Network Working Group
Request for Comments: 4127
Category: Experimental

F. Le Faucheur, Ed.
Cisco Systems, Inc.

June 2005

Russian Dolls Bandwidth Constraints Model for Diffserv-aware MPLS Traffic Engineering

Status of This Memo

Copyright Notice

Abstract

This document provides specifications for one Bandwidth Constraints Model for Diffserv-aware MPLS Traffic Engineering, which is referred to as the Russian Dolls Model.

Table of Contents

1. Introduction

[DSTE-REQ] presents the Service Providers requirements for support of 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.

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

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

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

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:

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.

A number of recovery mechanisms under investigation or specification in the IETF take advantage of the concept of bandwidth sharing across particular sets of LSPs. "Shared Mesh Restoration" in [GMPLS-RECOV] and "Facility-based Computation Model" in [MPLS-BACKUP] are example mechanisms that increase bandwidth efficiency by sharing bandwidth across backup LSPs protecting against independent failures. To ensure that the notion of "Reserved (CTc)" introduced in [DSTE-REQ] is compatible with such a concept of bandwidth sharing across multiple LSPs, the wording of the "Reserved (CTc)" definition provided in [DSTE-REQ] is generalized into the following:

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

With this generalization, the Russian Dolls Model definition provided in this document is compatible with Shared Mesh Restoration defined in [GMPLS-RECOV], so that DS-TE and Shared Mesh Protection can operate simultaneously. This assumes that Shared Mesh Restoration operates independently within each DS-TE Class-Type and does not operate across Class-Types (for example, backup LSPs protecting Primary LSPs of CTx also need to belong to CTx; Excess Traffic LSPs sharing bandwidth with Backup LSPs of CTx also need to belong to CTx).

We also introduce the following definition:

Reserved(CTb,q): Let us define "Reserved(CTb,q)" as the total amount of the bandwidth reserved by all the established LSPs that 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 cannot be any established LSPs that belongs to CTb and has a holding priority of q; therefore, in this case, Reserved(CTb,q) = 0.

4. Russian Dolls Model Definition

RDM is defined in the following manner:

- Maximum Number of Bandwidth Constraints (MaxBC)=
  Maximum Number of Class-Types (MaxCT) = 8

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

- BC0= Maximum Reservable Bandwidth, so that
  SUM (Reserved(CTc)) <= Max-Reservable-Bw,
  where the SUM is across all values of c in the
  range 0 <= c <= (MaxCT - 1)

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

Both preemption within a CT and across CTs is allowed.

Where 8 CTs are active, the RDM Bandwidth Constraints can also be expressed in the following way:

- All LSPs from CT7 use no more than BC7

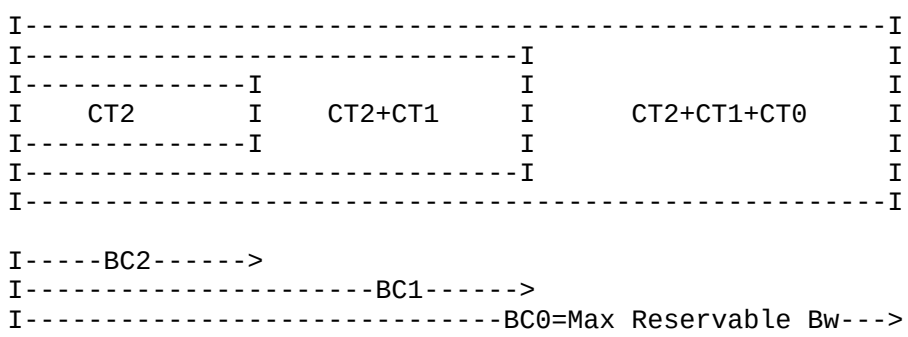
- All LSPs from CT6 and CT7 use no more than BC6

- All LSPs from CT5, CT6 and CT7 use no more than BC5

- etc.

- All LSPs from CT0, CT1, ..., CT7 use no more than BC0 = "Maximum Reservable Bandwidth"

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



While simpler Bandwidth Constraints models or, conversely, more flexible/sophisticated Bandwidth Constraints models can be defined, the Russian Dolls Model is attractive in some DS-TE environments for the following reasons:

- Although it is a little less intuitive than the Maximum Allocation Model (see [DSTE-MAM]), RDM is still a simple model to conceptualize.

- RDM can be used simultaneously to ensure bandwidth efficiency and to protect against QoS degradation of all CTs, whether preemption is used or not.

- RDM can be used in conjunction with preemption to simultaneously achieve (i) isolation across CTs (so that each CT is guaranteed its share of bandwidth no matter the level of contention by other classes), (ii) bandwidth efficiency, and (iii) protection against QoS degradation of all CTs.

- RDM 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 across Class-Types.

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

As a simple example usage of the "Russian Dolls" 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 = Max-Reservable - Bw = 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 RDM MUST reflect the RDM Bandwidth Constraints defined in section 4 above when computing "Unreserved TE-Class [i]".

As explained in [DSTE-PROTO], the details of admission control algorithms, as well as formulas for computing "Unreserved TE-Class [i]", are outside the scope of the IETF work. Keeping that in mind, we provide in this section an example for illustration purposes, of how values for the unreserved bandwidth for TE-Class[i] might be computed with RDM. In the example, we assume 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.

We assume that:

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

in the configured TE-Class mapping.

For readability, formulas are first shown assuming only 3 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 <= 2

If CTc = CT1, then "Unreserved TE-Class [i]" =
MIN [
[BC1 - SUM (Reserved(CTb,q))] for q <= p and 1 <= b <= 2,
[BC0 - SUM (Reserved(CTb,q))] for q <= p and 0 <= b <= 2
]

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

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,
 [ BC(c-1) - SUM ( Reserved(CTb,q) ) ] for q <= p and (c-1)<= 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. Receiving Both Maximum Reservable Bandwidth and Bandwidth Constraints sub-TLVs

[DSTE-PROTO] states that "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."

With RDM, BC0 is equal to the Maximum Reservable Bandwidth because they both represent the aggregate constraint across all CTs. Thus, a DS-TE LSR, receiving both the "Maximum Reservable Bw" sub-TLV and the new "Bandwidth Constraints" sub-TLV (which contains BC0) for a given link where the RDM model is used, MAY ignore the "Maximum Reservable Bw" sub-TLV.

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. IANA Considerations

[DSTE-PROTO] defines a new name space for "Bandwidth Constraints Model Id". The guidelines for allocation of values in that name space are detailed in section 13.1 of [DSTE-PROTO]. In accordance

with these guidelines, the IANA has assigned a Bandwidth Constraints Model Id for RDM from the range 0-239 (which is to be managed as per the "Specification Required" policy defined in [IANA-CONS]).

Bandwidth Constraints Model Id 0 was allocated by IANA to RDM.

9. Acknowledgements

We thank Martin Tatham for his key contribution in this work. Tatiana Renko is also warmly thanked for her instantiation of the Russian Doll.

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].

A.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

A.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 RDM will ensure that the amount of traffic of each CT established on a link is within acceptable levels as compared to the resources allocated to the corresponding Diffserv Per Hop Behaviors (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.

By also configuring:

- every CT2/Voice TE-LSP with preemption = 0

- every CT1/Premium Data TE-LSP with preemption = 1

- every CT0/Best-Effort TE-LSP with preemption = 2

DS-TE with the Russian Dolls Model will also ensure that:

- CT2 Voice LSPs always have first preemption priority in order to use the CT2 capacity

- CT1 Premium Data LSPs always have second preemption priority in order to use the CT1 capacity

- Best-Effort can use up to link capacity of what is left by CT2 and CT1.

Optional automatic adjustment of Diffserv scheduling configuration could be used for maintaining very strict relationships between the amounts of established traffic of each Class Type and corresponding Diffserv resources.

A.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.

Normative References

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

[DSTE-PROTO] Le Faucheur, F., Ed., "Protocol Extensions for Support of Diffserv-aware MPLS Traffic Engineering", RFC 4124, June 2005.

[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

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

[BC-MODEL] Lai, W., "Bandwidth Constraints Models for Differentiated Services (Diffserv)-aware MPLS Traffic Engineering: Performance Evaluation", RFC 4128, June 2005.

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

[GMPLS-RECOV] Lang, et al., "Generalized MPLS Recovery Functional Specification", Work in Progress.

[MPLS-BACKUP] Vasseur, et al., "MPLS Traffic Engineering Fast Reroute: Bypass Tunnel Path Computation for Bandwidth Protection", Work in Progress.

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

Full Copyright Statement

Intellectual Property

Acknowledgement

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