Francois Le Faucheur, Editor
Cisco Systems, Inc.

IETF Internet Draft
Expires: April, 2002
Document: draft-ietf-tewg-diff-te-russian-00.txt October, 2002

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

Status of this Memo

Abstract

This document provides a detailed description of a candidate 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 in the process of selecting bandwidth constraints model for Diff-Serv-aware MPLS Traffic Engineering. This document describes one bandwidth constraints model under consideration.

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 one default bandwidth constraint model which must be supported by any DS-TE implementation. However, additional bandwidth constraint models may also be specified."

This document provides a more detailed description of one particular Bandwidth Constraint model introduced in [DSTE-REQ] which is called the Russian Dolls model. The Russian Dolls model is one of the candidate models under consideration by the TEWG for selection of the default bandwidth constraints model and/or of optional additional models.

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

Considerations on the Russian Dolls model are provided in [RUSSIAN-CONS] and [BCMODEL].

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.

4. Russian Dolls Model Definition

The "Russian Dolls" model 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 All LSPs supporting Traffic Trunks from CTc (with b<=c<=7) use no more than BCb i.e.:
    - 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 or, conversely, more flexible/sophisticated Bandwidth Constraints models can be defined, the Russian Dolls model is an attractive trade-off for the following reasons:

- Network administrators generally find it superior to the basic Maximum Allocation model consisting of a single independent BC per CT (which, in typical deployment scenarios may result in either capacity wastage, low priority Traffic Trunk starvation and/or degradation of QoS objectives).
- network administrators generally find it sufficient for the real life deployments currently anticipated (e.g. it addresses all the scenarios described in [DSTE-REQ] as illustrated in Appendix A below).
- it remains simple and only requires limited protocol extensions, while more sophisticated Bandwidth Constraints model may require more complex extensions.

More details on the properties of the Russian Dolls model can be found in [RUSSIAN-CONS] and [BCMODEL].

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.

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 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 not used (.i.e. LOM[c]=1, 0<= c <=7).

We assume that:

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

in the configured TE-Class mapping.

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

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 the Russian Dolls model 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

Let us define "Reserved(CTc)" as the sum of the bandwidth reserved by all established LSPs which belong to CTc.

Let us define "Normalised(CTc)" as "Reserved(CTc)/LOM(c)".

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 MUST be extended in the following manner:

- Maximum Number of Bandwidth Constraints (MaxBC)= Maximum Number of Class-Types (MaxCT) = 8
- SUM [Normalised(CTc), for b<=c<=7] <= BCb i.e.:


```
o [ Normalised(CT7) ] <= BC7
o [ Normalised(CT6) + Normalised(CT7) ] <= BC6
o [ Normalised(CT5) + Normalised(CT6) +
    Normalised(CT7) ] <= BC5
o etc.
o [ Normalised(CT0) + Normalised(CT1) +
    ...
    Normalised(CT6) + Normalised(CT7) ] <= 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 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

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.
- Normalised(CTb,q) = Reserved(CTb/q)/LOM(b)

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

No new security considerations are raised by this document; they are the same as in the DS-TE requirements document [DSTE-REQ].

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-06.txt, September 2002.

[DSTE-PROTO] Le Faucheur et al, Protocol extensions for support of Diff-Serv-aware MPLS Traffic Engineering, draft-ietf-tewg-diff-te-proto-02.txt, October 2002.

10. Informative References

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

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

[OSPF-TE] Katz, Yeung, Traffic Engineering Extensions to OSPF, draft-katz-yeung-ospf-traffic-08.txt, September 2002.

[ISIS-TE] Smit, Li, IS-IS extensions for Traffic Engineering, draft-ietf-isis-traffic-04.txt, August 2001.

[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. 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 Model 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-Sev 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 Trafic 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.