Francois Le Faucheur, Editor
Cisco Systems, Inc.

IETF Internet Draft
Expires: August, 2003
Document: draft-ietf-tewg-diff-te-russian-01.txt   February, 2003

Russian Dolls Bandwidth Constraints Model for
Diff-Serv-aware MPLS Traffic Engineering

Status of this Memo

  This document is an Internet-Draft and is in full conformance with
  all provisions of Section 10 of RFC2026. Internet-Drafts are
  Working documents of the Internet Engineering Task Force (IETF), its
  areas, and its working groups. Note that other groups may also
  distribute working documents as Internet-Drafts.

  Internet-Drafts are draft documents valid for a maximum of six months
  and may be updated, replaced, or obsoleted by other documents at any
  time. It is inappropriate to use Internet-Drafts as reference
  material or to cite them other than as "work in progress."

  The list of current Internet-Drafts can be accessed at
  http://www.ietf.org/ietf/1id-abstracts.txt.
  The list of Internet-Draft Shadow Directories can be accessed at
  http://www.ietf.org/shadow.html.

Abstract

  This document provides specification for one Bandwidth Constraints
  model for Diff-Serv-aware MPLS Traffic Engineering, which is referred
  to as the Russian Dolls Model.

Summary for Sub-IP related Internet Drafts

  RELATED DOCUMENTS:
  draft-ietf-tewg-diff-te-reqts-06.txt
  draft-ietf-tewg-diff-te-proto-02.txt

  WHERE DOES IT FIT IN THE PICTURE OF THE SUB-IP WORK
  This ID is a Working Group document of the TE Working Group.

WHY IS IT TARGETED AT THIS WG(s)
TEWG is responsible for specifying protocol extensions for support of
Diff-Serv-aware MPLS Traffic Engineering.

    JUSTIFICATION
    The TEWG charter states that "This will entail verification and
    review of the Diffserv requirements in the WG Framework document and
    initial specification of how these requirements can be met through
    use and potentially expansion of existing protocols."
    In line with this, the TEWG is specifying bandwidth constraints model
    for Diff-Serv-aware MPLS Traffic Engineering. This document describes
    one particular bandwidth constraints model.

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].

1.   Introduction

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

    [DSTE-REQ] also defines the concept of Bandwidth Constraint Models
    for DS-TE and states that "The DS-TE technical solution MUST specify
    at least one bandwidth constraint model and MAY specify multiple
    bandwidth constraint."

    This document provides a detailed description of one particular
    Bandwidth Constraint model for DS-TE which is introduced in [DSTE-
    REQ] and called the Russian Dolls Model (RDM).

[DSTE-PROTO] specifies the IGP and RSVP-TE signaling extensions for support of DS-TE. These extensions support RDM.

2. Contributing Authors

This document was the collective work of several. The text and content of this document was contributed by the editor and the co-authors listed below. (The contact information for the editor appears in Section 11, and is not repeated below.)

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:

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

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

The following definition from [DSTE-PROTO] is also repeated here:

Normalised(CTc) : let us define "Normalised(CTc)" as
"Reserved(CTc)/LOM(c)", where LOM (c) is the Local Overbooking
Multiplier for CTc defined in [DSTE-PROTO].

We also introduce the following definitions:

Reserved(CTb,q) : let us define "Reserved(CTb,q)" as the sum of the
bandwidth reserved by all established LSPs which belong to CTb and
have a holding priority of q. Note that if q and CTb do not form one
of the 8 possible configured TE-Classes, then there can not be any

Le Faucheur et. al   3

established LSP which belong to CTb and have a holding priority of q,
so in that case Reserved(CTb,q)=0.

Normalised(CTc,q) : let us define "Normalised(CTc,q)" as

"Reserved(CTc/q) / LOM(c)", where LOM (c) is the Local Overbooking
Multiplier for CTc defined in [DSTE-PROTO].

4.   Russian Dolls Model Definition

RDM is defined in the following manner (assuming for now that the
optional per-CT Local Overbooking Multipliers defined in [DSTE-PROTO]
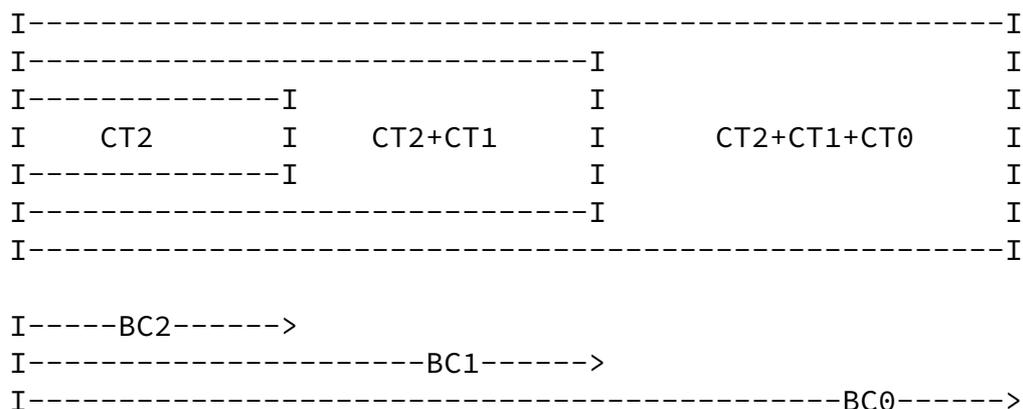are not used – i.e. LOM[c]=1 , 0<=c<=7):
            o Maximum Number of Bandwidth Constraints (MaxBC)= Maximum
              Number of Class-Types (MaxCT) = 8
            o for each value of b in the range 0 <= b <= (MaxCT – 1):
                  SUM (Reserved (CTc)) <= BCb,
                  for all "c" in the range b <= c <= (MaxCT – 1)

A DS-TE LSR implementing RDM MUST support enforcement of bandwidth
constraints in compliance with this definition.

Where 8 Class-Types are active, the RDM bandwidth constraints can
also be expressed in the following way:
       - All LSPs from CT7 use no more than BC7
       - All LSPs from CT6 and CT7 use no more than BC6
       - All LSPs from CT5, CT6 and CT7 use no more than BC5
       - etc.
       - All LSPs from CT0, CT1,... CT7 use no more than BC0

Purely for illustration purposes, the diagram below represents the
Russian Doll Bandwidth Constraints model in a pictorial manner when 3
Class-Types are active:

```
I-------------------------------------------------------I
I----------------------------I                          I
I-------------I              I                          I
I   CT2       I   CT2+CT1    I      CT2+CT1+CT0          I
I-------------I              I                          I
I----------------------------I                          I
I-------------------------------------------------------I


I-----BC2------>
I--------------------BC1------>
I-----------------------------------------BC0------>
```


While simpler Bandwidth Constraints models (see [MAM]) or,
conversely, more flexible/sophisticated Bandwidth Constraints models
can be defined, the Russian Dolls Modelis attractive in some DS-TE
environments for the following reasons:

            - Although less intuitive than MAM, RDM is still a simple model
               to conceptualize.
            - RDM can be used in conjunction with preemption to ensure
               strict isolation across Class-Types, so that each Class-Type
               is guaranteed its share of bandwidth no matter the level of
               contention by other classes, whether preemption is used or
               not.
            - RDM can simultaneously achieve isolation across Class-Types,
               bandwidth efficiency and protection against QoS degradation
               of all Class-Types.
            - MAM only requires limited protocol extensions such as the
               ones defined in [DSTE-PROTO].

      RDM may not be attractive in some DS-TE environments for the
      following reasons:
            - if the usage of preemption is precluded for some
               administrative reason, while RDM can still ensure bandwidth
               efficiency and protection against QoS degradation of all CTs,
               RDM cannot guarantee isolation.

      Additional considerations on the properties of RDM can be found in
      [BC-CONS] and [BC-MODEL].

      As a simple example usage of the "Russian Doll" Bandwidth Constraints
      Model, a network administrator using one CT for Voice (CT1) and one
      CT for data (CT0) might configure on a given link:
            - BC0 = 2.5 Gb/s (i.e. Voice + Data is limited to 2.5 Gb/s)
            - BC1= 1.5 Gb/s (i.e. Voice is limited to 1.5 Gb/s).

5.   Example Formulas for Computing "Unreserved TE-Class [i]" with
      Russian Dolls Model

      As specified in [DSTE-PROTO], formulas for computing "Unreserved TE-
      Class [i]" MUST reflect all of the Bandwidth Constraints relevant to
      the CT associated with TE-Class[i], and thus, depend on the Bandwidth
      Constraints Model. Thus, a DS-TE LSR implementing MAM MUST reflect
      the MAM bandwidth constraints defined in section 4 above when
      computing "Unreserved TE-Class [i]".

Keeping in mind, as explained in [DSTE-PROTO], that details of
admission control algorithms as well as formulas for computing
"Unreserved TE-Class [i]" are outside the scope of the IETF work, we
provide in this section, for illustration purposes, an example of how
values for the unreserved bandwidth for TE-Class[i] might be computed
with RDM, assuming:
>   - the basic admission control algorithm which simply deducts
>   the exact bandwidth of any established LSP from all of the
>   Bandwidth Constraints relevant to the CT associated with that
>   LSP.
>   - the optional per-CT Local Overbooking Multipliers are not
>   used (.i.e. LOM[c]=1, 0<= c <=7).

Le Faucheur et. al   5

We assume that:
     TE-Class [i] <--> < CTc , preemption p>
in the configured TE-Class mapping.

For readability, formulas are first shown assuming only 4 CTs are
active. The formulas are then extended to cover the cases where more
CTs are used.

If CTc = CT0, then "Unreserved TE-Class [i]" =
     [BC0 - SUM (Reserved(CTb,q))] for q <= p and 0 <= b <= 3

If CTc = CT1, then "Unreserved TE-Class [i]" =
     MIN [
     [BC1 - SUM (Reserved(CTb,q))] for q <= p and 1 <= b <= 3,
     [BC0 - SUM (Reserved(CTb,q))] for q <= p and 0 <= b <= 3
          ]

If CTc = CT2, then "Unreserved TE-Class [i]" =
     MIN [
     [BC2 - SUM (Reserved(CTb,q))] for q <= p and 2 <= b <= 3,
     [BC1 - SUM (Reserved(CTb,q))] for q <= p and 1 <= b <= 3,
     [BC0 - SUM (Reserved(CTb,q))] for q <= p and 0 <= b <= 3

```
          ]


    If CTc = CT3, then "Unreserved TE-Class [i]" =
          MIN   [
          [ BC3 - SUM ( Reserved(CTb,q) ) ] for q <= p and 3 <= b <= 3,
          [ BC2 - SUM ( Reserved(CTb,q) ) ] for q <= p and 2 <= b <= 3,
          [ BC1 - SUM ( Reserved(CTb,q) ) ] for q <= p and 1 <= b <= 3,
          [ BC0 - SUM ( Reserved(CTb,q) ) ] for q <= p and 0 <= b <= 3
              ]


    The formula can be generalized to 8 active CTs and expressed in a
    more compact way in the following:

           "Unreserved TE-Class [i]" =
          MIN   [
          [ BCc - SUM ( Reserved(CTb,q) ) ] for q <= p and c <= b <= 7,
          . . .
          [ BC0 - SUM ( Reserved(CTb,q) ) ] for q <= p and 0 <= b <= 7,
              ]

    where:
          TE-Class [i] <--> < CTc , preemption p>
    in the configured TE-Class mapping.
```

6.   Support of Optional Local Overbooking Method

   We remind the reader that, as discussed in [DSTE-PROTO], the
   "LSP/link size overbooking" method (which does not use the Local
   Overbooking Multipliers - LOMs-) is expected to be sufficient in many
   DS-TE environments. It is expected that the optional Local
   Overbooking method (and LOMs) would only be used in specific
   environments, in particular where different overbooking ratios need
   to be enforced on different links of the DS-TE domain and cross-
   effect of overbooking across CTs needs to be accounted for very

accurately.

This section discusses the impact of the optional local overbooking
method on RDM and associated rules and formula. This is only
applicable in the cases where the optional local overbooking method
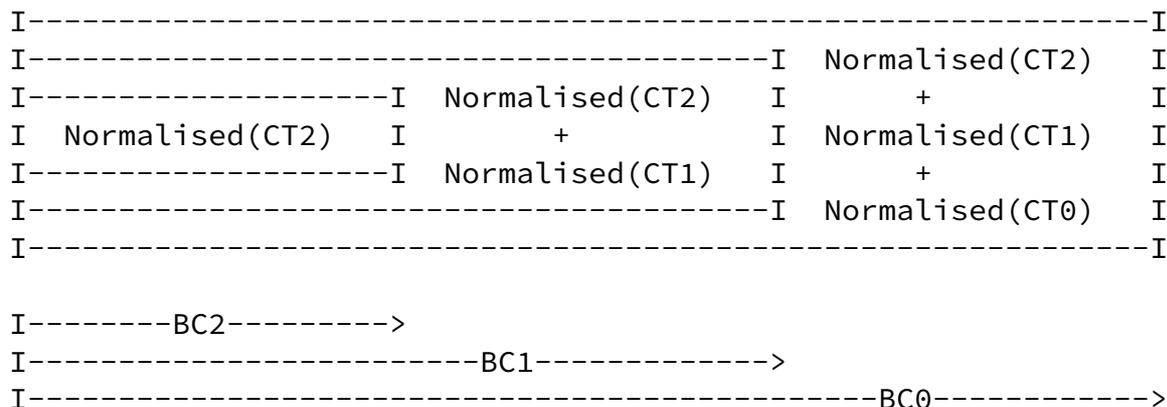is indeed supported by the DS-TE LSRs and actually deployed.

6.1.   Russian Dolls Model Definition With Local Overbooking

As specified in [DSTE-PROTO], when the optional Local Overbooking
method is supported, the bandwidth constraints MUST be applied to
"Normalised(CTc)" rather than to "Reserved(CTc)". Thus, when the
optional Local Overbooking method is supported, the Russian Doll
Model definition is extended in the following manner:
          o Maximum Number of Bandwidth Constraints (MaxBC)=
             Maximum Number of Class-Types (MaxCT) = 8
          o for each value of b in the range 0 <= b <= (MaxCT – 1):
                SUM (Normalised (CTc)) <= BCb,
                for all "c" in the range b <= c <= (MaxCT – 1)

A DS-TE LSR implementing RDM and implementing the optional Local
Overbooking method MUST support enforcement of bandwidth constraints
in compliance with this extended definition.

Purely for illustration purposes, the diagram below represents the
Russian Doll Bandwidth Constraints model in a pictorial manner when 3
Class-Types are active and the local overbooking method is used:

```
I----------------------------------------------------------I
I-----------------------------------------I  Normalised(CT2)   I
I-------------------I  Normalised(CT2)   I       +           I
I  Normalised(CT2)   I         +          I  Normalised(CT1)   I
I-------------------I  Normalised(CT1)   I       +           I
I-----------------------------------------I  Normalised(CT0)   I
I----------------------------------------------------------I


I--------BC2--------->
I----------------------BC1------------->
I-------------------------------------------BC0----------->
```

6.2.   Example Formulas for Computing "Unreserved TE-Class [i]" With
        Local Overbooking

   A DS-TE LSR implementing RDM and implementing the optional Local
   Overbooking method MUST reflect the RDM bandwidth constraints defined
   in section 6.1 above when computing "Unreserved TE-Class [i]".

   Again, keeping in mind that details of admission control algorithms
   as well as formulas for computing "Unreserved TE-Class [i]" are
   outside the scope of the IETF work, we provide in this section, for
   illustration purposes, an example of how values for the unreserved
   bandwidth for TE-Class[i] might be computed with the Russian Dolls
   Model, assuming:
        - the basic admission control algorithm which simply deducts
          the exact bandwidth of any established LSP from all of the
          Bandwidth Constraints relevant to the CT associated with that
          LSP.
        - the optional per-CT Local Overbooking Multipliers are used.

   When the optional local overbooking method is supported, the example
   generalized formula of section 5 becomes:

        "Unreserved TE-Class [i]" =
        LOM(c) x MIN [
        [BCc - SUM (Normalised(CTb,q))] for q <= p and c <= b <= 7,
        . . .
        [BC0 - SUM (Normalised(CTb,q))] for q <= p and 0 <= b <= 7,
            ]

   where:
        - TE-Class [i] <--> < CTc , preemption p>
          in the configured TE-Class mapping.

6.3.   Example Usage of LOM

   To illustrate usage of the local overbooking method with the Russian
   Dolls model, let's consider a DS-TE deployment where two CTs (CT0 for
   data and CT1 for voice) and a single preemption priority are used.

   The TE-Class mapping is the following:

        TE-Class <--> CT, preemption
        ==============================
            0   CT0, 0
            1   CT1, 0

```
         rest           unused
```

Let's assume that on a given link, BCs and LOMs are configured in the
following way:
```
     BC0 = 200
     BC1 = 100
```

```
     LOM(0) = 4  (i.e. = 400%)
     LOM(1) = 2  (i.e. = 200%)
```

Let's further assume that the DS-TE LSR uses the example formulas
presented above for computing unreserved bandwidth values.

If there is no established LSP on the considered link, the LSR will
advertise for that link in IGP :
```
        Unreserved TE-Class [0] = 4 x (200 - 0/4 - 0/2 )= 800
        Unreserved TE-Class [1] = 2 x (100- 0/2) = 200
```
Note again that these values advertised for Unreserved Bandwidth are
larger than BC1 and BC0.

If there is only a single established LSP, with CT=CT0 and BW=100,
the LSR will advertise:
```
        Unreserved TE-Class [0] = 4 x (200 - 100/4 - 0/2 )=700
        Unreserved TE-Class [1] = 2 x (100- 0/2) = 200
```

If there is only a single established LSP, with CT=CT1 and BW=100,
the LSR will advertise:
```
        Unreserved TE-Class [0] = 4 x (200 - 0/4 - 100/2 )= 600
        Unreserved TE-Class [1] = 2 x (100- 100/2) = 100
```
Note that the impact of an LSP on the unreserved bandwidth of a CT
does not depend only on the LOM for that CT: it also depends on the
LOM for the CT of the LSP. This can be seen in our example. A BW=100
tunnel affects Unreserved
CT0 twice as much if it is a CT1 tunnel, than if it is a CT0 tunnel.
It reduces Unreserved CTO by 200 (800->600) rather than by 100
(800->700). This is because LOM(1) is half as big as LOM(0). This
illustrates why the local overbooking method allows very fine
accounting of cross-effect of overbooking across CTs, as compared

with the LSP/link size overbooking method.

If there are two established LSPs, one with CT=CT1 and BW=100 and one
with CT=CT0 and BW=100, the LSR will advertise:
      Unreserved TE-Class [0] = 4 x (200 - 100/4 - 100/2) = 500
      Unreserved TE-Class [1] = 2 x (100 - 100/2) = 100

If there are two LSPs established, one with CT=CT1 and BW=100, and
one with CT=CT0 and BW=480, the LSR will advertise:
          Unreserved TE-Class [0] = 4 x (200 - 480/4 - 100/2) = 120
          Unreserved TE-Class [1] = 2 x MIN [(200 - 480/4 - 100/2),
                                              (100 - 100/2)]
                                  = 2 x MIN [30, 50]
                                  = 60

7.   Security Considerations

   Security considerations related to the use of DS-TE are discussed in
   [DSTE-PROTO]. Those apply independently of the Bandwidth Constraints
   model, including RDM specified in this document.

8.   Acknowledgments

   We thank Martin Tatham for his earlier contribution in this work.

9.   Normative References

   [DSTE-REQ] Le Faucheur et al, Requirements for support of Diff-Serv-
   aware MPLS Traffic Engineering, draft-ietf-tewg-diff-te-reqts-07.txt,
   February 2003.

   [DSTE-PROTO] Le Faucheur et al, Protocol extensions for support of
   Diff-Serv-aware MPLS Traffic Engineering, draft-ietf-tewg-diff-te-

proto-03.txt, February 2003.

   [RFC2119] S. Bradner, Key words for use in RFCs to Indicate
   Requirement Levels, RFC2119, March 1997.

10.   Informative References

   [BC-CONS] Le Faucheur, "Considerations on Bandwidth Constraints Model
   for DS-TE", draft-lefaucheur-tewg-russian-dolls-00.txt, June 2002.

   [BC-MODEL] Lai, "Bandwidth Constraints Models for DS-TE",
   draft-wlai-tewg-bcmodel-00.txt, June 2002.

   [MAM] Le Faucheur, "Maximum Allocation Bandwidth Constraints Model
   for Diff-Serv-aware MPLS Traffic Engineering", draft-lefaucheur-diff-
   tet-mam-00.txt, February 2003.

   [OSPF-TE] Katz et al., "Traffic Engineering Extensions to OSPF",
   draft-katz-yeung-ospf-traffic-09.txt, October 2002.

   [ISIS-TE] Smit et al., "IS-IS extensions for Traffic Engineering",
   draft-ietf-isis-traffic-04.txt, December 2002.

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

   [DIFF-MPLS] Le Faucheur et al, "MPLS Support of Diff-Serv", RFC3270,
   May 2002.

11.   11. Intellectual Property Considerations

   Cisco Systems, Inc. may seek patent or other intellectual property
   protection for some of all of the technologies disclosed in this
   document. If any standards arising from this document are or become

 Le Faucheur et. al   10

                  Russian Dolls model for DS-TE   February 2003

   protected by one or more patents assigned to Cisco Systems, Cisco

Systems intends to disclose those patents and license them on
reasonable and non-discriminatory terms.

12.   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

Appendix A - Addressing [DSTE-REQ] Scenarios

   This Appendix provides examples of how the Russian Dolls Bandwidth
   Constraints model can be used to support each of the scenarios
   described in [DSTE-REQ].

1.   Scenario 1: Limiting Amount of Voice

   By configuring on every link:
        - Bandwidth Constraint 1 (for CT1=Voice) = "certain percentage"
          of link capacity
        - BC0 (for CT1=Voice + CT0= Data) = link capacity

   By configuring:
        - every CT1/Voice TE-LSP with preemption =0
        - every CT0/Data TE-LSP with preemption =1

   DS-TE with the Russian Dolls Model will address all the requirements:
        - amount of Voice traffic limited to desired percentage on
          every link
        - data traffic capable of using all remaining link capacity
        - voice traffic capable of preempting other traffic

2.   Scenario 2: Maintain Relative Proportion of Traffic Classes

   By configuring on every link:
        - BC2 (for CT2) = e.g. 45%
        - BC1 (for CT1+CT2) = e.g. 80%
        - BC0 (for CT0+CT1+CT2) = e.g.100%

   DS-TE with the Russian Dolls Model will ensure that the amount of
   traffic of each Class Type established on a link is within acceptable
   levels as compared to the resources allocated to the corresponding
   Diff-Serv PHBs regardless of which order the LSPs are routed in,
   regardless of which preemption priorities are used by which LSPs and

regardless of failure situations. Optional automatic adjustment of
Diff-Serv scheduling configuration could be used for maintaining very
strict relationship between amount of established traffic of each
Class Type and corresponding Diff-Serv resources.

[3]. Scenario 3: Guaranteed Bandwidth Services

By configuring on every link:
- BC1 (for CT1) = "given" percentage of link bandwidth
  (appropriate to achieve the Guaranteed Bandwidth service's
  QoS objectives)
- BC0 (for CT0+CT1) = 100% of link bandwidth

DS-TE with the Russian Dolls Model will ensure that the amount of
Guaranteed Bandwidth Traffic established on every link remains below
the given percentage so that it will always meet its QoS objectives.
At the same time it will allow traffic engineering of the rest of the
traffic such that links can be filled up.