RTGWG
Internet-Draft
Intended status: Informational
Expires: August 2, 2012

C. Villamizar, Ed.
OCCNC, LLC
D. McDysan, Ed.
S. Ning
A. Malis
Verizon
L. Yong
Huawei USA

January 30, 2012

Requirements for MPLS Over a Composite Link
draft-ietf-rtgwg-cl-requirement-05

Abstract

There is often a need to provide large aggregates of bandwidth that are best provided using parallel links between routers or MPLS LSR. In core networks there is often no alternative since the aggregate capacities of core networks today far exceed the capacity of a single physical link or single packet processing element.

The presence of parallel links, with each link potentially comprised of multiple layers has resulted in additional requirements. Certain services may benefit from being restricted to a subset of the component links or a specific component link, where component link characteristics, such as latency, differ. Certain services require that an LSP be treated as atomic and avoid reordering. Other services will continue to require only that reordering not occur within a microflow as is current practice.

Current practice related to multipath is described briefly in an appendix.

This Internet-Draft will expire on August 2, 2012.

Table of Contents

1. Introduction

The purpose of this document is to describe why network operators require certain functions in order to solve certain business problems (Section 2). The intent is to first describe why things need to be done in terms of functional requirements that are as independent as possible of protocol specifications (Section 4). For certain functional requirements this document describes a set of derived protocol requirements (Section 5). Three appendices provide supporting details as a summary of existing/prior operator approaches (Appendix A), a summary of implementation techniques and relevant protocol standards (Appendix B), and a summary of G.800 terminology used to define a composite link (Appendix C).

1.1. Requirements Language

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 RFC 2119 [RFC2119].

2. Assumptions

The services supported include L3VPN RFC 4364 [RFC4364], RFC 4797 [RFC4797]L2VPN RFC 4664 [RFC4664] (VPWS, VPLS (RFC 4761 [RFC4761], RFC 4762 [RFC4762]) and VPMS VPMS Framework [I-D.ietf-l2vpn-vpms-frmwk-requirements]), Internet traffic encapsulated by at least one MPLS label, and dynamically signaled MPLS or MPLS-TP LSPs and pseudowires. The MPLS LSPs supporting these services may be pt-pt, pt-mpt, or mpt-mpt.

The locations in a network where these requirements apply are a Label Edge Router (LER) or a Label Switch Router (LSR) as defined in RFC 3031 [RFC3031].

The IP DSCP cannot be used for flow identification since L3VPN requires Diffserv transparency (see RFC 4031 5.5.2 [RFC4031]), and in general network operators do not rely on the DSCP of Internet packets.

3. Definitions

ITU-T G.800 Based Composite and Component Link Definitions:
Section 6.9.2 of ITU-T-G.800 [ITU-T.G.800] defines composite and component links as summarized in Appendix C. The following definitions for composite and component links are derived from and intended to be consistent with the cited ITU-T G.800

terminology.

Composite Link: A composite link is a logical link composed of a set of parallel point-to-point component links, where all links in the set share the same endpoints. A composite link may itself be a component of another composite link, but only a strict hierarchy of links is allowed.

Component Link: A point-to-point physical or logical link that preserves ordering in the steady state. A component link may have transient out of order events, but such events must not exceed the network's specific NPO. Examples of a physical link are: Lambda, Ethernet PHY, and OTN. Examples of a logical link are: MPLS LSP, Ethernet VLAN, and MPLS-TP LSP.

Flow: A sequence of packets that must be transferred in order on one component link.

Flow identification: The label stack and other information that uniquely identifies a flow. Other information in flow identification may include an IP header, PW control word, Ethernet MAC address, etc. Note that an LSP may contain one or more Flows or an LSP may be equivalent to a Flow. Flow identification is used to locally select a component link, or a path through the network toward the destination.

Network Performance Objective (NPO): Numerical values for performance measures, principally availability, latency, and delay variation. See Appendix A for more details.

4. Network Operator Functional Requirements

The Functional Requirements in this section are grouped in subsections starting with the highest priority.

4.1. Availability, Stability and Transient Response

Limiting the period of unavailability in response to failures or transient events is extremely important as well as maintaining stability. The transient period between some service disrupting event and the convergence of the routing and/or signaling protocols MUST occur within a time frame specified by NPO values. Appendix A provides references and a summary of service types requiring a range of restoration times.

FR#1 The solution SHALL provide a means to summarize some routing advertisements regarding the characteristics of a composite link such that the routing protocol converges within the timeframe needed to meet the network performance objective. A composite link CAN be announced in conjunction with detailed parameters about its component links, such as bandwidth and latency. The composite link SHALL behave as a single IGP adjacency.

FR#2 The solution SHALL ensure that all possible restoration operations happen within the timeframe needed to meet the NPO. The solution may need to specify a means for aggregating signaling to meet this requirement.

FR#3 The solution SHALL provide a mechanism to select a path for a flow across a network that contains a number of paths comprised of pairs of nodes connected by composite links in such a way as to automatically distribute the load over the network nodes connected by composite links while meeting all of the other mandatory requirements stated above. The solution SHOULD work in a manner similar to that of current networks without any composite link protocol enhancements when the characteristics of the individual component links are advertised.

FR#4 If extensions to existing protocols are specified and/or new protocols are defined, then the solution SHOULD provide a means for a network operator to migrate an existing deployment in a minimally disruptive manner.

FR#5 Any automatic LSP routing and/or load balancing solutions MUST not oscillate such that performance observed by users changes such that an NPO is violated. Since oscillation may cause reordering, there MUST be means to control the frequency of changing the component link over which a flow is placed.

FR#6 Management and diagnostic protocols MUST be able to operate over composite links.

4.2. Component Links Provided by Lower Layer Networks

Case 3 as defined in [ITU-T.G.800] involves a component link supporting an MPLS layer network over another lower layer network (e.g., circuit switched or another MPLS network (e.g., MPLS-TP)). The lower layer network may change the latency (and/or other performance parameters) seen by the MPLS layer network. Network Operators have NPOs of which some components are based on performance parameters. Currently, there is no protocol for the lower layer network to inform the higher layer network of a change in a

performance parameter. Communication of the latency performance parameter is a very important requirement. Communication of other performance parameters (e.g., delay variation) is desirable.

FR#7 In order to support network NPOs and provide acceptable user experience, the solution SHALL specify a protocol means to allow a lower layer server network to communicate latency to the higher layer client network.

FR#8 The precision of latency reporting SHOULD be at least 10% of the one way latencies for latency of 1 ms or more.

FR#9 The solution SHALL provide a means to limit the latency on a per LSP basis between nodes within a network to meet an NPO target when the path between these nodes contains one or more pairs of nodes connected via a composite link.

The NPOs differ across the services, and some services have different NPOs for different QoS classes, for example, one QoS class may have a much larger latency bound than another. Overload can occur which would violate an NPO parameter (e.g., loss) and some remedy to handle this case for a composite link is required.

FR#10 If the total demand offered by traffic flows exceeds the capacity of the composite link, the solution SHOULD define a means to cause the LSPs for some traffic flows to move to some other point in the network that is not congested. These "preempted LSPs" may not be restored if there is no uncongested path in the network.

4.3. Parallel Component Links with Different Characteristics

Corresponding to Case 1 of [ITU-T.G.800], as one means to provide high availability, network operators deploy a topology in the MPLS network using lower layer networks that have a certain degree of diversity at the lower layer(s). Many techniques have been developed to balance the distribution of flows across component links that connect the same pair of nodes. When the path for a flow can be chosen from a set of candidate nodes connected via composite links, other techniques have been developed.

FR#11 The solution SHALL measure traffic on a labeled traffic flow and dynamically select the component link on which to place this flow in order to balance the load so that no component link in the composite link between a pair of nodes is overloaded.

FR#12 When a traffic flow is moved from one component link to another in the same composite link between a set of nodes (or sites), it MUST be done so in a minimally disruptive manner.

When a flow is moved from a current link to a target link with different latency, reordering can occur if the target link latency is less than that of the current or clumping can occur if target link latency is greater than that of the current. Therefore, some flows (e.g., timing distribution, PW circuit emulation) are quite sensitive to these effects, which may be specified in an NPO or are needed to meet a user experience objective (e.g. jitter buffer under/overrun).

FR#13 The solution SHALL provide a means to identify flows whose rearrangement frequency needs to be bounded by a configured value.

FR#14 The solution SHALL provide a means that communicates whether the flows within an LSP can be split across multiple component links. The solution SHOULD provide a means to indicate the flow identification field(s) which can be used along the flow path which can be used to perform this function.

FR#15 The solution SHALL provide a means to indicate that a traffic flow shall select a component link with the minimum latency value.

FR#16 The solution SHALL provide a means to indicate that a traffic flow shall select a component link with a maximum acceptable latency value as specified by protocol.

FR#17 The solution SHALL provide a means to indicate that a traffic flow shall select a component link with a maximum acceptable delay variation value as specified by protocol.

FR#18 The solution SHALL provide a means local to a node that automatically distributes flows across the component links in the composite link such that NPOs are met.

FR#19 The solution SHALL provide a means to distribute flows from a single LSP across multiple component links to handle at least the case where the traffic carried in an LSP exceeds that of any component link in the composite link. As defined in section 3, a flow is a sequence of packets that must be transferred on one component link.

FR#20 The solution SHOULD support the use case where a composite link itself is a component link for a higher order composite link. For example, a composite link comprised of MPLS-TP bi-directional tunnels viewed as logical links could then be used as a component link in yet another composite link that connects MPLS routers.

FR#21 The solution MUST support an optional means for LSP signaling to bind an LSP to a particular component link within a composite link. If this option is not exercised, then an LSP that is bound to a composite link may be bound to any component link matching all other signaled requirements, and different directions of a bidirectional LSP can be bound to different component links.

FR#22 The solution MUST support a means to indicate that both directions of co-routed bidirectional LSP MUST be bound to the same component link.

5. Derived Requirements

This section takes the next step and derives high-level requirements on protocol specification from the functional requirements.

DR#1 The solution SHOULD attempt to extend existing protocols wherever possible, developing a new protocol only if this adds a significant set of capabilities.

DR#2 A solution SHOULD extend LDP capabilities to meet functional requirements (without using TE methods as decided in [RFC3468]).

DR#3 Coexistence of LDP and RSVP-TE signaled LSPs MUST be supported on a composite link. Other functional requirements should be supported as independently of signaling protocol as possible.

DR#4 When the nodes connected via a composite link are in the same MPLS network topology, the solution MAY define extensions to the IGP.

DR#5 When the nodes are connected via a composite link are in different MPLS network topologies, the solution SHALL NOT rely on extensions to the IGP.

DR#6 The Solution SHOULD support composite link IGP advertisement that results in convergence time better than that of advertising the individual component links. The solution SHALL be designed so that it represents the range of capabilities of the individual component links such that functional requirements are met, and also minimizes the frequency of advertisement updates which may cause IGP convergence to occur.

Examples of advertisement update triggering events to be considered include: LSP establishment/release, changes in component link characteristics (e.g., latency, up/down state), and/or bandwidth utilization.

DR#7 When a worst case failure scenario occurs, the number of RSVP-TE LSPs to be resignaled will cause a period of unavailability as perceived by users. The resignaling time of the solution MUST meet the NPO objective for the duration of unavailability. The resignaling time of the solution MUST not increase significantly as compared with current methods.

6. Management Requirements

MR#1 Management Plane MUST support polling of the status and configuration of a composite link and its individual composite link and support notification of status change.

MR#2 Management Plane MUST be able to activate or de-activate any component link in a composite link in order to facilitate operation maintenance tasks. The routers at each end of a composite link MUST redistribute traffic to move traffic from a de-activated link to other component links based on the traffic flow TE criteria.

MR#3 Management Plane MUST be able to configure a LSP over a composite link and be able to select a component link for the LSP.

MR#4 Management Plane MUST be able to trace which component link a LSP is assigned to and monitor individual component link and composite link performance.

MR#5 Management Plane MUST be able to verify connectivity over each individual component link within a composite link.

MR#6 Management Plane SHOULD provide the means for an operator to initiate an optimization process.

7. Acknowledgements

Frederic Jounay of France Telecom and Yuji Kamite of NTT Communications Corporation co-authored a version of this document.

A rewrite of this document occurred after the IETF77 meeting. Dimitri Papadimitriou, Lou Berger, Tony Li, the WG chairs John Scuder and Alex Zinin, and others provided valuable guidance prior to and at the IETF77 RTGWG meeting.

Tony Li and John Drake have made numerous valuable comments on the RTGWG mailing list that are reflected in versions following the IETF77 meeting.

8. IANA Considerations

This memo includes no request to IANA.

9. Security Considerations

This document specifies a set of requirements. The requirements themselves do not pose a security threat. If these requirements are met using MPLS signaling as commonly practiced today with authenticated but unencrypted OSPF-TE, ISIS-TE, and RSVP-TE or LDP, then the requirement to provide additional information in this communication presents additional information that could conceivably be gathered in a man-in-the-middle confidentiality breach. Such an attack would require a capability to monitor this signaling either through a provider breach or access to provider physical transmission infrastructure. A provider breach already poses a threat of numerous tpes of attacks which are of far more serious consequence. Encrption of the signaling can prevent or render more difficult any confidentiality breach that otherwise might occur by means of access to provider physical transmission infrastructure.

10. References

10.1. Normative References

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

10.2. Informative References

[I-D.ietf-l2vpn-vpms-frmwk-requirements]
Kamite, Y., JOUNAY, F., Niven-Jenkins, B., Brungard, D., and L. Jin, "Framework and Requirements for Virtual Private Multicast Service (VPMS)", draft-ietf-l2vpn-vpms-frmwk-requirements-03 (work in progress), July 2010.

[ITU-T.G.800]
ITU-T, "Unified functional architecture of transport networks", 2007, <http://www.itu.int/rec/T-REC-G/recommendation.asp?parent=T-REC-G.800>.

[RFC2702] Awduche, D., Malcolm, J., Agogbua, J., O'Dell, M., and J. McManus, "Requirements for Traffic Engineering Over MPLS", RFC 2702, September 1999.

[RFC3031] Rosen, E., Viswanathan, A., and R. Callon, "Multiprotocol Label Switching Architecture", RFC 3031, January 2001.

[RFC3468] Andersson, L. and G. Swallow, "The Multiprotocol Label Switching (MPLS) Working Group decision on MPLS signaling protocols", RFC 3468, February 2003.

[RFC3809] Nagarajan, A., "Generic Requirements for Provider Provisioned Virtual Private Networks (PPVPN)", RFC 3809, June 2004.

[RFC4031] Carugi, M. and D. McDysan, "Service Requirements for Layer 3 Provider Provisioned Virtual Private Networks (PPVPNs)", RFC 4031, April 2005.

[RFC4364] Rosen, E. and Y. Rekhter, "BGP/MPLS IP Virtual Private Networks (VPNs)", RFC 4364, February 2006.

[RFC4664] Andersson, L. and E. Rosen, "Framework for Layer 2 Virtual Private Networks (L2VPNs)", RFC 4664, September 2006.

[RFC4665] Augustyn, W. and Y. Serbest, "Service Requirements for Layer 2 Provider-Provisioned Virtual Private Networks", RFC 4665, September 2006.

[RFC4761] Kompella, K. and Y. Rekhter, "Virtual Private LAN Service (VPLS) Using BGP for Auto-Discovery and Signaling", RFC 4761, January 2007.

[RFC4762] Lasserre, M. and V. Kompella, "Virtual Private LAN Service

(VPLS) Using Label Distribution Protocol (LDP) Signaling", RFC 4762, January 2007.

[RFC4797] Rekhter, Y., Bonica, R., and E. Rosen, "Use of Provider Edge to Provider Edge (PE-PE) Generic Routing Encapsulation (GRE) or IP in BGP/MPLS IP Virtual Private Networks", RFC 4797, January 2007.

[RFC5254] Bitar, N., Bocci, M., and L. Martini, "Requirements for Multi-Segment Pseudowire Emulation Edge-to-Edge (PWE3)", RFC 5254, October 2008.

10.3. Appendix References

[I-D.ietf-pwe3-fat-pw] Bryant, S., Filsfils, C., Drafz, U., Kompella, V., Regan, J., and S. Amante, "Flow Aware Transport of Pseudowires over an MPLS PSN", draft-ietf-pwe3-fat-pw-03 (work in progress), January 2010.

[RFC1717] Sklower, K., Lloyd, B., McGregor, G., and D. Carr, "The PPP Multilink Protocol (MP)", RFC 1717, November 1994.

[RFC2475] Blake, S., Black, D., Carlson, M., Davies, E., Wang, Z., and W. Weiss, "An Architecture for Differentiated Services", RFC 2475, December 1998.

[RFC2615] Malis, A. and W. Simpson, "PPP over SONET/SDH", RFC 2615, June 1999.

[RFC2991] Thaler, D. and C. Hopps, "Multipath Issues in Unicast and Multicast Next-Hop Selection", RFC 2991, November 2000.

[RFC2992] Hopps, C., "Analysis of an Equal-Cost Multi-Path Algorithm", RFC 2992, November 2000.

[RFC3260] Grossman, D., "New Terminology and Clarifications for Diffserv", RFC 3260, April 2002.

[RFC4201] Kompella, K., Rekhter, Y., and L. Berger, "Link Bundling in MPLS Traffic Engineering (TE)", RFC 4201, October 2005.

[RFC4301] Kent, S. and K. Seo, "Security Architecture for the Internet Protocol", RFC 4301, December 2005.

[RFC4385] Bryant, S., Swallow, G., Martini, L., and D. McPherson, "Pseudowire Emulation Edge-to-Edge (PWE3) Control Word for Use over an MPLS PSN", RFC 4385, February 2006.

[RFC4928] Swallow, G., Bryant, S., and L. Andersson, "Avoiding Equal Cost Multipath Treatment in MPLS Networks", BCP 128, RFC 4928, June 2007.

Appendix A. Existing Network Operator Practices and Protocol Usage

The network operator practices appendix has been moved to a separate document. When that document has an XML I-D tag the references to this appendix will be changed to that document and this appendix will be deleted.

Appendix B. Existing Multipath Standards and Techniques

The multipath standards and techniques appendix has been moved to a separate document. When that document has an XML I-D tag the references to this appendix will be changed to that document and this appendix will be deleted.

Appendix C. ITU-T G.800 Composite Link Definitions and Terminology

Composite Link:

Section 6.9.2 of ITU-T-G.800 [ITU-T.G.800] defines composite link in terms of three cases, of which the following two are relevant (the one describing inverse (TDM) multiplexing does not apply). Note that these case definitions are taken verbatim from section 6.9, "Layer Relationships".

Case 1: "Multiple parallel links between the same subnetworks can be bundled together into a single composite link. Each component of the composite link is independent in the sense that each component link is supported by a separate server layer trail. The composite link conveys communication information using different server layer trails thus the sequence of symbols crossing this link may not be preserved. This is illustrated in Figure 14."

Case 3: "A link can also be constructed by a concatenation of component links and configured channel forwarding relationships. The forwarding relationships must have a 1:1 correspondence to the link connections that will be provided by the client link. In this case, it is not possible to fully infer the status of the link by observing the server layer trails visible at the ends of the link. This is illustrated in Figure 16."

Subnetwork: A set of one or more nodes (i.e., LER or LSR) and links. As a special case it can represent a site comprised of multiple nodes.

Forwarding Relationship: Configured forwarding between ports on a subnetwork. It may be connectionless (e.g., IP, not considered in this draft), or connection oriented (e.g., MPLS signaled or configured).

Component Link: A topolological relationship between subnetworks (i.e., a connection between nodes), which may be a wavelength, circuit, virtual circuit or an MPLS LSP.

Authors' Addresses

Curtis Villamizar (editor)
OCCNC, LLC

Email: curtis@occnc.com

Dave McDysan (editor)
Verizon
22001 Loudoun County PKWY
Ashburn, VA 20147

Email: dave.mcdysan@verizon.com

So Ning
Verizon
2400 N. Glenville Ave.
Richardson, TX 75082

Phone: +1 972-729-7905
Email: ning.so@verizonbusiness.com

Andrew Malis
Verizon
117 West St.
Waltham, MA 02451

Phone: +1 781-466-2362
Email: andrew.g.malis@verizon.com

Lucy Yong
Huawei USA
1700 Alma Dr. Suite 500
Plano, TX 75075

Phone: +1 469-229-5387
Email: lucyyong@huawei.com