             Requirements for Support of Differentiated Services-aware
                        MPLS Traffic Engineering

Status of this Memo

   This memo provides information for the Internet community. It does
   not specify an Internet standard of any kind. Distribution of this
   memo is unlimited.

Copyright Notice

   Copyright (C) The Internet Society (2003). All Rights Reserved.

Abstract

   This document presents Service Provider requirements for support of
   Differentiated Services (Diff-Serv)-aware MPLS Traffic Engineering
   (DS-TE).

   Its objective is to provide guidance for the definition, selection
   and specification of a technical solution addressing these
   requirements. Specification for this solution itself is outside the
   scope of this document.

   A problem statement is first provided. Then, the document describes
   example applications scenarios identified by Service Providers where
   existing MPLS Traffic Engineering mechanisms fall short and
   Diff-Serv-aware Traffic Engineering can address the needs. The
   detailed requirements that need to be addressed by the technical
   solution are also reviewed. Finally, the document identifies the
   evaluation criteria that should be considered for selection and
   definition of the technical solution.

Table of Contents

Specification 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].

1. Introduction

1.1. Problem Statement

   Diff-Serv is used by some Service Providers to achieve scalable
   network designs supporting multiple classes of services.

   In some such Diff-Serv networks, where optimization of transmission
   resources on a network-wide basis is not sought, MPLS Traffic
   Engineering (TE) mechanisms may not be used.

   In other networks, where optimization of transmission resources is
   sought, Diff-Serv mechanisms [DIFF-MPLS] may be complemented by
   MPLS Traffic Engineering mechanisms [TE-REQ] [ISIS-TE] [OSPF-TE]
   [RSVP-TE] which operate on an aggregate basis across all
   Diff-Serv classes of service. In this case, Diff-Serv and MPLS TE
   both provide their respective benefits.

   To achieve fine-grained optimization of transmission resources and
   further enhanced network performance and efficiency, as discussed in
   [TEWG-FW], it may be desirable to perform traffic engineering at a
   per-class level instead of at an aggregate level. By mapping the
   traffic from a given Diff-Serv class of service on a separate LSP, it
   allows this traffic to utilize resources available to the given class
   on both shortest paths and non-shortest paths, and follow paths that
   meet engineering constraints which are specific to the given class.
   This is what we refer to as "Diff-Serv-aware Traffic Engineering
   (DS-TE)".

   This document focuses exclusively on the specific environments which
   would benefit from DS-TE. Some examples include:

      -    networks where bandwidth is scarce (e.g., transcontinental
           networks)
      -    networks with significant amounts of delay-sensitive traffic
      -    networks where the relative proportion of traffic across
           classes of service is not uniform

   This document focuses on intra-domain operation. Inter-domain
   operation is not considered.

1.2. Definitions

   For the convenience of the reader, relevant Diff-Serv ([DIFF-ARCH],
   [DIFF-NEW] and [DIFF-PDB]) definitions are repeated herein.

      Behavior Aggregate (BA): a collection of packets with the same
      (Diff-Serv) codepoint crossing a link in a particular direction.

Per-Hop-Behavior (PHB): the externally observable forwarding
behavior applied at a DS-compliant node to a Diff-Serv behavior
aggregate.

PHB Scheduling Class (PSC): A PHB group for which a common
constraint is that ordering of at least those packets belonging to
the same microflow must be preserved.

Ordered Aggregate (OA): a set of BAs that share an ordering
constraint. The set of PHBs that are applied to this set of
Behavior Aggregates constitutes a PHB scheduling class.

Traffic Aggregate (TA): a collection of packets with a codepoint
that maps to the same PHB, usually in a DS domain or some subset
of a DS domain. A traffic aggregate marked for the foo PHB is
referred to as the "foo traffic aggregate" or "foo aggregate"
interchangeably. This generalizes the concept of Behavior
Aggregate from a link to a network.

Per-Domain Behavior (PDB): the expected treatment that an
identifiable or target group of packets will receive from
"edge-to-edge" of a DS domain. A particular PHB (or, if
applicable, list of PHBs) and traffic conditioning requirements
are associated with each PDB.

We also repeat the following definition from [TE-REQ]:

Traffic Trunk: an aggregation of traffic flows of the same class
which are placed inside a Label Switched Path.

In the context of the present document, "flows of the same class" is
to be interpreted as "flows from the same Forwarding Equivalence
Class which are to be treated equivalently from the DS-TE
perspective".

We refer to the set of TAs corresponding to the set of PHBs of a
given PSC, as a {TA}PSC. A given {TA}PSC will receive the
treatment of the PDB associated with the corresponding PSC. In
this document, we also loosely refer to a {TA}PSC as a "Diff-Serv
class of service", or a "class of service". As an example, the
set of packets within a DS domain with a codepoint that maps to
the EF PHB may form one {TA}PSC in that domain. As another
example, the set of packets within a DS domain with a codepoint
that maps to the AF11 or AF12 or AF13 PHB may form another {TA}PSC
in that domain.

We refer to the collection of packets which belong to a given Traffic
Aggregate and are associated with a given MPLS Forwarding Equivalence
Class (FEC) ([MPLS-ARCH]) as a <FEC/TA>.

We refer to the set of <FEC/TA> whose TAs belong to a given {TA}PSC
as a <FEC/{TA}PSC>.

1.3. Mapping of traffic to LSPs

A network may have multiple Traffic Aggregates (TAs) it wishes to
service. Recalling from [DIFF-MPLS], there are several options on
how the set of <FEC/{TA}PSC> of a given FEC can be split into Traffic
Trunks for mapping onto LSPs when running MPLS Traffic Engineering.

One option is to not split this set of <FEC/{TA}PSC> so that each
Traffic Trunk comprises traffic from all the {TA}/PSC. This option
is typically used when aggregate traffic engineering is deployed
using current MPLS TE mechanisms. In that case, all the
<FEC/{TA}PSC> of a given FEC are routed collectively according to a
single shared set of constraints and will follow the same path. Note
that the LSP transporting such a Traffic Trunk is, by definition, an
E-LSP as defined in [DIFF-MPLS].

Another option is to split the different <FEC/{TA}PSC> of a given FEC
into multiple Traffic Trunks on the basis of the {TA}PSC. In other
words, traffic, from one given node to another, is split, based on
the "classes of service", into multiple Traffic Trunks which are
transported over separate LSP and can potentially follow different
paths through the network. DS-TE takes advantage of this and
computes a separate path for each LSP. In so doing, DS-TE can take
into account the specific requirements of the Traffic Trunk
transported on each LSP (e.g., bandwidth requirement, preemption
priority). Moreover DS-TE can take into account the specific
engineering constraints to be enforced for these sets of Traffic
Trunks (e.g., limit all Traffic Trunks transporting a particular
{TA}PSC to x% of link capacity). DS-TE achieves per LSP constraint
based routing with paths that match specific objectives of the
traffic while forming the corresponding Traffic Trunk.

For simplicity, and because this is the specific topic of this
document, the above paragraphs in this section only considered
splitting traffic of a given FEC into multiple Traffic Aggregates on
the basis of {TA}PSC. However, it should be noted that, in addition
to this, traffic from every {TA}PSC may also be split into multiple
Traffic Trunks for load balancing purposes.

2. Application Scenarios

2.1. Scenario 1: Limiting Proportion of Classes on a Link

   An IP/MPLS network may need to carry a significant amount of VoIP
   traffic compared to its link capacity. For example, 10,000
   uncompressed calls at 20ms packetization result in about 1Gbps of IP
   traffic, which is significant on an OC-48c based network. In case of
   topology changes such as link/node failure, VoIP traffic levels can
   even approach the full bandwidth on certain links.

   For delay/jitter reasons, some network administrators see it as
   undesirable to carry more than a certain percentage of VoIP traffic
   on any link. The rest of the available link bandwidth can be used to
   route other "classes of service" corresponding to delay/jitter
   insensitive traffic (e.g., Best Effort Internet traffic). The exact
   determination of this "certain" percentage is outside the scope of
   this requirements document.

   During normal operations, the VoIP traffic should be able to preempt
   other "classes of service" (if these other classes are designated as
   preemptable and they have lower preemption priority), so that it will
   be able to use the shortest available path, only constrained by the
   maximum defined link utilization ratio/percentage of the VoIP class.

   Existing TE mechanisms only allow constraint based routing of traffic
   based on a single bandwidth constraint common to all "classes of
   service", which does not satisfy the needs described here. This
   leads to the requirement for DS-TE to be able to enforce a different
   bandwidth constraint for different "classes of service". In the
   above example, the bandwidth constraint to be enforced for VoIP
   traffic may be the "certain" percentage of each link capacity, while
   the bandwidth constraint to be enforced for the rest of the "classes
   of service" might have their own constraints or have access to the
   rest of the link capacity.

2.2. Scenario 2: Maintain relative proportion of traffic

   Suppose an IP/MPLS network supports 3 "classes of service". The
   network administrator wants to perform Traffic Engineering to
   distribute the traffic load. Also assume that proportion across
   "classes of service" varies significantly depending on the
   source/destination POPs.

With existing TE mechanisms, the proportion of traffic from each
"class of service" on a given link will vary depending on multiple
factors including:

- in which order the different TE-LSPs are established
- the preemption priority associated with the different TE-LSPs
- link/node failure situations

This may make it difficult or impossible for the network
administrator to configure the Diff-Serv PHBs (e.g., queue bandwidth)
to ensure that each "class of service" gets the appropriate
treatment. This leads again to the requirement for DS-TE to be able
to enforce a different bandwidth constraint for different "classes of
service". This could be used to ensure that, regardless of the order
in which tunnels are routed, regardless of their preemption priority
and regardless of the failure situation, the amount of traffic of
each "class of service" routed over a link matches the Diff-Serv
scheduler configuration on that link to the corresponding class
(e.g., queue bandwidth).

As an illustration of how DS-TE would address this scenario, the
network administrator may configure the service rate of Diff-Serv
queues to (45%,35%,20%) for "classes of service" (1,2,3)
respectively. The administrator would then split the traffic into
separate Traffic Trunks for each "class of service" and associate a
bandwidth to each LSP transporting those Traffic Trunks. The network
administrator may also want to configure preemption priorities of
each LSP in order to give highest restoration priority to the highest
priority "class of service" and medium priority to the medium "class
of service". Then DS-TE could ensure that after a failure, "class of
service" 1 traffic would be rerouted with first access at link
capacity without exceeding its service rate of 45% of the link
bandwidth. "Class of service" 2 traffic would be rerouted with
second access at the link capacity without exceeding its allotment.
Note that where "class of service" 3 is the Best-Effort service, the
requirement on DS-TE may be to ensure that the total amount of
traffic routed across all "classes of service" does not exceed the
total link capacity of 100% (as opposed to separately limiting the
amount of Best Effort traffic to 20 even if there was little "class
of service" 1 and "class of service" 2 traffic).

In this scenario, DS-TE would allow for the maintenance of a more
steady distribution of "classes of service", even during rerouting.
This would rely on the required capability of DS-TE to adjust the
amount of traffic of each "class of service" routed on a link based
on the configuration of the scheduler and the amount of bandwidth
available for each "class of service".

Alternatively, some network administrators may want to solve the
problem by having the scheduler dynamically adjusted based on the
amount of bandwidth of the LSPs admitted for each "class of service".
This is an optional additional requirement on the DS-TE solution.

2.3. Scenario 3: Guaranteed Bandwidth Services

In addition to the Best effort service, an IP/MPLS network operator
may desire to offer a point-to-point "guaranteed bandwidth" service
whereby the provider pledges to provide a given level of performance
(bandwidth/delay/loss...) end-to-end through its network from an
ingress port to an egress port. The goal is to ensure that all the
"guaranteed" traffic under the scope of a subscribed service level
specification, will be delivered within the tolerances of this
service level specification.

One approach for deploying such "guaranteed" service involves:

- dedicating a Diff-Serv PHB (or a Diff-Serv PSC as defined in
  [DIFF-NEW]) to the "guaranteed" traffic
- policing guaranteed traffic on ingress against the traffic contract
  and marking the "guaranteed" packets with the corresponding
  DSCP/EXP value

Where a very high level of performance is targeted for the
"guaranteed" service, it may be necessary to ensure that the amount
of "guaranteed" traffic remains below a given percentage of link
capacity on every link. Where the proportion of "guaranteed" traffic
is high, constraint based routing can be used to enforce such a
constraint.

However, the network operator may also want to simultaneously perform
Traffic Engineering for the rest of the traffic (i.e.,
non-guaranteed traffic) which would require that constraint based
routing is also capable of enforcing a different bandwidth
constraint, which would be less stringent than the one for guaranteed
traffic.

Again, this combination of requirements can not be addressed with
existing TE mechanisms. DS-TE mechanisms allowing enforcement of a
different bandwidth constraint for guaranteed traffic and for
non-guaranteed traffic are required.

3. Detailed Requirements for DS-TE

   This section specifies the functionality that the above scenarios
   require out of the DS-TE solution. Actual technical protocol
   mechanisms and procedures to achieve such functionality are outside
   the scope of this document.

3.1. DS-TE Compatibility

   Since DS-TE may impact scalability (as discussed later in this
   document) and operational practices, DS-TE is expected to be used
   when existing TE mechanisms combined with Diff-Serv cannot address
   the network design requirements (i.e., where constraint based routing
   is required and where it needs to enforce different bandwidth
   constraints for different "classes of service", such as in the
   scenarios described above in section 2). Where the benefits of DSTE
   are only required in a topological subset of their network, some
   network operators may wish to only deploy DS-TE in this topological
   subset.

   Thus, the DS-TE solution MUST be developed in such a way that:

   (i)   it raises no interoperability issues with existing deployed TE
          mechanisms.
   (ii)   it allows DS-TE deployment to the required level of
          granularity and scope (e.g., only in a subset of the topology,
          or only for the number of classes required in the considered
          network)

3.2. Class-Types

   The fundamental requirement for DS-TE is to be able to enforce
   different bandwidth constraints for different sets of Traffic Trunks.

   [TEWG-FW] introduces the concept of Class-Types when discussing
   operations of MPLS Traffic Engineering in a Diff-Serv environment.

   We refine this definition into the following:

          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.

   Note that different LSPs transporting Traffic Trunks from the same CT
   may be using the same or different preemption priorities as explained
   in more details in section 3.4 below.

Mapping of {TA}PSC to Class-Types is flexible. Different {TA}PSC can
be mapped to different CTs, multiple {TA}PSC can be mapped to the
same CT and one {TA}PSC can be mapped to multiple CTs.

For illustration purposes, let's consider the case of a network
running 4 Diff-Serv PDBs which are respectively based on the EF PHB
[EF], the AF1x PSC [AF], the AF2x PSC and the Default (i.e.,
Best-Effort) PHB [DIFF-FIELD]. The network administrator may decide
to deploy DS-TE in the following way:

     o from every DS-TE Head-end to every DS-TE Tail-end, split the
        traffic into 4 Traffic Trunks: one for traffic of each
        {TA}PSC
     o because the QoS objectives for the AF1x PDB and for the AF2x
        PDB may be of similar nature (e.g., both targeting low loss
        albeit at different levels perhaps), the same (set of)
        Bandwidth Constraint(s) may be applied collectively over the
        AF1x Traffic Trunks and the AF2x Traffic Trunks. Thus, the
        network administrator may only define three CTs: one for the
        EF Traffic Trunks, one for the AF1x and AF2x Traffic Trunks
        and one for the Best Effort Traffic Trunks.

As another example of mapping of {TA}PSC to CTs, a network operator
may split the traffic from the {TA}PSC associated with EF into two
different sets of traffic trunks, so that each set of traffic trunks
is subject to different constraints on the bandwidth it can access.
In this case, two distinct CTs are defined for the EF {TA}PSC
traffic: one for the traffic subset subject to the first (set of)
bandwidth constraint(s), the other for the traffic subset subject to
the second (set of) bandwidth constraint(s).

The DS-TE solution MUST support up to 8 CTs. Those are referred to
as CTc, 0 <= c <= MaxCT-1 = 7.
The DS-TE solution MUST be able to enforce a different set of
Bandwidth Constraints for each CT.
A DS-TE implementation MUST support at least 2 CTs, and MAY support
up to 8 CTs.

In a given network, the DS-TE solution MUST NOT require the network
administrator to always deploy the maximum number of CTs. The DS-TE
solution MUST allow the network administrator to deploy only the
number of CTs actually utilized.

3.3. Bandwidth Constraints

   We refer to a Bandwidth Constraint Model as the set of rules
   defining:

   - the maximum number of Bandwidth Constraints; and
   - which CTs each Bandwidth Constraint applies to and how.

   By definition of CT, each CT is assigned either a Bandwidth
   Constraint, or a set of Bandwidth Constraints.

   We refer to the Bandwidth Constraints as BCb, 0 <= b <= MaxBC-1

   For a given Class-Type CTc, 0 <= c <= MaxCT-1, let us define
   "Reserved(CTc)" as the sum of the bandwidth reserved by all
   established LSPs which belong to CTc.

   Different models of Bandwidth Constraints are conceivable for control
   of the CTs.

   For example, a model with one separate Bandwidth Constraint per CT
   could be defined. This model is referred to as the "Maximum
   Allocation Model" and is defined by:

       - MaxBC= MaxCT
       - for each value of b in the range 0 <= b <= (MaxCT - 1):
               Reserved (CTb) <= BCb

   For illustration purposes, on a link of 100 unit of bandwidth where
   three CTs are used, the network administrator might then configure
   BC0=20, BC1= 50, BC2=30 such that:

   - All LSPs supporting Traffic Trunks from CT2 use no more than 30
     (e.g., Voice <= 30)
   - All LSPs supporting Traffic Trunks from CT1 use no more than 50
     (e.g., Premium Data <= 50)
   - All LSPs supporting Traffic Trunks from CT0 use no more than 20
     (e.g., Best Effort <= 20)

   As another example, a "Russian Doll" model of Bandwidth Constraints
   may be defined whereby:

       - MaxBC= MaxCT
       - for each value of b in the range 0 <= b <= (MaxCT - 1):
               SUM (Reserved (CTc)) <= BCb,
               for all "c" in the range b <= c <= (MaxCT - 1)

For illustration purposes, on a link of 100 units of bandwidth where
three CTs are used, the network administrator might then configure
BC0=100, BC1= 80, BC2=60 such that:

- All LSPs supporting Traffic Trunks from CT2 use no more than 60
  (e.g., Voice <= 60)
- All LSPs supporting Traffic Trunks from CT1 or CT2 use no more than
  80 (e.g., Voice + Premium Data <= 80)
- All LSPs supporting Traffic Trunks from CT0 or CT1 or CT2 use no
  more than 100 (e.g., Voice + Premium Data + Best Effort <= 100).

Other Bandwidth Constraints model can also be conceived. Those could
involve arbitrary relationships between BCb and CTc. Those could
also involve additional concepts such as associating minimum
reservable bandwidth to a CT.

The DS-TE technical solution MUST have the capability to support
multiple Bandwidth Constraints models. The DS-TE technical solution
MUST specify at least one bandwidth constraint model and MAY specify
multiple Bandwidth Constraints models. Additional Bandwidth
Constraints models MAY also be specified at a later stage if deemed
useful based on operational experience from DS-TE deployments. The
choice of which (or which set of) Bandwidth Constraints model(s) is
to be supported by a given DS-TE implementation, is an implementation
choice. For simplicity, a network operator may elect to use the same
Bandwidth Constraints Model on all the links of his/her network.
However, if he/she wishes/needs to do so, the network operator may
elect to use different Bandwidth Constraints models on different
links in a given network.

Regardless of the Bandwidth Constraint Model, the DS-TE solution MUST
allow support for up to 8 BCs.

3.4. Preemption and TE-Classes

[TEWG-FW] defines the notion of preemption and preemption priority.
The DS-TE solution MUST retain full support of such preemption.
However, a network administrator preferring not to use preemption for
user traffic MUST be able to disable the preemption mechanisms
described below.

The preemption attributes defined in [TE-REQ] MUST be retained and
applicable across all Class Types. The preemption attributes of
setup priority and holding priority MUST retain existing semantics,
and in particular these semantics MUST not be affected by the Ordered
Aggregate transported by the LSP or by the LSP's Class Type. This
means that if LSP1 contends with LSP2 for resources, LSP1 may preempt
LSP2 if LSP1 has a higher set-up preemption priority (i.e., lower

numerical priority value) than LSP2's holding preemption priority
regardless of LSP1's OA/CT and LSP2's OA/CT.

We introduce the following definition:

    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 set-up priority, as the
                   holding priority or both.

Note that by definition:

- for a given Class-Type, there may be one or multiple
  TE-classes using that Class-Type, each using a different preemption
  priority
- for a given preemption priority, there may be one or multiple
  TE-Class(es) using that preemption priority, each using a different
  Class-Type.

The DS-TE solution MUST allow all LSPs transporting Traffic Trunks of
a given Class-Type to use the same preemption priority. In other
words, the DS-TE solution MUST allow a Class-Type to be used by
single TE-Class. This effectively allows the network administrator
to ensure that no preemption happens within that Class-Type, when so
desired.

As an example, the DS-TE solution MUST allow the network
administrator to define a Class-Type comprising a single TE-class
using preemption 0.

The DS-TE solution MUST allow two LSPs transporting Traffic Trunks of
the same Class-Type to use different preemption priorities, and allow
the LSP with higher (numerically lower) set-up priority to preempt
the LSP with lower (numerically higher) holding priority when they
contend for resources. In other words, the DS-TE solution MUST allow
multiple TE-Classes to be defined for a given Class-Type. This
effectively allows the network administrator to enable preemption
within a Class-Type, when so desired.

As an example, the DS-TE solution MUST allow the network
administrator to define a Class-Type comprising three TE-Classes; one
using preemption 0, one using preemption 1 and one using preemption
4.

The DS-TE solution MUST allow two LSPs transporting Traffic Trunks
from different Class-Types to use different preemption priorities,
and allow the LSP with higher setup priority to preempt the one with
lower holding priority when they contend for resources.

As an example, the DS-TE solution MUST allow the network
administrator to define two Class-Types (CT0 and CT1) each comprising
two TE-Classes where say:

    -one TE-Class groups CT0 and preemption 0
    -one TE-Class groups CT0 and preemption 2
    -one TE-Class groups CT1 and preemption 1
    -one TE-Class groups CT1 and preemption 3

The network administrator would then, in particular, be able to:

- transport a CT0 Traffic Trunk over an LSP with setup priority=0 and
  holding priority=0
- transport a CT0 Traffic Trunk over an LSP with setup priority=2 and
  holding priority=0
- transport a CT1 Traffic Trunk over an LSP with setup priority=1 and
  holding priority=1
- transport a CT1 Traffic Trunk over an LSP with setup priority=3 and
  holding priority=1.

The network administrator would then, in particular, NOT be able to:

- transport a CT0 Traffic Trunk over an LSP with setup priority=1 and
  holding priority=1
- transport a CT1 Traffic Trunk over an LSP with setup priority=0 and
  holding priority=0

The DS-TE solution MUST allow two LSPs transporting Traffic Trunks
from different Class-Types to use the same preemption priority. In
other words, the DS-TE solution MUST allow TE-classes using different
CTs to use the same preemption priority. This effectively allows the
network administrator to ensure that no preemption happens across
Class-Types, if so desired.

As an example, the DS-TE solution MUST allow the network
administrator to define three Class-Types (CT0, CT1 and CT2) each
comprising one TE-Class which uses preemption 0. In that case, no
preemption will ever occur.

Since there are 8 preemption priorities and up to 8 Class-Types,
there could theoretically be up to 64 TE-Classes in a network. This
is felt to be beyond current practical requirements. The current
practical requirement is that the DS-TE solution MUST allow support

for up to 8 TE-classes. The DS-TE solution MUST allow these
TE-classes to comprise any arbitrary subset of 8 (or less) from the
(64) possible combinations of (8) Class-Types and (8) preemption
priorities.

As with existing TE, an LSP which gets preempted is torn down at
preemption time. The Head-end of the preempted LSP may then attempt
to reestablish that LSP, which involves re-computing a path by
Constraint Based Routing based on updated available bandwidth
information and then signaling for LSP establishment along the new
path. It is to be noted that there may be cases where the preempted
LSP cannot be reestablished (e.g., no possible path satisfying LSP
bandwidth constraints as well as other constraints). In such cases,
the Head-end behavior is left to implementation. It may involve
periodic attempts at reestablishing the LSP, relaxing of the LSP
constraints, or other behaviors.

3.5. Mapping of Traffic to LSPs

The DS-TE solution MUST allow operation over E-LSPs onto which a
single <FEC/{TA}PSC> is transported.

The DS-TE solution MUST allow operation over L-LSPs.

The DS-TE solution MAY allow operation over E-LSPs onto which
multiple <FEC/{TA}PSC> of a given FEC are transported, under the
condition that those multiple <FEC/{TA}PSC> can effectively be
treated by DS-TE as a single atomic traffic trunk (in particular this
means that those multiple <FEC/{TA}PSC> are routed as a whole based
on a single collective bandwidth requirement, a single affinity
attribute, a single preemption level, a single Class-Type, etc.). In
that case, it is also assumed that the multiple {TA}PSCs are grouped
together in a consistent manner throughout the DS-TE domain (e.g., if
<FECx/{TA}PSC1> and <FECx/{TA}PSC2> are transported together on an
E-LSP, then there will not be any L-LSP transporting <FECy/{TA}PSC1>
or <FECy/{TA}PSC2> on its own, and there will not be any E-LSP
transporting <FECz/{TA}PSC1> and/or <FECz/{TA}PSC2> with
<FECz/{TA}PSC3>).

3.6. Dynamic Adjustment of Diff-Serv PHBs

As discussed in section 2.2, the DS-TE solution MAY support
adjustment of Diff-Serv PHBs parameters (e.g., queue bandwidth) based
on the amount of TE-LSPs established for each OA/Class-Type. Such
dynamic adjustment is optional for DS-TE implementations.

Where this dynamic adjustment is supported, it MUST allow for
disabling via configuration (thus reverting to PHB treatment with
static scheduler configuration independent of DS-TE operations). It
MAY involve a number of configurable parameters which are outside the
scope of this specification. Those MAY include configurable
parameters controlling how scheduling resources (e.g., service rates)
need to be apportioned across multiple OAs when those belong to the
same Class-Type and are transported together on the same E-LSP.

Where supported, the dynamic adjustment MUST take account of the
performance requirements of each PDB when computing required
adjustments.

3.7. Overbooking

Existing TE mechanisms allow overbooking to be applied on LSPs for
Constraint Based Routing and admission control. Historically, this
has been achieved in TE deployment through factoring overbooking
ratios at the time of sizing the LSP bandwidth and/or at the time of
configuring the Maximum Reservable Bandwidth on links.

The DS-TE solution MUST also allow overbooking and MUST effectively
allow different overbooking ratios to be enforced for different CTs.

The DS-TE solution SHOULD optionally allow the effective overbooking
ratio of a given CT to be tweaked differently in different parts of
the network.

3.8. Restoration

With existing TE, restoration policies use standard priority
mechanisms such as, for example, the preemption priority to
effectively control the order/importance of LSPs for restoration
purposes.

The DS-TE solution MUST ensure that similar application of the use of
standard priority mechanisms for implementation of restoration policy
are not prevented since those are expected to be required for
achieving the survivability requirements of DS-TE networks.

Further discussion of restoration requirements are presented in the
output document of the TEWG Requirements Design Team [SURVIV-REQ].

4. Solution Evaluation Criteria

A range of solutions is possible for the support of the DS-TE
requirements discussed above. For example, some solutions may
require that all current TE protocols syntax (IGP, RSVP-TE,) be

extended in various ways. For instance, current TE protocols could
be modified to support multiple bandwidth constraints rather than the
existing single aggregate bandwidth constraint. Alternatively, other
solutions may keep the existing TE protocols syntax unchanged but
modify their semantics to allow for the multiple bandwidth
constraints.

This section identifies the evaluation criteria that MUST be used to
assess potential DS-TE solutions for selection.

4.1. Satisfying detailed requirements

The solution MUST address all the scenarios described in section 2
and satisfy all the requirements listed in section 3.

4.2. Flexibility

   - number of Class-Types that can be supported, compared to number
     identified in Requirements section
   - number of PDBs within a Class-Type

4.3. Extendibility

   - how far can the solution be extended in the future if requirements
     for more Class-Types are identified in the future.

4.4. Scalability

   - impact on network scalability in what is propagated, processed,
     stored and computed (IGP signaling, IGP processing, IGP database,
     TE-Tunnel signaling ,...).
   - how does scalability impact evolve with number of
     Class-Types/PDBs actually deployed in a network. In particular,
     is it possible to keep overhead small for a large networks which
     only use a small number of
     Class-Types/PDBs, while allowing higher number of
     Class-Types/PDBs in smaller networks which can bear higher
     overhead)

4.5. Backward compatibility/Migration

   - backward compatibility/migration with/from existing TE mechanisms
   - backward compatibility/migration when increasing/decreasing the
     number of Class-Types actually deployed in a given network.

4.6. Bandwidth Constraints Model

   Work is currently in progress to investigate the performance and
   trade-offs of different operational aspects of Bandwidth Constraints
   models (for example see [BC-MODEL], [BC-CONS] and [MAR]). In this
   investigation, at least the following criteria are expected to be
   considered:

        (1) addresses the scenarios in Section 2
        (2) works well under both normal and overload conditions
        (3) applies equally when preemption is either enabled or disabled
        (4) minimizes signaling load processing requirements
        (5) maximizes efficient use of the network
        (6) Minimizes implementation and deployment complexity.

   In selection criteria (2), "normal condition" means that the network
   is attempting to establish a volume of DS-TE LSPs for which it is
   designed; "overload condition" means that the network is attempting
   to establish a volume of DS-TE LSPs beyond the one it is designed
   for; "works well" means that under these conditions, the network
   should be able to sustain the expected performance, e.g., under
   overload it is x times worse than its normal performance.

5. Security Considerations

   The solution developed to address the DS-TE requirements defined in
   this document MUST address security aspects. DS-TE does not raise
   any specific additional security requirements beyond the existing
   security requirements of MPLS TE and Diff-Serv. The solution MUST
   ensure that the existing security mechanisms (including those
   protecting against DOS attacks) of MPLS TE and Diff-Serv are not
   compromised by the protocol/procedure extensions of the DS-TE
   solution or otherwise MUST provide security mechanisms to address
   this.

6. Acknowledgment

   We thank David Allen for his help in aligning with up-to-date
   Diff-Serv terminology.

7. Normative References

   [AF]   Heinanen, J., Baker, F., Weiss, W. and J. Wroclawski,
               "Assured Forwarding PHB Group", RFC 2597, June 1999.

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

   [DIFF-FIELD] Nichols, K., Blake, S., Baker, F. and D. Black,
                "Definition of the Differentiated Services Field (DS
                Field) in the IPv4 and IPv6 Headers", RFC 2474, December
                1998.

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

   [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.

   [DIFF-NEW]   Grossman, D., "New Terminology and Clarifications for
                Diffserv", RFC 3260, April 2002.

   [EF]   Davie, B., Charny, A., Bennet, J.C.R., Benson, K., Le
                Boudec, J.Y., Davari, S., Courtney, W., Firioiu, V. and
                D. Stiliadis, "An Expedited Forwarding PHB (Per-Hop
                Behavior)", RFC 3246, March 2002.

   [TEWG-FW]   Awduche, D., Chiu, A., Elwalid, A., Widjaja, I. and X.
                Xiao, "Overview and Principles of Internet Traffic
                Engineering", RFC 3272, May 2002.

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

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

8.   Informative References

   [DIFF-PDB]   Nichols, K. and B. Carpenter, "Definition of
                Differentiated Services Per Domain Behaviors and Rules
                for their Specification", RFC 3086, April 2001.

   [ISIS-TE]   Smit, Li, "IS-IS extensions for Traffic Engineering",
                Work in Progress, December 2002.

   [OSPF-TE]   Katz, et al., "Traffic Engineering Extensions to OSPF",
                Work in Progress, October 2002.

   [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.

   [SURVIV-REQ] Lai, W. and D. McDysan, "Network Hierarchy and
                Multilayer Survivability", RFC 3386, November 2002.

   [BC-MODEL]   Lai, W., "Bandwidth Constraints Models for
                Diffserv-aware MPLS Traffic Engineering: Performance
                Evaluation", Work in Progress, June 2002.

   [BC-CONS]   F. Le Faucheur, "Considerations on Bandwidth Constraints
                Models for DS-TE", Work in Progress, June 2002.

   [MAR]   Ash, J., "Max Allocation with Reservation Bandwidth
                Constraint Model for MPLS/DiffServ TE & Performance
                Comparisons", Work in Progress, May 2003.

9. Contributing Authors

   This document was the collective work of several people. The text
   and content of this document was contributed by the editors and the
   co-authors listed below. (The contact information for the editors
   appears below.)

   Martin Tatham   Thomas Telkamp
   BT   Global Crossing
   Adastral Park, Martlesham Heath, Oudkerkhof 51, 3512 GJ Utrecht
   Ipswich IP5 3RE, UK   The Netherlands
   Phone: +44-1473-606349   Phone: +31 30 238 1250
   EMail: martin.tatham@bt.com   EMail: telkamp@gblx.net

   David Cooper   Jim Boyle
   Global Crossing   Protocol Driven Networks, Inc.
   960 Hamlin Court   1381 Kildaire Farm Road #288
   Sunnyvale, CA 94089, USA   Cary, NC 27511, USA
   Phone: (916) 415-0437   Phone: (919) 852-5160
   EMail: dcooper@gblx.net   EMail: jboyle@pdnets.com

   Luyuan Fang   Gerald R. Ash
   AT&T Labs   AT&T Labs
   200 Laurel Avenue   200 Laurel Avenue
   Middletown, New Jersey 07748, USA Middletown, New Jersey 07748,USA
   Phone: (732) 420-1921   Phone: (732) 420-4578
   EMail: luyuanfang@att.com   EMail: gash@att.com

Pete Hicks   Angela Chiu
CoreExpress, Inc   AT&T Labs-Research
12655 Olive Blvd, Suite 500   200 Laurel Ave. Rm A5-1F13
St. Louis, MO 63141, USA   Middletown, NJ 07748, USA
Phone: (314) 317-7504   Phone: (732) 420-9061
EMail: pete.hicks@coreexpress.net EMail: chiu@research.att.com

William Townsend   Thomas D. Nadeau
Tenor Networks   Cisco Systems, Inc.
100 Nagog Park   300 Beaver Brook Road
Acton, MA 01720, USA   Boxborough, MA 01719
Phone: +1 978-264-4900   Phone: +1-978-936-1470
EMail:btownsend@tenornetworks.com EMail: tnadeau@cisco.com

Darek Skalecki
Nortel Networks
3500 Carling Ave,
Nepean K2H 8E9,
Phone: (613) 765-2252
EMail: dareks@nortelnetworks.com

10. Editors' Addresses

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

Wai Sum Lai
AT&T Labs
200 Laurel Avenue
Middletown, New Jersey 07748, USA

Phone: (732) 420-3712
EMail: wlai@att.com

11. Full Copyright Statement

   Copyright (C) The Internet Society (2003). All Rights Reserved.

Acknowledgement