

Network Working Group
Internet-Draft
Intended status: Standards Track
Expires: April 25, 2013

X. Fu
M. Betts
Q. Wang
ZTE
D. McDysan
A. Malis
Verizon
V. Manral
Hewlett-Packard Corp.
October 22, 2012

RSVP-TE extensions for Loss and Delay Traffic Engineering
draft-fuxh-mpls-delay-loss-rsvp-te-ext-02

Abstract

With more and more enterprises using cloud based services, the distances between the user and the applications are growing. For multiple applications such as High Performance Computing and Electronic Financial markets, the response times are critical as is packet loss, while other applications require more throughput. For example, financial or trading companies are very focused on end-to-end private pipe line delay optimizations that improve things 2-3 ms. Delay, jitter, loss and SLA (Service Level Agreement) are key parameters that these "high value" customers use to select a private pipe line provider. This document extends RSVP-TE protocol to promote SLA experience of delay and packet loss application.

Status of this Memo

This Internet-Draft is submitted in full conformance with the provisions of [BCP 78](#) and [BCP 79](#).

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at <http://datatracker.ietf.org/drafts/current/>.

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."

This Internet-Draft will expire on April 25, 2013.

Copyright Notice

Internet-Draft

RSVP-TE for services aware MPLS

October 2012

Copyright (c) 2012 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to [BCP 78](http://trustee.ietf.org/license-info) and the IETF Trust's Legal Provisions Relating to IETF Documents (<http://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

Table of Contents

1.	Introduction	3
1.1.	Conventions Used in This Document	4
2.	Performance Accumulation and Verification	4
2.1.	New RSVP ADSPEC sub-object	4
2.2.	New RSVP SENDER_TSPEC sub-object	6
2.3.	New RSVP FLOWSPEC sub-object	7
2.4.	Signaling Procedures	8
3.	Performance SLA Parameters Conveying	9
3.1.	Performance SLA Parameters ERO sub-object	10
3.2.	Signaling Procedure	14
4.	Security Considerations	14
5.	IANA Considerations	14
6.	References	15
6.1.	Normative References	15
6.2.	Informative References	15
	Authors' Addresses	16

1. Introduction

End-to-end service optimization based on delay, jitter and loss is a key requirement for service provider. So communicating delay, jitter and packet loss as traffic engineering performance metrics is a very important requirement. [DELAY-LOSS-PS] describes the requirement of delay and loss traffic engineering application. [DELAY-LOSS-FRAMEWORK] describes the framework and architecture to meet requirements. Delay, jitter and loss metrics are sent in IGP protocol as defined in [[OSPF-TE-EXPRESS-PATH](#)] and [ISIS-TE-EXPRESS-PATH]. [[EXPRESS-PATH](#)] describes how to use these traffic engineering metrics to compute explicit paths at path computation entity. So source node could predict end-to-end delay or loss performance before an end-to-end LSP is established.

In the case of multi-domain (e.g., Inter-AS, Inter-Area or Multi-Layer), a full path of TE LSP can't be or is not determined at the ingress node of TE LSP. This is most likely to arise owing to TE visibility limitations. If not all domains support to communicate delay, jitter and loss as traffic engineering metric parameters, one end-to-end optimized path with performance constraint (e.g., less than 10 ms) may not be computed by BRPC [[RFC5441](#)] in PCE architecture.

This document extend RSVP-TE to accumulate (e.g., sum) delay and loss information of links and nodes along one LSP across multi-domain (e.g., Inter-AS, Inter-Area or Multi-Layer) so that end points can verify whether total amount of delay or loss could meet the agreement between operator and his user. When RSVP-TE signaling is used, source node can determine if delay and loss requirement is met much more rapidly than performing actual end-to-end performance measurement. delay, jitter and loss are part of service/QoS description/characterization and that as such belongs in a flowspec/tspec/adspec. This document modify IntServ (as represented by [RFC210](#)) to provide new parameters.

One end-to-end LSP may go across some Composite Links [[CL-REQ](#)]. RSVP-TE message needs to carry a indication for selection of component links based on delay, jitter or loss constraint. When one end-to-end LSP traverse a server layer, there will be some delay, jitter or loss constraint requirements for server layer. So RSVP-TE message needs to carry a indication for FA selection or FA-LSP creation. This document defines a new ERO sub-object to indicate that a component links, FA or FA-LSP should meet maximum acceptable delay, jitter or loss value.

[1.1.](#) Conventions Used in This Document

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.](#) Performance Accumulation and Verification

Delay, jitter and loss accumulation and verification applies where a full path of multi-domain (e.g., Inter-AS, Inter-Area or Multi-Layer) TE LSP can't be or is not determined at ingress node of the TE LSP. This is most likely to arise owing to TE visibility limitations. If all domains support to communicate delay, jitter and loss as traffic engineering metric parameters, one end-to-end optimized path with performance constraint (e.g., less than 10 ms) could be computed by BRPC [[RFC5441](#)] in PCE. Otherwise, it could use the mechanism defined in this section to accumulat delay, jitter and loss along a path which goes across multi-domain.

E2E performance requirement (e.g., delay isn't larger than 10ms) could be signaled by RSVP-TE (e.g., Path and Resv message). Intermediate nodes could reject the request (Path or Resv message) if the accumulated delay, jitter or loss is not achievable. This is essential in multiple AS use cases, but may not be needed in a single IGP level/area if the IGP is extended to convey delay, jitter and loss information.

Node delay for a WAN could be ignored or even an average, however

that was not true for the LAN cases. Whether the node delay should be accumulated or not depends on the implementation.

One domain may need to know that other domains support performance accumulation. It could be discovered in some automatic way. PCEs in different domains may play a role here. It is for further study.

2.1. New RSVP ADSPEC sub-object

This document defines a new RSVP ADSPEC [\[RFC2210\]](#) sub-object to support end-to-end accumulation of delay, jitter or loss. The new RSVP ADSPEC sub-object has the following format.

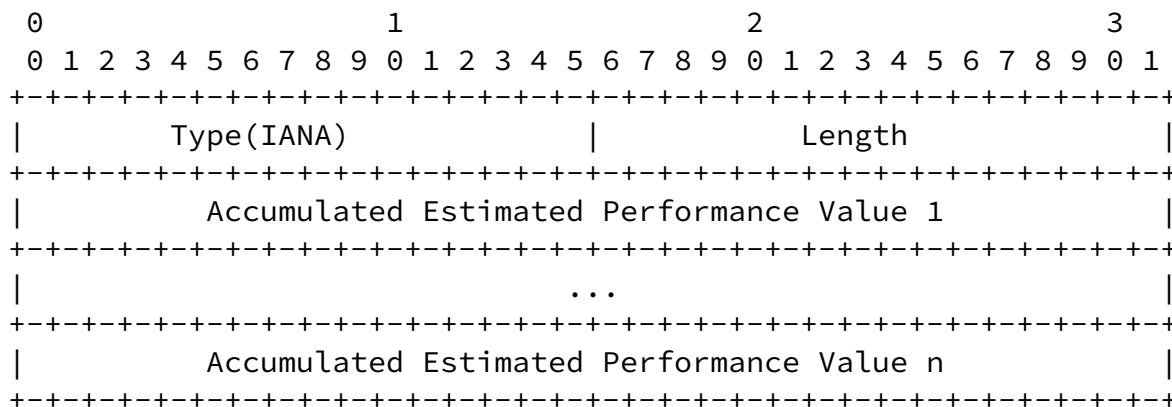


Figure 1: New RSVP ADSPEC sub-Object

- o Type(TBD): It indicates performance accumulation from source to sink.
- o Length: length of performance accumulation value.

Accumulated Estimated Performance Value format is defined in the next picture.

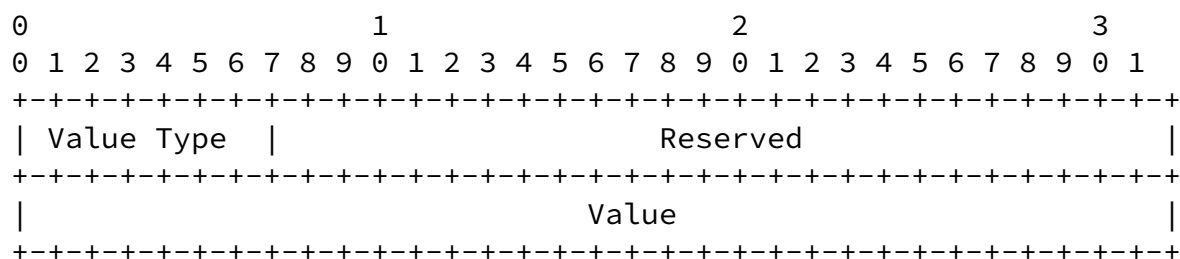


Figure 2: Format of Accumulated Estimated Performance Value

o Value Type:

- * 0: It indicates Performance Accumulation Value is for end-to-end delay accumulation along the LSP.
- * 1: It indicates Performance Accumulation Value is for end-to-end jitter accumulation along the LSP.
- * 2: It indicates Performance Accumulation Value is for end-to-end loss accumulation along the LSP.

o Value:

- * If it is accumulated estimated delay value, it MUST be quantified in units of micro-seconds and encoded as an float point value.

- * If it is accumulated estimated delay variation value, it MUST be quantified in units of micro-seconds and encoded as an float point value. Since latency variation is accumulated non-linearly, delay variation accumulation should be in a lower priority.
- * If it is accumulated estimated loss value, it MUST be quantified in units of the number of packets per million packets. For link loss, the path loss is not the sum of the used links' losses. Instead, the path loss percentage is $(100 - \text{loss}_{L1}) * (100 - \text{loss}_{L2}) * \dots * (100 - \text{loss}_{Ln})$, where the links along the path are L1 to Ln. For example, assume packet loss is 10% for two hops of a link. The measurements will come to 19% total packet loss. Because of 10% loss on the first link only 90% packet reach the second link where another 10% of 90%

are lost, which is 9% of total packets.

2.2. New RSVP SENDER_TSPEC sub-object

This document defines a new RSVP SENDER_TSPEC [RFC2210] sub-object which indicates end-to-end latency, jitter or loss performance requirement. Intermediate nodes could reject the request (Path or Resv message) if the accumulated delay, jitter or loss value exceeds required performance value in the SENDER_TSPEC sub-object.

If the accumulated delay, jitter or loss is not achievable, there is no necessary to accumulate delay, jitter or loss for remaining domain or nodes.

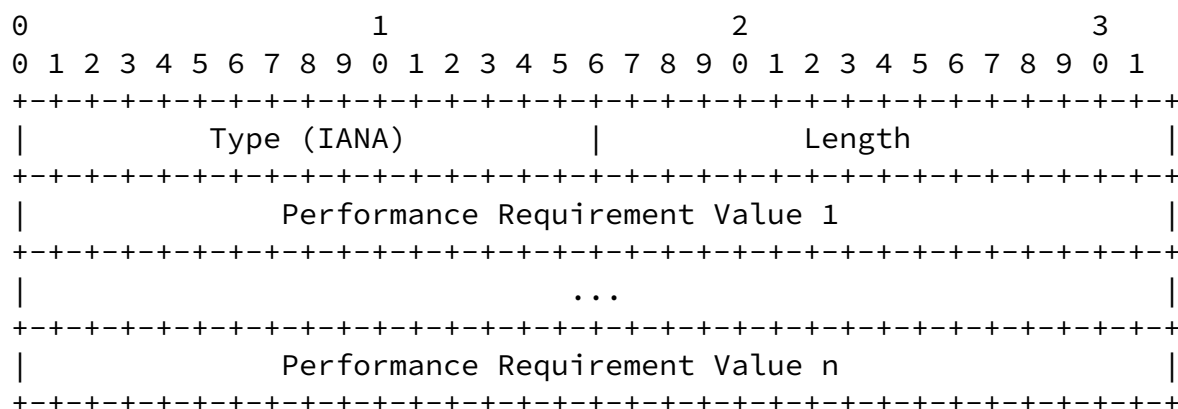


Figure 3: New RSVP SENDER_TSPEC sub-object

The Performance Requirement Value format is defined in the next picture.

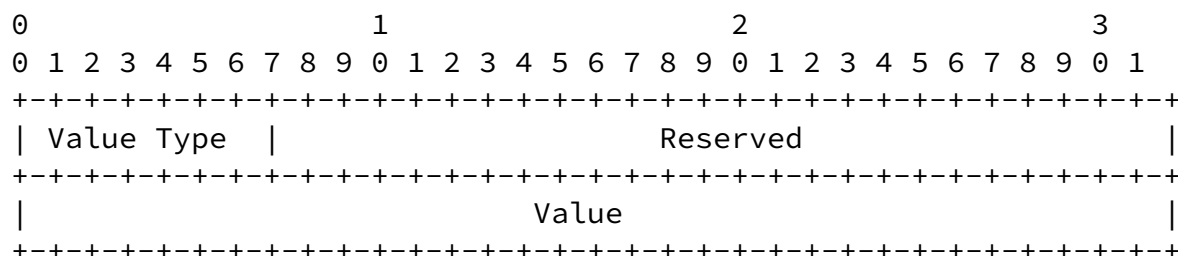


Figure 4: Format of Performance Requirement Value

- o Value Type:
 - * 0: It indicates Performance Value is end-to-end delay requirement. The accumulated estimated delay value should not exceed this value. It MUST be quantified in units of micro-seconds and encoded as an float point value.
 - * 1: It indicates Performance Value is end-to-end jitter requirement. The accumulated estimated delay variation value should not exceed this value. It MUST be quantified in units of micro-seconds and encoded as an float point value.
 - * 2: It indicates Performance Value is end-to-end loss requirement. The accumulated estimated loss value should not exceed this value. It MUST be quantified in units of the number of packets per million packets.
- o Value:
 - * If it is end-to-end delay requirement, accumulated estimated delay value should not exceed this value. It MUST be quantified in units of micro-seconds and encoded as an float point value.
 - * If it is end-to-end jitter requirement, accumulated estimated delay variation value should not exceed this value. It MUST be quantified in units of micro-seconds and encoded as an float point value.
 - * If it is end-to-end loss requirement, the accumulated estimated loss value should not exceed this value. It MUST be quantified in units of the number of packets per million packets.

[2.3.](#) New RSVP FLOWSPEC sub-object

In order to make source node get performance accumulation result from source to sink in multi-domain (e.g., Inter-AS, Inter-Area or Multi-Layer) application, this document defines a new RSVP FLOWSPEC

[RFC2210] sub-object. It has the same format and TLV Type as defined

in section "2.1 New RSVP ADSPEC sub-object". When sink node receive Path message, it must copy the accumulation result from ADSPEC to FLOWSPEC.

When LSP goes across a Composite Links [[CL-REQ](#)], traffic from LSR A to B and B to A may go across different Component Links so that performance from sink to source may not be identical to the one from source to sink. This document defines another new RSVP FLOWSPEC [[RFC2210](#)] sub-object. It has the same format as defined in section "2.1 New RSVP ADSPEC sub-object" except a different TLV Type which indicates performance accumulation from sink to source. So source node can get performance accumulated value from sink to source.

This document also defined a new FLOWSPEC sub-object which has the same format, TLV Type and value as defined in section "2.2 New RSVP SENDER_TSPEC sub-object". It indicates end-to-end delay, jitter or loss performance requirement.

[2.4.](#) Signaling Procedures

When source node desires to accumulate delay, jitter or loss of one end-to-end LSP, "Delay Accumulating desired", "Jitter Accumulation desired" or "Loss Accumulation desired" flag (value TBD) should be set in LSP_ATTRIBUTES object of Path/Resv message [[RFC5420](#)]. If source node makes intermediate node have the capability to verify accumulated performance, "Delay Verifying desired", "Jitter Verifying desired" or "Loss Verifying desired" flag (value TBD) should be also set in LSP_ATTRIBUTES object of Path/Resv message.

A source node initiates performance accumulation for a given LSP by adding a new ADSPEC sub-object as defined in [section 2.1](#) to Path message. If performance verifying is desired, source node also adds a new SENDER_TSPEC sub-object as defined in [section 2.2](#) to Path message.

When downstream node receives Path message and if performance (e.g., delay, jitter or loss) accumulating desired is set in the LSP_ATTRIBUTES, it accumulates delay, jitter or loss and updates ADSPEC sub-object before it sends Path message to downstream.

If performance verifying desired is set in LSP_ATTRIBUTES, downstream node will check whether accumulated estimated performance value exceeds the value carried in SENDER_TSPEC sub-object. If the accumulated performance is not achievable, there is no necessary to accumulate performance for remaining domain or nodes. It MUST generate a error message with a indication which means that accumulated performance (i.e., delay, jitter or loss) couldn't meet

end-to-end performance requirement (TBD by IANA).

If intermediate node (e.g., entry node of one domain) couldn't support performance accumulation function, it MUST generate a error message with a indication which means that performance accumulation is unsupported (TBD by IANA).

When sink node of LSP receives Path message and performance accumulating desired is set in LSP_ATTRIBUTES, it copy accumulated estimated performance value from ADSPEC to FLOWSPEC before it send Resv message to upstream. Then source node can get performance accumulated value from source to sink for unidirectional and bidirectional LSP.

If LSP is a bidirectional one and performance accumulating desired is set in LSP_ATTRIBUTES, it adds a FLOWSPEC sub-object as defined in [section 2.3](#) to accumulate end-to-end performance value from sink to source.

If LSP is a bidirectional one and performance verifying desired is set in LSP_ATTRIBUTES, it copy end-to-end performance requirement value from SENDER_TSPEC sub-object to FLOWSPEC sub-object.

When upstream node receives Resv message and if performance accumulating desired is set in LSP_ATTRIBUTES, it accumulates performance of each hops and updates FLOWSPEC sub-object (i.e., from sink to source) before it sends Resv message to upstream.

If performance verifying desired is set in LSP_ATTRIBUTES, it will check whether accumulated estimated performance value from sink to source exceeds end-to-end performance requirement value. If accumulated performance is not achievable, there is no necessary to accumulate performance for remaining domain or nodes. It MUST generate a error message with a indication which means that accumulated performance couldn't meet end-to-end performance requirement (TBD by IANA).

After source node receive Resv message, it will get accumulated performance value, it can confirm whether performance value meet SLA or not.

[3.](#) Performance SLA Parameters Conveying

[CL-REQ] introduces Composite Link into MPLS network. In order to assign LSP to one of component links with different performance

characteristics (e.g., delay, jitter and loss), RSVP-TE message MUST convey performance SLA parameter to end points of Composite Links.

So it can select one of component links or trigger the creation of lower layer connection based on performance SLA parameter.

One end-to-end LSP (e.g., in IP/MPLS or MPLS-TP network) may traverse a FA-LSP of server layer (e.g., OTN rings). There will be some performance constraint requirement for server layer. RSVP-TE message also needs to carry a indication for FA selection or FA-LSP creation.

So this document extend RSVP-TE to carry a indication of request maximum acceptable delay, jitter or loss value. The end point will take these parameters into account for selection or creation of component link, FA selection or FA-LSP.

This document defines extensions to and describes the use of RSVP-TE [[RFC3209](#)], [[RFC3471](#)], [[RFC3473](#)] to explicitly convey the performance SLA parameter for selection or creation of component link or FA/FA-LSP. Specifically, in this document, Performance SLA Parameters TLV are defined and added into ERO as a sub-object.

3.1. Performance SLA Parameters ERO sub-object

A new OPTIONAL sub-object of EXPLICIT_ROUTE Object (ERO) is used to specify performance SLA parameters including a indication of request maximum acceptable delay, jitter or loss value. It can be used for following scenarios.

- o One end-to-end LSP may traverse a server layer FA-LSP. This sub-object of ERO can indicate that FA selection or FA-LSP creation shall be based on delay, jitter or loss constraint. The boundary nodes of multi-layer will take these parameters into account for FA selection or FA-LSP creation.
- o One end-to-end LSP may be across some Composite Links [[CL-REQ](#)]. This subobject of ERO can indicate that a traffic flow shall select a component link with some delay, jitter or loss constraint values as specified in this subobject.

Performance SLA Parameters ERO subobject has the following format. It follows a subobject containing the IP address, or the link

identifier [[RFC3477](#)], associated with the TE link on which it is to be used.

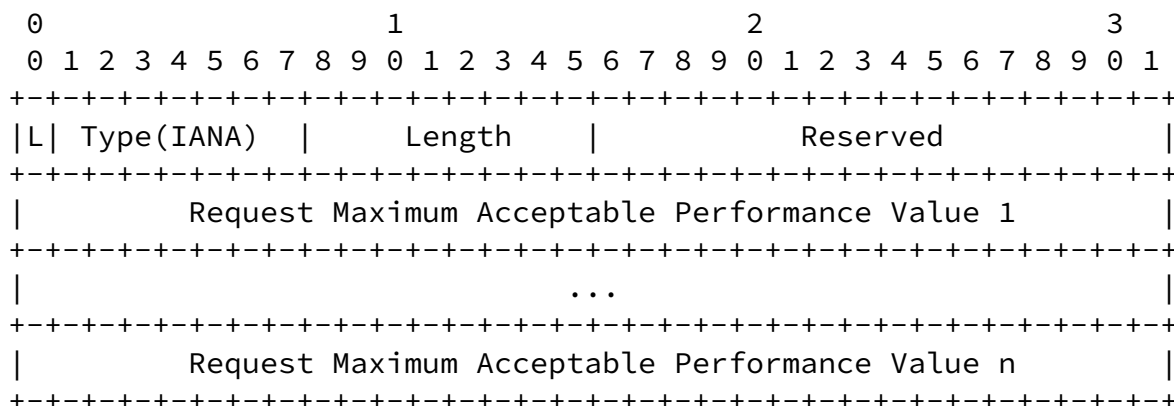


Figure 5: Format of Performance SLA Parameters TLV

Request Maximum Acceptable Performance Value format is defined in the next picture.

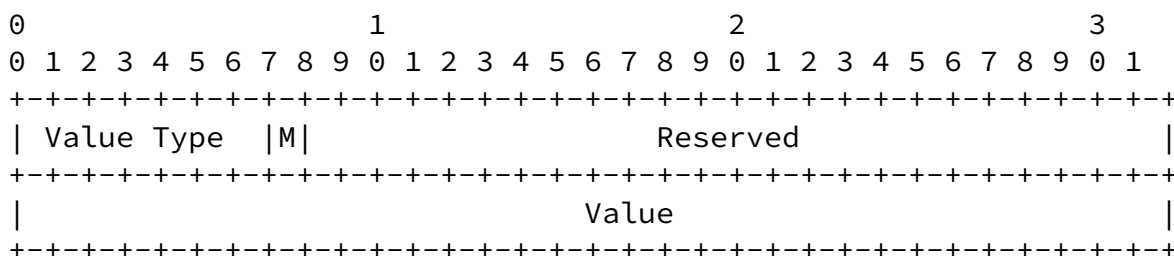


Figure 6: Format of Request Maximum Acceptable Performance Value

o Value Type:

- * 0: It is maximum acceptable delay value.
- * 1: It is maximum acceptable delay variation value.

- * 2: It is maximum acceptable loss value.
- o M bit: a one bit field indicates whether a traffic flow shall select a component link with minimum performance (i.e., delay, jitter or loss) value or not. It can also indicate whether one end-to-end LSP shall select a FA with minimum performance value or not when it traverse a server layer.
- o Value:
 - * If it is maximum acceptable delay or jitter value, it MUST be quantified in units of micro-seconds and encoded as an float point value.

- * If it is maximum acceptable loss value, it MUST be quantified in units of the number of packets per million packets.

Following is an example about how to use these parameters. Assume there are following component links within one composite link.

- o Component link1: delay=50 ms, delay variation=15 ns, loss=8%
- o Component link2: delay=100 ms, delay variation=6 ns, loss=9%
- o Component link3: delay=200 ms, delay variation=3 ns, loss=8%
- o Component link4: delay=300 ms, delay variation=1 ns, loss=7%

Assume there are following request information.

- o Request minimum delay=FALSE
- o Request minimum delay variation=FALSE
- o Request minimum loss=FALSE
- o Maximum Acceptable delay Value=150 ms

- o Maximum Acceptable delay Variation Value=10 ns
- o Maximum Acceptable Loss Value=10%

Only Component link2 could be qualified.

- o Request minimum delay=FALSE
- o Request minimum delay variation=FALSE
- o Request minimum loss=FALSE
- o Maximum Acceptable Delay Value=350 ms
- o Maximum Acceptable Delay Variation Value=10 ns
- o Maximum Acceptable Loss Value=8%

Component link 3 and 4 could be qualified. Which component link is selected depends on local policy.

- o Request minimum delay=FALSE
- o Request minimum delay variation=TRUE
- o Request minimum loss=FALSE
- o Maximum Acceptable Delay Value=350 ms
- o Maximum Acceptable Delay Variation Value=10 ns
- o Maximum Acceptable Loss Value=10%

Only Component link4 could be qualified.

- o Request minimum delay=TRUE
- o Request minimum delay variation=FALSE

- o Request minimum loss=FALSE
- o Maximum Acceptable Delay Value=350 ms
- o Maximum Acceptable Delay Variation Value=10 ns
- o Maximum Acceptable Loss Value=10%

Only Component link2 could be qualified.

- o Request minimum delay=FALSE
- o Request minimum delay variation=FALSE
- o Request minimum loss=TRUE
- o Maximum Acceptable Delay Value= 350 ms
- o Maximum Acceptable Delay Variation Value=10 ns
- o Maximum Acceptable Loss Value=10%

Only Component link4 could be qualified.

Request minimum delay=TRUE

Request minimum delay variation=TRUE

Request minimum loss=TRUE

Maximum Acceptable delay Value=350 ms

Maximum Acceptable delay Variation Value=10 ns

Maximum Acceptable Loss Value=10%

In this case, there is no any qualified component links. But priority may be used for delay and variation, so one of component links could be still selected.

[3.2.](#) Signaling Procedure

When an intermediate node receives a PATH message containing ERO and finds that there is a Performance SLA Parameters ERO sub-object immediately behind the IP address or link address sub-object related to itself, if the node determines that it's a region edge node of FA-LSP or an end point of a Composite Link [[CL-REQ](#)], then this node extracts Performance SLA parameters (i.e., request maximum acceptable delay, jitter and loss value) from Performance SLA Parameters ERO sub-object. This node used these performance parameters for FA selection, FA-LSP creation or component link selection.

If intermediate node couldn't support performance SLA, it MUST generate a PathErr message with a "Performance SLA unsupported" indication (TBD by IANA). If intermediate node couldn't select a FA or component link, or create a FA-LSP which meet performance constraint defined in Performance SLA Parameters ERO sub-object, it must generate a PathErr message with a "Performance SLA parameters couldn't be met" indication (TBD by IANA).

[4.](#) Security Considerations

This document raises no new security issues.

[5.](#) IANA Considerations

TBD

[6.](#) References

[6.1.](#) Normative References

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

- [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.
- [RFC3473] Berger, L., "Generalized Multi-Protocol Label Switching (GMPLS) Signaling Resource ReserVation Protocol-Traffic Engineering (RSVP-TE) Extensions", [RFC 3473](#), January 2003.
- [RFC3477] Kompella, K. and Y. Rekhter, "Signalling Unnumbered Links in Resource ReSerVation Protocol - Traffic Engineering (RSVP-TE)", [RFC 3477](#), January 2003.
- [RFC3630] Katz, D., Kompella, K., and D. Yeung, "Traffic Engineering (TE) Extensions to OSPF Version 2", [RFC 3630](#), September 2003.
- [RFC4203] Kompella, K. and Y. Rekhter, "OSPF Extensions in Support of Generalized Multi-Protocol Label Switching (GMPLS)", [RFC 4203](#), October 2005.

[6.2](#). Informative References

- [CL-REQ] C. Villamizar, "Requirements for MPLS Over a Composite Link", [draft-ietf-rtgwg-cl-requirement-07](#) .
- [DELAY-LOSS-FRAMEWORK]
X.Fu, D. McDysan et al., "Loss and Delay Traffic Engineering Framework for MPLS",
[draft-fuxh-mpls-delay-loss-te-framework-05](#) .
- [EXPRESS-PATH]
A. Atlas, "Performance-based Path Selection for Explicitly Routed LSPs", [draft-atlas-mpls-te-express-path-01](#) .
- [ISIS-TE-EXPRESS-PATH]
S. Previdi, "IS-IS Traffic Engineering (TE) Metric Extensions", [draft-previdi-isis-te-metric-extensions-01](#) .
- [LOSS-DELAY-PS]
X.Fu, D. McDysan et al., "Delay and Loss Traffic Engineering Problem Statement for MPLS",
[draft-fuxh-mpls-delay-loss-te-problem-statement-00](#) .

[OSPF-TE-EXPRESS-PATH]

S. Giacalone, "OSPF Traffic Engineering (TE) Metric Extensions", [draft-ietf-ospf-te-metric-extensions-01](#) .

[ietf-mpls-loss-delay]

D. Frost, "Packet Loss and Delay Measurement for MPLS Networks", [RFC6374](#) .

Authors' Addresses

Xihua Fu
ZTE

Email: fu.xihua@zte.com.cn

Malcolm Betts
ZTE

Email: malcolm.betts@zte.com.cn

Qilei Wang
ZTE

Email: wang.qilei@zte.com.cn

Dave McDysan
Verizon

Email: dave.mcdysan@verizon.com

Andrew Malis
Verizon

Email: andrew.g.malis@verizon.com

Vishwas Manral
Hewlett-Packard Corp.

Email: vishwas.manral@hp.com

