: CT = CT1, setup priority = 0, holding priority = 0.

Small-size Voice LSPs could be configured with: CT = CT1, setup priority = 1, holding priority = 0.

Large-size Data LSPs could be configured with: CT = CT0, setup priority = 2, holding priority = 1.

Small-size Data LSPs could be configured with: CT = CT0, setup priority = 2, holding priority = 2.

A new large-size Voice LSP would be able to preempt a Data LSP in case they contend for resources, but it would not be able to preempt any Voice LSP even a small-size Voice LSP.

A new small-size Voice LSP would be able to preempt a small-size Data LSP in case they contend for resources, but it would not be able to preempt a large-size Data LSP or any Voice LSP.

A Data LSP would not be able to preempt any other LSP.

5. IGP Extensions for DS-TE

This section only discusses the differences with the IGP advertisement supported for (aggregate) MPLS Traffic Engineering as per [OSPF-TE] and [ISIS-TE]. The rest of the IGP advertisement is unchanged.

5.1. Bandwidth Constraints

As detailed above in Section 4.1.1, up to 8 BCs (BCb, 0 <= b <= 7) are configurable on any given link.

With DS-TE, the existing "Maximum Reservable Bandwidth" sub-TLV ([OSPF-TE], [ISIS-TE]) is retained with a generalized semantics so that it MUST now be interpreted as the aggregate bandwidth constraint across all Class-Types; i.e., SUM (Reserved (CTc)) <= Max Reservable Bandwidth, independently of the Bandwidth Constraints Model.

This document also defines the following new optional sub-TLV to advertise the eight potential BCs (BC0 to BC7):

"Bandwidth Constraints" sub-TLV:

- Bandwidth Constraints Model Id (1 octet)
- Reserved (3 octets)
- Bandwidth Constraints (N x 4 octets)

Where:

- With OSPF, the sub-TLV is a sub-TLV of the "Link TLV" and its sub-TLV type is 17.

- With ISIS, the sub-TLV is a sub-TLV of the "extended IS reachability TLV" and its sub-TLV type is 22.

- Bandwidth Constraints Model Id: a 1-octet identifier for the Bandwidth Constraints Model currently in use by the LSR initiating the IGP advertisement. See the IANA Considerations section for assignment of values in this name space.

- Reserved: a 3-octet field. This field should be set to zero by the LSR generating the sub-TLV and should be ignored by the LSR receiving the sub-TLV.

- Bandwidth Constraints: contains BC0, BC1,... BC(N-1). Each BC is encoded on 32 bits in IEEE floating point format. The units are bytes (not bits!) per second. Where the configured TE-Class mapping and the Bandwidth Constraints model in use are such that BCh+1, BCh+2, ...and BC7 are not relevant to any of the Class-Types associated with a configured TE-Class, it is RECOMMENDED that only the Bandwidth Constraints from BC0 to BCh be advertised, in order to minimize the impact on IGP scalability.

All relevant generic TLV encoding rules (including TLV format, padding and alignment, as well as IEEE floating point format encoding) defined in [OSPF-TE] and [ISIS-TE] are applicable to this new sub-TLV.

The "Bandwidth Constraints" sub-TLV format is illustrated 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
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
| BC Model Id   |           Reserved                            |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                       BC0 value                               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
//                       . . .                                 //
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                       BCh value                               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

A DS-TE LSR MAY optionally advertise BCs.

A DS-TE LSR, which does advertise BCs, MUST use the new "Bandwidth Constraints" sub-TLV (in addition to the existing Maximum Reservable Bandwidth sub-TLV) to do so. For example, in the case where a service provider deploys DS-TE with TE-Classes associated with CT0 and CT1 only, and where the Bandwidth Constraints Model is such that only BC0 and BC1 are relevant to CT0 and CT1, a DS-TE LSR which does advertise BCs would include in the IGP advertisement the Maximum Reservable Bandwidth sub-TLV, as well as the "Bandwidth Constraints" sub-TLV. The former should contain the aggregate bandwidth constraint across all CTs, and the latter should contain BC0 and BC1.

A DS-TE LSR receiving the "Bandwidth Constraints" sub-TLV with a Bandwidth Constraints Model Id that does not match the Bandwidth Constraints Model it currently uses SHOULD generate a warning to the operator/management system, reporting the inconsistency between Bandwidth Constraints Models used on different links. Also, in that case, if the DS-TE LSR does not support the Bandwidth Constraints

Model designated by the Bandwidth Constraints Model Id, or if the DS-TE LSR does not support operations with multiple simultaneous Bandwidth Constraints Models, the DS-TE LSR MAY discard the corresponding TLV. If the DS-TE LSR does support the Bandwidth Constraints Model designated by the Bandwidth Constraints Model Id, and if the DS-TE LSR does support operations with multiple simultaneous Bandwidth Constraints Models, the DS-TE LSR MAY accept the corresponding TLV and allow operations with different Bandwidth Constraints Models used in different parts of the DS-TE domain.

5.2. Unreserved Bandwidth

With DS-TE, the existing "Unreserved Bandwidth" sub-TLV is retained as the only vehicle to advertise dynamic bandwidth information necessary for Constraint-Based Routing on head-ends, except that it is used with a generalized semantics. The Unreserved Bandwidth sub-TLV still carries eight bandwidth values, but they now correspond to the unreserved bandwidth for each of the TE-Classes (instead of for each preemption priority, as per existing TE).

More precisely, a DS-TE LSR MUST support the Unreserved Bandwidth sub-TLV with a definition that is generalized into the following:

The Unreserved Bandwidth sub-TLV specifies the amount of bandwidth not yet reserved for each of the eight TE-Classes, in IEEE floating point format arranged in increasing order of TE-Class index. Unreserved bandwidth for TE-Class [0] occurs at the start of the sub-TLV, and unreserved bandwidth for TE-Class [7] at the end of the sub-TLV. The unreserved bandwidth value for TE-Class [i] (0 <= i <= 7) is referred to as "Unreserved TE-Class [i]". It indicates the bandwidth that is available, for reservation, to an LSP that:

- transports a Traffic Trunk from the Class-Type of TE-Class[i], and

- has a setup priority corresponding to the preemption priority of TE-Class[i].

The units are bytes per second.

Because the bandwidth values are now ordered by TE-class index and thus can relate to different CTs with different BCs and to any arbitrary preemption priority, a DS-TE LSR MUST NOT assume any ordered relationship among these bandwidth values.

With existing TE, because all preemption priorities reflect the same (and only) BCs and bandwidth values are advertised in preemption priority order, the following relationship is always true, and is often assumed by TE implementations:

```
      If i < j, then "Unreserved Bw [i]" >= "Unreserved Bw [j]"
```

With DS-TE, no relationship is to be assumed such that:

```
      If i < j, then any of the following relationships may be true:
                "Unreserved TE-Class [i]" = "Unreserved TE-Class [j]"
                    OR
                "Unreserved TE-Class [i]" > "Unreserved TE-Class [j]"
                    OR
                "Unreserved TE-Class [i]" < "Unreserved TE-Class [j]".
```

Rules for computing "Unreserved TE-Class [i]" are specified in Section 11.

If TE-Class[i] is unused, the value advertised by the IGP in "Unreserved TE-Class [i]" MUST be set to zero by the LSR generating the IGP advertisement, and MUST be ignored by the LSR receiving the IGP advertisement.

6. RSVP-TE Extensions for DS-TE

In this section, we describe extensions to RSVP-TE for support of Diffserv-aware MPLS Traffic Engineering. These extensions are in addition to the extensions to RSVP defined in [RSVP-TE] for support of (aggregate) MPLS Traffic Engineering and to the extensions to RSVP defined in [DIFF-MPLS] for support of Diffserv over MPLS.

6.1. DS-TE-Related RSVP Messages Format

One new RSVP object is defined in this document: the CLASSTYPE object. Detailed description of this object is provided below. This new object is applicable to Path messages. This specification only defines the use of the CLASSTYPE object in Path messages used to establish LSP Tunnels in accordance with [RSVP-TE] and thus containing a session object with a CT equal to LSP_TUNNEL_IPv4 and containing a LABEL_REQUEST object.

Restrictions defined in [RSVP-TE] for support of establishment of LSP Tunnels via RSVP-TE are also applicable to the establishment of LSP Tunnels supporting DS-TE. For instance, only unicast LSPs are supported, and multicast LSPs are for further study.

This new CLASSTYPE object is optional with respect to RSVP so that general RSVP implementations not concerned with MPLS LSP setup do not have to support this object.

An LSR supporting DS-TE MUST support the CLASSTYPE object.

6.1.1. Path Message Format

The format of the Path message is as follows:

```
<Path Message> ::=      <Common Header> [ <INTEGRITY> ]
                        <SESSION> <RSVP_HOP>
                        <TIME_VALUES>
                        [ <EXPLICIT_ROUTE> ]
                        <LABEL_REQUEST>
                        [ <SESSION_ATTRIBUTE> ]
                        [ <DIFFSERV> ]
                        [ <CLASSTYPE> ]
                        [ <POLICY_DATA> ... ]
                        [ <sender descriptor> ]

<sender descriptor> ::=  <SENDER_TEMPLATE> [ <SENDER_TSPEC> ]
                        [ <ADSPEC> ]
                        [ <RECORD_ROUTE> ]
```

6.2. CLASSTYPE Object

The CLASSTYPE object Class Name is CLASSTYPE. Its Class Number is 66. Currently, there is only one defined C-Type which is C-Type 1. The CLASSTYPE object format is shown below.

6.2.1. CLASSTYPE object

Class Number = 66
Class-Type = 1

```
 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
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|        Reserved                                         | CT  |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Reserved: 29 bits
    This field is reserved. It MUST be set to zero on transmission and MUST be ignored on receipt.

CT: 3 bits
    Indicates the Class-Type. Values currently allowed are 1, 2, ... , 7. Value of 0 is Reserved.

6.3. Handling CLASSTYPE Object

To establish an LSP tunnel with RSVP, the sender LSR creates a Path message with a session type of LSP_Tunnel_IPv4 and with a

LABEL_REQUEST object as per [RSVP-TE]. The sender LSR may also include the DIFFSERV object as per [DIFF-MPLS].

If the LSP is associated with Class-Type 0, the sender LSR MUST NOT include the CLASSTYPE object in the Path message. This allows backward compatibility with non-DSTE-configured or non-DSTE-capable LSRs as discussed below in Section 10 and Appendix C.

If the LSP is associated with Class-Type N (1 <= N <=7), the sender LSR MUST include the CLASSTYPE object in the Path message with the Class-Type (CT) field set to N.

If a Path message contains multiple CLASSTYPE objects, only the first one is meaningful; subsequent CLASSTYPE object(s) MUST be ignored and MUST NOT be forwarded.

Each LSR along the path MUST record the CLASSTYPE object, when it is present, in its path state block.

If the CLASSTYPE object is not present in the Path message, the LSR MUST associate the Class-Type 0 to the LSP.

The destination LSR responding to the Path message by sending a Resv message MUST NOT include a CLASSTYPE object in the Resv message (whether or not the Path message contained a CLASSTYPE object).

During establishment of an LSP corresponding to the Class-Type N, the LSR MUST perform admission control over the bandwidth available for that particular Class-Type.

An LSR that recognizes the CLASSTYPE object and that receives a Path message that:

- contains the CLASSTYPE object, but

- does not contain a LABEL_REQUEST object or does not have a session type of LSP_Tunnel_IPv4,

MUST send a PathErr towards the sender with the error code "Diffserv-aware TE Error" and an error value of "Unexpected CLASSTYPE object". These codes are defined in Section 6.5.

An LSR receiving a Path message with the CLASSTYPE object that:

- recognizes the CLASSTYPE object, but

- does not support the particular Class-Type,

MUST send a PathErr towards the sender with the error code "Diffserv-aware TE Error" and an error value of "Unsupported Class-Type". These codes are defined in Section 6.5.

An LSR receiving a Path message with the CLASSTYPE object that:

- recognizes the CLASSTYPE object, but

- determines that the Class-Type value is not valid (i.e., Class-Type value 0),

MUST send a PathErr towards the sender with the error code "Diffserv-aware TE Error" and an error value of "Invalid Class-Type value". These codes are defined in Section 6.5.

An LSR receiving a Path message with the CLASSTYPE object, which:

- recognizes the CLASSTYPE object and

- supports the particular Class-Type, but

- determines that the tuple formed by (i) this Class-Type and (ii) the setup priority signaled in the same Path message, is not one of the eight TE-Classes configured in the TE-class mapping,

MUST send a PathErr towards the sender with the error code "Diffserv-aware TE Error" and an error value of "CT and setup priority do not form a configured TE-Class". These codes are defined in Section 6.5.

An LSR receiving a Path message with the CLASSTYPE object that:

- recognizes the CLASSTYPE object and

- supports the particular Class-Type, but

- determines that the tuple formed by (i) this Class-Type and (ii) the holding priority signaled in the same Path message, is not one of the eight TE-Classes configured in the TE-class mapping, MUST send a PathErr towards the sender with the error code "Diffserv-aware TE Error" and an error value of "CT and holding priority do not form a configured TE-Class". These codes are defined in Section 6.5.

An LSR receiving a Path message with the CLASSTYPE object that:

- recognizes the CLASSTYPE object and

- supports the particular Class-Type, but

- determines that the tuple formed by (i) this Class-Type and (ii) the setup priority signaled in the same Path message, is not one of the eight TE-Classes configured in the TE-class mapping, AND

- determines that the tuple formed by (i) this Class-Type and (ii) the holding priority signaled in the same Path message, is not one of the eight TE-Classes configured in the TE-class mapping

MUST send a PathErr towards the sender with the error code "Diffserv-aware TE Error" and an error value of "CT and setup priority do not form a configured TE-Class AND CT and holding priority do not form a configured TE-Class". These codes are defined in Section 6.5.

An LSR receiving a Path message with the CLASSTYPE object and with the DIFFSERV object for an L-LSP that:

- recognizes the CLASSTYPE object,

- has local knowledge of the relationship between Class-Types and Per Hop Behavior (PHB) Scheduling Class, e.g., via configuration, and

- determines, based on this local knowledge, that the PHB Scheduling Class (PSC) signaled in the DIFFSERV object is inconsistent with the Class-Type signaled in the CLASSTYPE object,

MUST send a PathErr towards the sender with the error code "Diffserv-aware TE Error" and an error value of "Inconsistency between signaled PSC and signaled CT". These codes are defined below in Section 6.5.

An LSR receiving a Path message with the CLASSTYPE object and with the DIFFSERV object for an E-LSP that:

- recognizes the CLASSTYPE object,

- has local knowledge of the relationship between Class-Types and PHBs (e.g., via configuration)

- determines, based on this local knowledge, that the PHBs signaled in the MAP entries of the DIFFSERV object are inconsistent with the Class-Type signaled in the CLASSTYPE object,

MUST send a PathErr towards the sender with the error code "Diffserv-aware TE Error" and an error value of "Inconsistency between signaled PHBs and signaled CT". These codes are defined in Section 6.5.

An LSR MUST handle situations in which the LSP cannot be accepted for reasons other than those already discussed in this section, in accordance with [RSVP-TE] and [DIFF-MPLS] (e.g., a reservation is rejected by admission control, and a label cannot be associated).

6.4. Non-support of the CLASSTYPE Object

An LSR that does not recognize the CLASSTYPE object Class-Num MUST behave in accordance with the procedures specified in [RSVP] for an unknown Class-Num whose format is 0bbbbbbb (i.e., it MUST send a PathErr with the error code "Unknown object class" toward the sender).

An LSR that recognizes the CLASSTYPE object Class-Num but that does not recognize the CLASSTYPE object C-Type, MUST behave in accordance with the procedures specified in [RSVP] for an unknown C-type (i.e., it MUST send a PathErr with the error code "Unknown object C-Type" toward the sender).

Both of the above situations cause the path setup to fail. The sender SHOULD notify the operator/management system that an LSP cannot be established and might take action to retry reservation establishment without the CLASSTYPE object.

6.5. Error Codes for Diffserv-aware TE

In the procedures described above, certain errors are reported as a "Diffserv-aware TE Error". The value of the "Diffserv-aware TE Error" error code is 28.

The following table defines error values for the Diffserv-aware TE Error:

Value	Error
1	Unexpected CLASSTYPE object
2	Unsupported Class-Type
3	Invalid Class-Type value
4	Class-Type and setup priority do not form a configured TE-Class
5	Class-Type and holding priority do not form a configured TE-Class
6	Class-Type and setup priority do not form a configured TE-Class AND Class-Type and holding priority do not form a configured TE-Class
7	Inconsistency between signaled PSC and signaled Class-Type
8	Inconsistency between signaled PHBs and signaled Class-Type

See the IANA Considerations section for allocation of additional values.

7. DS-TE Support with MPLS Extensions

There are a number of extensions to the initial base specification for signaling [RSVP-TE] and IGP support for TE [OSPF-TE][ISIS-TE]. Those include enhancements for generalization ([GMPLS-SIG] and [GMPLS-ROUTE]), as well as for additional functionality, such as LSP hierarchy [HIERARCHY], link bundling [BUNDLE], and fast restoration [REROUTE]. These specifications may reference how to encode information associated with certain preemption priorities, how to treat LSPs at different preemption priorities, or they may otherwise specify encodings or behavior that have a different meaning for a DS-TE router.

In order for an implementation to support both this specification for Diffserv-aware TE and a given MPLS enhancement, such as those listed above (but not limited to those), it MUST treat references to "preemption priority" and to "Maximum Reservable Bandwidth" in a generalized manner, i.e., the manner in which this specification uses those terms.

Additionally, current and future MPLS enhancements may include more precise specification for how they interact with Diffserv-aware TE.

7.1. DS-TE Support and References to Preemption Priority

When a router supports both Diffserv-aware TE and one of the MPLS protocol extensions such as those mentioned above, encoding of values of preemption priority in signaling or encoding of information associated with preemption priorities in IGP defined for the MPLS extension, MUST be considered an encoding of the same information for the corresponding TE-Class. For instance, if an MPLS enhancement specifies advertisement in IGP of a parameter for routing information at preemption priority N, in a DS-TE environment it MUST actually be interpreted as specifying advertisement of the same routing information but for TE-Class [N]. On receipt, DS-TE routers MUST also interpret it as such.

When there is discussion on how to comparatively treat LSPs of different preemption priority, a DS-TE LSR MUST treat the preemption priorities in this context as those associated with the TE-Classes of the LSPs in question.

7.2. DS-TE Support and References to Maximum Reservable Bandwidth

When a router supports both Diffserv-aware TE and MPLS protocol extensions such as those mentioned above, advertisements of Maximum Reservable Bandwidth MUST be done with the generalized interpretation defined in Section 4.1.1 as the aggregate bandwidth constraint across all Class-Types. It MAY also allow the optional advertisement of all BCs.

8. Constraint-Based Routing

Let us consider the case where a path needs to be computed for an LSP whose Class-Type is configured to CTc and whose setup preemption priority is configured to p.

Then the pair of CTc and p will map to one of the TE-Classes defined in the TE-Class mapping. Let us refer to this TE-Class as TE-Class[i].

The Constraint-Based Routing algorithm of a DS-TE LSR is still only required to perform path computation satisfying a single BC which is to fit in "Unreserved TE-Class [i]" as advertised by the IGP for every link. Thus, no changes to the existing TE Constraint-Based Routing algorithm itself are required.

The Constraint-Based Routing algorithm MAY also take into account, when used, the optional additional information advertised in IGP such as the BCs and the Maximum Reservable Bandwidth. For example, the

BCs MIGHT be used as tie-breaker criteria in situations where multiple paths, otherwise equally attractive, are possible.

9. Diffserv Scheduling

The Class-Type signaled at LSP establishment MAY optionally be used by DS-TE LSRs to dynamically adjust the resources allocated to the Class-Type by the Diffserv scheduler. In addition, the Diffserv information (i.e., the PSC) signaled by the TE-LSP signaling protocols as specified in [DIFF-MPLS], if used, MAY optionally be used by DS-TE LSRs to dynamically adjust the resources allocated by the Diffserv scheduler to a PSC/OA within a CT.

10. Existing TE as a Particular Case of DS-TE

We observe that existing TE can be viewed as a particular case of DS-TE where:

```
   (i)   a single Class-Type is used,
   (ii)  all 8 preemption priorities are allowed for that Class-Type,
         and
   (iii) the following TE-Class mapping is used:
               TE-Class[i]  <-->  < CT0 , preemption i >
               Where 0 <= i <= 7.
```

In that case, DS-TE behaves as existing TE.

As with existing TE, the IGP advertises:
- Unreserved Bandwidth for each of the 8 preemption priorities.

As with existing TE, the IGP may advertise:
- Maximum Reservable Bandwidth containing a BC applying across all LSPs .

Because all LSPs transport traffic from CT0, RSVP-TE signaling is done without explicit signaling of the Class-Type (which is only used for Class-Types other than CT0, as explained in Section 6) as with existing TE.

11. Computing "Unreserved TE-Class [i]" and Admission Control Rules

11.1. Computing "Unreserved TE-Class [i]"

We first observe that, for existing TE, details on admission control algorithms for TE LSPs, and consequently details on formulas for computing the unreserved bandwidth, are outside the scope of the current IETF work. This is left for vendor differentiation. Note that this does not compromise interoperability across various implementations because the TE schemes rely on LSRs to advertise their local view of the world in terms of Unreserved Bw to other LSRs. This way, regardless of the actual local admission control algorithm used on one given LSR, Constraint-Based Routing on other LSRs can rely on advertised information to determine whether an additional LSP will be accepted or rejected by the given LSR. The only requirement is that an LSR advertises unreserved bandwidth values that are consistent with its specific local admission control algorithm and take into account the holding preemption priority of established LSPs.

In the context of DS-TE, again, details on admission control algorithms are left for vendor differentiation, and formulas for computing the unreserved bandwidth for TE-Class[i] are outside the scope of this specification. However, DS-TE places the additional requirement on the LSR that the unreserved bandwidth values advertised MUST reflect all the BCs relevant to the CT associated with TE-Class[i] in accordance with the Bandwidth Constraints Model. Thus, formulas for computing "Unreserved TE-Class [i]" depend on the Bandwidth Constraints Model in use and MUST reflect how BCs apply to CTs. Example formulas for computing "Unreserved TE-Class [i]" Model are provided for the Russian Dolls Model and Maximum Allocation Model respectively in [DSTE-RDM] and [DSTE-MAM].

As with existing TE, DS-TE LSRs MUST consider the holding preemption priority of established LSPs (as opposed to their setup preemption priority) for the purpose of computing the unreserved bandwidth for TE-Class [i].

11.2. Admission Control Rules

A DS-TE LSR MUST support the following admission control rule:

Regardless of how the admission control algorithm actually computes the unreserved bandwidth for TE-Class[i] for one of its local links, an LSP of bandwidth B, of setup preemption priority p and of Class-Type CTc is admissible on that link if, and only if,:

    B <= Unreserved Bandwidth for TE-Class[i]

where TE-Class [i] maps to < CTc , p > in the TE-Class mapping configured on the LSR.

12. Security Considerations

This document does not introduce additional security threats beyond those described for Diffserv ([DIFF-ARCH]) and MPLS Traffic Engineering ([TE-REQ], [RSVP-TE], [OSPF-TE], [ISIS-TE]) and the same security measures and procedures described in these documents apply here. For example, the approach for defense against theft- and denial-of-service attacks discussed in [DIFF-ARCH], which consists of the combination of traffic conditioning at DS boundary nodes along with security and integrity of the network infrastructure within a Diffserv domain, may be followed when DS-TE is in use. Also, as stated in [TE-REQ], it is specifically important that manipulation of administratively configurable parameters (such as those related to DS-TE LSPs) be executed in a secure manner by authorized entities.

13. IANA Considerations

This document creates two new name spaces that are to be managed by IANA. Also, a number of assignments from existing name spaces have been made by IANA in this document. They are discussed below.

13.1. A New Name Space for Bandwidth Constraints Model Identifiers

This document defines in Section 5.1 a "Bandwidth Constraints Model Id" field (name space) within the "Bandwidth Constraints" sub-TLV, both for OSPF and ISIS. The new name space has been created by the IANA and they will maintain this new name space. The field for this namespace is 1 octet, and IANA guidelines for assignments for this field are as follows:

- o values in the range 0-239 are to be assigned according to the "Specification Required" policy defined in [IANA-CONS].

- o values in the range 240-255 are reserved for "Private Use" as defined in [IANA-CONS].

13.2. A New Name Space for Error Values under the "Diffserv-aware TE Error"

An Error Code is an 8-bit quantity defined in [RSVP] that appears in an ERROR_SPEC object to define an error condition broadly. With each Error Code there may be a 16-bit Error Value (which depends on the Error Code) that further specifies the cause of the error.

This document defines in Section 6.5 a new RSVP error code, the "Diffserv-aware TE Error" (see Section 13.3.4). The Error Values for the "Diffserv-aware TE Error" constitute a new name space to be managed by IANA.

This document defines, in Section 6.5, values 1 through 7 in that name space (see Section 13.3.5).

Future allocations of values in this name space are to be assigned by IANA using the "Specification Required" policy defined in [IANA-CONS].

13.3. Assignments Made in This Document

13.3.1. Bandwidth Constraints sub-TLV for OSPF Version 2

[OSPF-TE] creates a name space for the sub-TLV types within the "Link TLV" of the Traffic Engineering Link State Advertisement (LSA) and rules for management of this name space by IANA.

This document defines in Section 5.1 a new sub-TLV, the "Bandwidth Constraints" sub-TLV, for the OSPF "Link" TLV. In accordance with the IANA considerations provided in [OSPF-TE], a sub-TLV type in the range 10 to 32767 was requested, and the value 17 has been assigned by IANA for the "Bandwidth Constraints" sub-TLV.

13.3.2. Bandwidth Constraints sub-TLV for ISIS

[ISIS-TE] creates a name space for the sub-TLV types within the ISIS "Extended IS Reachability" TLV and rules for management of this name space by IANA.

This document defines in Section 5.1 a new sub-TLV, the "Bandwidth Constraints" sub-TLV, for the ISIS "Extended IS Reachability" TLV. In accordance with the IANA considerations provided in [ISIS-TE], a sub-TLV type was requested, and the value 22 has been assigned by IANA for the "Bandwidth Constraints" sub-TLV.

13.3.3. CLASSTYPE Object for RSVP

[RSVP] defines the Class Number name space for RSVP object, which is managed by IANA. Currently allocated Class Numbers are listed at http://www.iana.org/assignments/rsvp-parameters.

This document defines in Section 6.2.1 a new RSVP object, the CLASSTYPE object. IANA has assigned a Class Number for this RSVP object from the range defined in Section 3.10 of [RSVP] for objects that, if not understood, cause the entire RSVP message to be rejected with an error code of "Unknown Object Class". Such objects are identified by a zero in the most significant bit of the class number (i.e., Class-Num = 0bbbbbbb).

IANA assigned Class-Number 66 to the CLASSTYPE object. C_Type 1 is defined in this document for the CLASSTYPE object.

13.3.4. "Diffserv-aware TE Error" Error Code

[RSVP] defines the Error Code name space and rules for management of this name space by IANA. Currently allocated Error Codes are listed at http://www.iana.org/assignments/rsvp-parameters.

This document defines in Section 6.5 a new RSVP Error Code, the "Diffserv-aware TE Error". In accordance with the IANA considerations provided in [RSVP], Error Code 28 was assigned by IANA to the "Diffserv-aware TE Error".

13.3.5. Error Values for "Diffserv-aware TE Error"

An Error Code is an 8-bit quantity defined in [RSVP] that appears in an ERROR_SPEC object to define an error condition broadly. With each Error Code there may be a 16-bit Error Value (which depends on the Error Code) that further specifies the cause of the error.

This document defines in Section 6.5 a new RSVP error code, the "Diffserv-aware TE Error" (see Section 13.3.4). The Error Values for the "Diffserv-aware TE Error" constitute a new name space to be managed by IANA.

This document defines, in Section 6.5, the following Error Values for the "Diffserv-aware TE Error":

Value	Error
1	Unexpected CLASSTYPE object
2	Unsupported Class-Type
3	Invalid Class-Type value
4	Class-Type and setup priority do not form a configured TE-Class
5	Class-Type and holding priority do not form a configured TE-Class
6	Class-Type and setup priority do not form a configured TE-Class AND Class-Type and holding priority do not form a configured TE-Class
7	Inconsistency between signaled PSC and signaled Class-Type
8	Inconsistency between signaled PHBs and signaled Class-Type

See Section 13.2 for allocation of other values in that name space.

14. Acknowledgements

We thank Martin Tatham, Angela Chiu, and Pete Hicks for their earlier contribution in this work. We also thank Sanjaya Choudhury for his thorough review and suggestions.

Appendix A: Prediction for Multiple Path Computation

There are situations where a head-end needs to compute paths for multiple LSPs over a short period of time. There are potential advantages for the head-end in trying to predict the impact of the n-th LSP on the unreserved bandwidth when computing the path for the (n+1)-th LSP, before receiving updated IGP information. For example, better load-distribution of the multiple LSPs would be performed across multiple paths. Also, when the (n+1)-th LSP would no longer fit on a link after establishment of the n-th LSP, the head-end would avoid Connection Admission Control (CAC) rejection. Although there are a number of conceivable scenarios where worse situations might result, doing such predictions is more likely to improve situations. As a matter of fact, a number of network administrators have elected to use such predictions when deploying existing TE.

Such predictions are local matters, are optional, and are outside the scope of this specification.

Where such predictions are not used, the optional BC sub-TLV and the optional Maximum Reservable Bandwidth sub-TLV need not be advertised in IGP for the purpose of path computation, since the information contained in the Unreserved Bw sub-TLV is all that is required by Head-Ends to perform Constraint-Based Routing.

Where such predictions are used on head-ends, the optional BCs sub-TLV and the optional Maximum Reservable Bandwidth sub-TLV MAY be advertised in IGP. This is in order for the head-ends to predict as accurately as possible how an LSP affects unreserved bandwidth values for subsequent LSPs.

Remembering that actual admission control algorithms are left for vendor differentiation, we observe that predictions can only be performed effectively when the head-end LSR predictions are based on the same (or a very close) admission control algorithm as that used by other LSRs.

Appendix B: Solution Evaluation

B.1. Satisfying Detailed Requirements

This DS-TE Solution addresses all the scenarios presented in [DSTE-REQ].

It also satisfies all the detailed requirements presented in [DSTE-REQ].

The objective set out in the last paragraph of Section 4.7 of [DSTE-REQ], "Overbooking", is only partially addressed by this DS-TE solution. Through support of the "LSP size Overbooking" and "Link Size Overbooking" methods, this DS-TE solution effectively allows CTs to have different overbooking ratios and simultaneously allows overbooking to be tweaked differently (collectively across all CTs) on different links. But, in a general sense, it does not allow the effective overbooking ratio of every CT to be tweaked differently in different parts of the network independently of other CTs, while maintaining accurate bandwidth accounting of how different CTs mutually affect each other through shared BCs (such as the Maximum Reservable Bandwidth).

B.2. Flexibility

This DS-TE solution supports 8 CTs. It is entirely flexible as to how Traffic Trunks are grouped together into a CT.

B.3. Extendibility

A maximum of 8 CTs is considered more than comfortable by the authors of this document. A maximum of 8 TE-Classes is considered sufficient by the authors of this document. However, this solution could be extended to support more CTs or more TE-Classes if deemed necessary in the future; this would necessitate additional IGP extensions beyond those specified in this document.

Although the prime objective of this solution is support of Diffserv-aware Traffic Engineering, its mechanisms are not tightly coupled with Diffserv. This makes the solution amenable, or more easily extendable, for support of potential other future Traffic Engineering applications.

B.4. Scalability

This DS-TE solution is expected to have a very small scalability impact compared to that of existing TE.

From an IGP viewpoint, the amount of mandatory information to be advertised is identical to that of existing TE. One additional sub-TLV has been specified, but its use is optional, and it only contains a limited amount of static information (at most 8 BCs).

We expect no noticeable impact on LSP Path computation because, as with existing TE, this solution only requires Constrained Shortest Path First (CSPF) to consider a single unreserved bandwidth value for any given LSP.

From a signaling viewpoint, we expect no significant impact due to this solution because it only requires processing of one additional item of information (the Class-Type) and does not significantly increase the likelihood of CAC rejection. Note that DS-TE has some inherent impact on LSP signaling in that it assumes that different classes of traffic are split over different LSPs so that more LSPs need to be signaled. However, this is due to the DS-TE concept itself and not to the actual DS-TE solution discussed here.

B.5. Backward Compatibility/Migration

This solution is expected to allow smooth migration from existing TE to DS-TE. This is because existing TE can be supported as a particular configuration of DS-TE. This means that an "upgraded" LSR with a DS-TE implementation can directly interwork with an "old" LSR supporting existing TE only.

This solution is expected to allow smooth migration when the number of CTs actually deployed is increased, as it only requires configuration changes. However, these changes need to be performed in a coordinated manner across the DS-TE domain.

Appendix C: Interoperability with Non-DS-TE Capable LSRs

This DSTE solution allows operations in a hybrid network where some LSRs are DS-TE capable and some are not, as may occur during migration phases. This appendix discusses the constraints and operations in such hybrid networks.

We refer to the set of DS-TE-capable LSRs as the DS-TE domain. We refer to the set of non-DS-TE-capable (but TE-capable) LSRs as the TE-domain.

Hybrid operations require that the TE-Class mapping in the DS-TE domain be configured so that:

- a TE-Class exists for CT0 for every preemption priority actually used in the TE domain, and

- the index in the TE-class mapping for each of these TE-Classes is equal to the preemption priority.

For example, imagine the TE domain uses preemption 2 and 3. Then, DS-TE can be deployed in the same network by including the following TE-Classes in the TE-Class mapping:

i <--->	CT	preemption
2	CT0	2
3	CT0	3

Another way to look at this is to say that although the whole TE-class mapping does not have to be consistent with the TE domain, the subset of this TE-Class mapping applicable to CT0 effectively has to be consistent with the TE domain.

Hybrid operations also require that:

- non-DS-TE-capable LSRs be configured to advertise the Maximum Reservable Bandwidth, and

- DS-TE-capable LSRs be configured to advertise BCs (using the Max Reservable Bandwidth sub-TLV as well as the BCs sub-TLV, as specified in Section 5.1).

This allows DS-TE-capable LSRs to identify non-DS-TE-capable LSRs unambiguously.

Finally, hybrid operations require that non-DS-TE-capable LSRs be able to accept Unreserved Bw sub-TLVs containing non decreasing bandwidth values (i.e., with Unreserved [p] < Unreserved [q] with p < q).

In such hybrid networks, the following apply:

- CT0 LSPs can be established by both DS-TE-capable LSRs and non-DS-TE-capable LSRs.

- CT0 LSPs can transit via (or terminate at) both DS-TE-capable LSRs and non-DS-TE-capable LSRs.

- LSPs from other CTs can only be established by DS-TE-capable LSRs.

- LSPs from other CTs can only transit via (or terminate at) DS-TE-capable LSRs.

Let us consider the following example to illustrate operations:

```
   LSR0--------LSR1----------LSR2
        Link01       Link12
```

```
   where:
      LSR0 is a non-DS-TE-capable LSR
      LSR1 and LSR2 are DS-TE-capable LSRs
```

Let's assume again that preemptions 2 and 3 are used in the TE-domain and that the following TE-Class mapping is configured on LSR1 and LSR2:

i <--->	CT	preemption
0	CT1	0
1	CT1	1
2	CT0	2
3	CT0	3
rest	unused	

LSR0 is configured with a Max Reservable Bandwidth = m01 for Link01. LSR1 is configured with a BC0 = x0, a BC1 = x1 (possibly = 0), and a Max Reservable Bandwidth = m10 (possibly = m01) for Link01.

In IGP for Link01, LSR0 will advertise:

- Max Reservable Bw sub-TLV = <m01>

- Unreserved Bw sub-TLV = <CT0/0, CT0/1, CT0/2, CT0/3, CT0/4, CT0/5, CT0/6, CT0/7>

On receipt of such advertisement, LSR1 will:

- understand that LSR0 is not DS-TE-capable because it advertised a Max Reservable Bw sub-TLV and no Bandwidth Constraints sub-TLV, and

- conclude that only CT0 LSPs can transit via LSR0 and that only the values CT0/2 and CT0/3 are meaningful in the Unreserved Bw sub-TLV. LSR1 may effectively behave as if the six other values contained in the Unreserved Bw sub-TLV were set to zero.

In IGP for Link01, LSR1 will advertise:

- Max Reservable Bw sub-TLV = <m10>

- Bandwidth Constraints sub-TLV = <BC Model ID, x0, x1>

- Unreserved Bw sub-TLV = <CT1/0, CT1/1, CT0/2, CT0/3, 0, 0, 0, 0>

On receipt of such advertisement, LSR0 will:

- ignore the Bandwidth Constraints sub-TLV (unrecognized)

- correctly process CT0/2 and CT0/3 in the Unreserved Bw sub-TLV and use these values for CTO LSP establishment

- incorrectly believe that the other values contained in the Unreserved Bw sub-TLV relate to other preemption priorities for CT0; but it will actually never use those since we assume that only preemptions 2 and 3 are used in the TE domain.

Normative References

[DSTE-REQ] Le Faucheur, F. and W. Lai, "Requirements for Support of Differentiated Services-aware MPLS Traffic Engineering", RFC 3564, July 2003.

[MPLS-ARCH] Rosen, E., Viswanathan, A. and R. Callon, "Multiprotocol Label Switching Architecture", RFC 3031, January 2001.

[TE-REQ] Awduche, D., Malcolm, J., Agogbua, J., O'Dell, M. and J. McManus, "Requirements for Traffic Engineering Over MPLS", RFC 2702, September 1999.

[OSPF-TE] Katz, D., Kompella, K. and D. Yeung, "Traffic Engineering (TE) Extensions to OSPF Version 2", RFC 3630, September 2003.

[ISIS-TE] Smit, H. and T. Li, "Intermediate System to Intermediate System (IS-IS) Extensions for Traffic Engineering (TE)", RFC 3784, June 2004.

[RSVP-TE] Awduche, D., Berger, L., Gan, D., Li, T., Srinivasan, V. and G. Swallow, "RSVP-TE: Extensions to RSVP for LSP Tunnels", RFC 3209, December 2001.

[RSVP] Braden, R., Zhang, L., Berson, S., Herzog, S. and S. Jamin, "Resource ReSerVation Protocol (RSVP) -- Version 1 Functional Specification", RFC 2205, September 1997.

[DIFF-MPLS] Le Faucheur, F., Wu, L., Davie, B., Davari, S., Vaananen, P., Krishnan, R., Cheval, P. and J. Heinanen, "Multi-Protocol Label Switching (MPLS) Support of Differentiated Services", RFC 3270, May 2002.

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

[IANA-CONS] Narten, T. and H. Alvestrand, "Guidelines for Writing an IANA Considerations Section in RFCs", BCP 26, RFC 2434, October 1998.

Informative References

[DIFF-ARCH] Blake, S., Black, D., Carlson, M., Davies, E., Wang, Z., and W. Weiss, "An Architecture for Differentiated Service", RFC 2475, December 1998.

[DSTE-RDM] Le Faucheur,F., Ed., "Russian Dolls Bandwidth Constraints Model for Diffserv-aware MPLS Traffic Engineering", RFC 4127, June 2005.

[DSTE-MAM] Le Faucheur, F. and W. Lai, "Maximum Allocation Bandwidth Constraints Model for Diffserv-aware Traffice Engineering", RFC 4125, June 2005.

[DSTE-MAR] Ash, J., "Max Allocation with Reservation Bandwidth Constraints Model for DiffServ-aware MPLS Traffic Engineering & Performance Comparisons", RFC 4126, June 2005.

[GMPLS-SIG] Berger, L., "Generalized Multi-Protocol Label Switching (GMPLS) Signaling Functional Description", RFC 3471, January 2003.

[GMPLS-ROUTE] Kompella, et al., "Routing Extensions in Support of Generalized MPLS", Work in Progress.

[BUNDLE] Kompella, Rekhter, Berger, "Link Bundling in MPLS Traffic Engineering", Work in Progress.

[HIERARCHY] Kompella, Rekhter, "LSP Hierarchy with Generalized MPLS TE", Work in Progress.

[REROUTE] Pan, P., Swallow, G., and A. Atlas, "Fast Reroute Extensions to RSVP-TE for LSP Tunnels", RFC 4090, May 2005.

Editor's Address

Francois Le Faucheur
Cisco Systems, Inc.
Village d'Entreprise Green Side - Batiment T3
400, Avenue de Roumanille
06410 Biot-Sophia Antipolis
France

Phone: +33 4 97 23 26 19
EMail: flefauch@cisco.com

Acknowledgement

Funding for the RFC Editor function is currently provided by the Internet Society.