TOC

Internet Engineering Task Force	J. Newton
Internet-Draft	Cable&Wireless
Intended status: BCP	M. Aissaoui
Expires: December 29, 2008	Alcatel-Lucent
	JP. Vasseur
	Cisco Systems, Inc.
	June 27, 2008

A Diffserv-TE Implementation Model to dynamically change booking factors during failure events
draft-newton-mpls-te-dynamic-overbooking-00.txt

Status of this Memo

Abstract

This document discusses the requirements for and describes an implementation model of Diffserv-TE that allows the booking factors applied to network resources to be dynamically changed during network failure events.

Table of Contents

1. Introduction

TOC

The IETF has developed RSVP-TE (Awduche, D., Berger, L., Gan, D., Li, T., Srinivasan, V., and G. Swallow, “RSVP-TE: Extensions to RSVP for LSP Tunnels,” December 2001.) [RFC3209] to provide the capability of signalling MPLS Traffic Engineered LSPs that reserve resources from the network. This was further developed with Diffserv-aware MPLS Traffic Engineering (Le Faucheur, F., “Protocol Extensions for Support of Diffserv-aware MPLS Traffic Engineering,” June 2005.) [RFC4124] that allowed the network to enforce different bandwidth constraints for different classes of traffic.
These developments allow network operators to optimise network resource utilisation. However, there is currently no defined method of applying different optimisation schemes to a network when in normal operation (with all resources available) and the same network when in failure mode (with one or more resources unavailable).

Although DSTE was developed to allow Traffic engineering at the per class level, where a class is generally considered to be a particular diff-serv class of service, it is proposed here to advertise different DSTE Bandwidth Constraints (BC) for use during normal network operation and failure modes. LSPs would then be signalled using one DSTE Class Type (CT) during normal network operation and a different CT during failure modes. This allows individual control over the booking factors for normal network operation and network failure modes.

2. Application scenarios

TOC

2.1. Scenario 1: Fair traffic loss during failure

TOC

An IP/MPLS network may be designed such that not all traffic can be delivered during a network failure event. All traffic on this network is of a single class of service and is treated as equal. Traffic engineering has been implemented in order to best optimise the network such that there is no traffic loss during normal operation. In order to do this, LSPs reserve their actual load and link booking factors are set to 100%.

In the current implementation, during a network failure some LSPs will not be able to signal their required resources and will therefore fail to be placed. There will consequently be an unfair distribution of packet loss with some LSPs having 100% loss and other with 0% loss.

In order to fairly distribute the traffic loss during failure, it is necessary to effectively increase the booking factors during a failure event to a value greater than 100% so that all LSPs can be placed and the excess load is distributed fairly across the network.

2.2. Scenario 2: Managing out-of-contract traffic

TOC

An IP/MPLS network is designed to carry traffic of different drop-precedence. Traffic with a high drop-precedence is considered as out-of-contract and is configured to be dropped first under congestion. Traffic with a low drop-precedence is considered in-contract and should always be delivered.

The network operator wishes to optimise the network such that all traffic can be delivered during normal network operation but only in-contract traffic is guaranteed during failure events. LSPs are configured to signal only the in-contract load and booking factors are

set to such that bandwidth is still available to the additional out-of-contract load during normal operation; this booking factor would normally be based upon the forecasted ratio of in-contract to out-of contract traffic.
During a failure event, there now may not be enough network resources to signal all LSPs. Consequently, the booking factors need to be increased in order to guarantee that all in-contract demand can be served. This new booking factor would normally be based upon the level of in-contract traffic. Network resources are likely to be congested with total demand, but queuing and scheduling mechanisms can be implemented such that only out-of-contract traffic will be discarded.

3. Considerations

TOC

3.1. Detection of failure

TOC

All network elements within the same IGP area will be aware of a link failure due to a link-state change notification in OSPF or ISIS if they are within the same area as the failure occurred. However the network implementation could be multi-area in which case the link-state change will not be propagated to the entire network.
The Head-end element of an LSP is always made aware of relevant network failures as an LSP passing over a remote failed resource will receive a RSVP Path Tear message. Alternatively, if FRR is implemented, this notification will be through a RSVP "Tunnel locally repaired" Path-Error Message or the "Local protection in use" flag within the RRO object of the RESV message. In the case that the failure is local to the HE element, then the failure is detected due to local link or control-plane failure.

3.2. Bandwidth Constraint Model

TOC

The IETF defines different Bandwidth constraint models for use with Diffserv TE. The Russian Dolls Bandwidth Constraint Model (Le Faucheur, F., “Russian Dolls Bandwidth Constraints Model for Diffserv-aware MPLS Traffic Engineering,” June 2005.) [RFC4127] (RDM) and the Maximum Allocation Bandwidth Constraints Model (Le Faucheur, F. and W. Lai, “Maximum Allocation Bandwidth Constraints Model for Diffserv-aware MPLS Traffic Engineering,” June 2005.) [RFC4125] (MAM) are examples of these different models for reserving bandwidth from network resources.

RDM defines a model where a reservation from BC1 also reserves resources from BC0 where a reservation from BC0 only reserves resources from BC0. For the purposes of using different BCs to represent different booking factors, it is clearly important that any reservation from the BC in use during normal network operation is also reserved from the BC used during network failure mode.
For the purposes of this implementation, an implementation of the RDM model is required with a BC[x] being used during normal network operation and BC[less than x] being used during failure mode.

3.3. Preemption

TOC

DSTE allows that different Preemption priorities can be applied to LSPs of different class-types in a flexible manner. The model of operation described here would generally dictate that an LSP would be signalled with the same Preemption priority whether the network is in normal or failure operation, but other implementations could be envisaged.

3.4. Booking Factor usage

TOC

A booking factor defines the ratio between the actual resource size and the resource size as advertised to the network. Additionally, there is no defined mapping between the resources reserved by a particular LSP and the resources actually consumed by the same LSP and different overbooking implementations are possible: The "LSP Size Overbooking" method or the "Link Size Overbooking" method [RFC4124] (Le Faucheur, F., “Protocol Extensions for Support of Diffserv-aware MPLS Traffic Engineering,” June 2005.). Booking factor implementations consequently vary widely from network to network; the model of operation described here effectively applies the "Link Size Overbooking" method, but could be used in conjunction with the "LSP Size Overbooking" method.

3.5. Admission control behaviour with shared resources

TOC

Section 4.6.4 of [RFC3209] (Awduche, D., Berger, L., Gan, D., Li, T., Srinivasan, V., and G. Swallow, “RSVP-TE: Extensions to RSVP for LSP Tunnels,” December 2001.) describes how to setup a tunnel that is capable of maintaining resource reservations (without double counting) while it is being rerouted. For this functionality to be maintained in the operational model described here, implementations must function as

described in this reference even though the LSP is also changing class type during the process.
More specifically, if we consider the RDM Bandwidth Constraint Model, the admission control implementation must allow:
An LSP originally signalled with CT1 must be able to share the original bandwidth component reserved from BC0 as is re-signalled with CT0. The BC1 reservation will be removed as and when the original LSP is torn down.
An LSP originally signalled with CT0 must be able to share the BC0 component of the reservation as it is re-signalled with CT1. There will be no change in reservation as and when the original LSP is torn down.

4. Operation

TOC

4.1. Normal network operation

TOC

4.1.1. Traffic Engineering advertisements

TOC

*We use BC1 to advertise the available bandwidth on all network resources for use in normal network operation.

*We use BC0 to advertise the available bandwidth on all network resources during failure modes.

*We use the RDM bandwidth constraint model.

The overbooking factor during normal operation is therefore equal to: BC1 / link_reservable_bandwidth.
The overbooking factor during failure modes is therefore equal to: BC0 / link_reservable_bandwidth.
Note that any BC could be used so long as the BC number for the failure mode is lower than the BC number for normal operation; we are using BC0 and BC1 as an example.

TOC

4.1.2. Head-end behaviour under normal network operation

The Head-End (HE) of an LSP signalled during normal network operation runs its CSPF using BC1 and signals the LSP using CT1. This allows the network to optimise itself based upon the booking factor for normal network operation.

BC0 is not used other than the fact that any network resources reserved from BC1 is also reserved from BC0 as we are using RDM.

4.2. Network failure mode operation

TOC

4.2.1. Operation at point of failure

TOC

At the point of failure, all LSPs that pass over the failed resource will be torn down - notifying the HE of the failure. If MPLS Fast ReRoute (FRR) is deployed, then the LSPs that pass over the failed resource will be subject to FRR procedures and the HE will consequently be notified through a path-error message.

4.2.2. Head-end Behaviour under Network Failure Mode

TOC

The HE of an LSP affected by the network failure will receive notification of the LSP failure. It runs a CSPF algorithm for the LSP path using BC1 and, if this fails, it considers the LSP in question to be operating under network failure mode. Additionally, if this CSPF finds an available BC1 path through the network, but LSP set-up is rejected by a downstream node, the LSP should also then be considered to be operating under network failure mode. In failure mode, the HE then runs a new CSPF algorithm for the LSP path using BC0, allowing the LSP to use the additional resource made available for failure mode operations. Only the specific LSPs affected by the failure are considered to be operating under network failure mode.

Where FRR has been implemented with Global Reversion, a Make-Before Break operation takes place instead of a complete resignal of the LSP in question. In this case, the bandwidth signalled by the new LSP in CT0 could be shared with the bandwidth signalled by the original LSP in BC1 during the MBB operation as per Section 3.5 (Admission control behaviour with shared resources).

4.2.3. Mid-point behaviour under network failure mode.

TOC

The LSP mid-points have no specific requirements during failure operation, however, they must be capable of allowing bandwidth sharing between class types as per Section 3.5 (Admission control behaviour with shared resources).

4.3. Reversion Behaviour

TOC

Consider that the failed resource has been returned to operation. Local implementation defines when and if the HE element will attempt to optimise an LSP. When this optimisation function is performed, the HE should run any CSPF based upon the BC and CT signalled for normal network operation rather than failure mode (BC1 and CT1 respectively in our example). In other words, the CT for failure mode operation should only be used directly after a failure event and when the LSP is down or undergoing FRR.

In the case that a second failure event occurs on a failure mode LSP before reversion takes place, then the LSP will go through the procedures in Section 4.2.2 (Head-end Behaviour under Network Failure Mode)

In the case that the new CSPF identifies that the optimum path is identical to the existing path, the LSP should still be re-signalled with the CT for normal network operation.

The HE element uses standard LSP signalling behaviour for reversion whilst allowing bandwidth sharing between class types as per Section 3.5 (Admission control behaviour with shared resources).

4.4. New LSPs signalled during failure.

TOC

Any new LSPs signalled during failure should initially be routed and signalled using the CT and BC for normal network operation. In the case that this fails, the LSP setup should fail in the normal way. The LSP should not make use of the BC and CT for failure mode operation.

5. IANA Considerations

TOC

This memo includes no request to IANA.

6. Security Considerations

TOC

None.

7. Normative References

TOC

[RFC2119]	Bradner, S., “Key words for use in RFCs to Indicate Requirement Levels,” BCP 14, RFC 2119, March 1997 (TXT, HTML, XML).
[RFC3209]	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 (TXT).
[RFC4124]	Le Faucheur, F., “Protocol Extensions for Support of Diffserv-aware MPLS Traffic Engineering,” RFC 4124, June 2005 (TXT).
[RFC4125]	Le Faucheur, F. and W. Lai, “Maximum Allocation Bandwidth Constraints Model for Diffserv-aware MPLS Traffic Engineering,” RFC 4125, June 2005 (TXT).
[RFC4127]	Le Faucheur, F., “Russian Dolls Bandwidth Constraints Model for Diffserv-aware MPLS Traffic Engineering,” RFC 4127, June 2005 (TXT).

Authors' Addresses

TOC

	Jonathan Newton
	Cable&Wireless
Email:	jonathan.newton@cw.com
	Mustapha Aissaoui
	Alcatel-Lucent
Email:	mustapha.aissaoui@alcatel-lucent.com
	JP Vasseur
	Cisco Systems, Inc.
Email:	jvasseur@cisco.com

Full Copyright Statement

TOC

This document and the information contained herein are provided on an “AS IS” basis and THE CONTRIBUTOR, THE ORGANIZATION HE/SHE REPRESENTS OR IS SPONSORED BY (IF ANY), THE INTERNET SOCIETY, THE IETF TRUST AND THE INTERNET ENGINEERING TASK FORCE DISCLAIM ALL WARRANTIES, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO ANY WARRANTY THAT THE USE OF THE INFORMATION HEREIN WILL NOT INFRINGE ANY RIGHTS OR ANY IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE.

Intellectual Property

The IETF takes no position regarding the validity or scope of any Intellectual Property Rights or other rights that might be claimed to pertain to the implementation or use of the technology described in this document or the extent to which any license under such rights might or might not be available; nor does it represent that it has made any independent effort to identify any such rights. Information on the procedures with respect to rights in RFC documents can be found in BCP 78 and BCP 79.

Copies of IPR disclosures made to the IETF Secretariat and any assurances of licenses to be made available, or the result of an attempt made to obtain a general license or permission for the use of such proprietary rights by implementers or users of this specification can be obtained from the IETF on-line IPR repository at http://www.ietf.org/ipr.

The IETF invites any interested party to bring to its attention any copyrights, patents or patent applications, or other proprietary rights that may cover technology that may be required to implement this standard. Please address the information to the IETF at ietf-ipr@ietf.org.