Network Working Group
Internet-Draft
Intended status: Informational
Expires: June 17, 2010

W. Sun, Ed.
SJTU
G. Zhang, Ed.
CATR

December 14, 2009

Label Switched Path (LSP) Data Path Delay Metric in Generalized MPLS/ MPLS-TE Networks

draft-sun-ccamp-dpm-01.txt

Abstract

When setting up a label switched path (LSP) in Generalized MPLS and MPLS/TE networks, the completion of the signaling process does not necessarily mean that the cross connection along the LSP have been programmed accordingly and in a timely manner. Meanwhile, the completion of signaling process may be used by applications as indication that data path has become usable. The existence of this delay and the possible failure of cross connection programming, if not properly treated, will result in data loss or even application failure. Characterization of this performance can thus help designers to improve the application model and to build more robust applications. This document defines a series of performance metrics to evaluate the availability of data path in the signaling process.

Copyright Notice

Table of Contents

1. Introduction

Ideally, the completion of the signaling process means that the signaled label switched path (LSP) is available and is ready to carry traffic. However, in actual implementations, vendors may choose to program the cross connection in a pipelined manner, so that the overall LSP provisioning delay can be reduced. In such situations, the data path may not be available instantly after the signaling process completes. Implementation deficiency may also cause the inconsistency in between the signaling process and data path provisioning. For example, if the data plane failed to program the cross connection accordingly but does not manage to report this to the control plane, the signaling process may complete successfully while the corresponding data path will never become functional at all.

On the other hand, the completion of the signaling process may be used in many cases as indication of data path availability. For example, when invoking through User Network Interface (UNI), a client device or an application may use the reception of the correct RESV message as indication that data path is fully functional and start to transmit traffic. This will results in data loss or even application failure.

Although RSVP(-TE) specifications have suggested that the cross connections are programmed before signaling messages are propagated upstream, it is still worthwhile to verify the conformance of an implementation and measure the delay, when necessary.

This document defines a series of performance metrics to evaluate the availability of data path when the signaling process completes. The metrics defined in this document complements the control plane metrics defined in [I-D.ietf-ccamp-lsp-dppm]. They can be used to verify the conformance of implementations against related specifications, as elaborated in [I-D.shiomoto-ccamp-switch-programming]. They also can be used to build more robust applications.

2. 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].

3. Overview of Performance Metrics

In this memo, we define three performance metrics to characterize the performance of data path provisioning with GMPLS/MPLS-TE signaling. These metrics complement the metrics defined in [I-D.ietf-ccamp-lsp-dppm], in the sense that the completion of the signaling process for a Label Switched Path (LSP) and the programming of cross connections along the LSP may not be consistent. The performance metrics in [I-D.ietf-ccamp-lsp-dppm] characterize the performance of LSP provisioning from the pure signaling point of view, while the metric in this document takes into account the validity of the data path.

The three metrics are:

o RRFD - the delay between RESV message received by ingress node and forward data path becomes available.

o RSRD - the delay between RESV message sent by egress node and reverse data path becomes available.

o PRFD - the delay between PATH message received by egress node and forward data path becomes available.

As in [I-D.ietf-ccamp-lsp-dppm], we continue to use the structures and notions introduced and discussed in the IPPM Framework document, [RFC2330] [RFC2679] [RFC2681]. The reader is assumed to be familiar with the notions in those documents. The readers are assumed to be familiar with the definitions in [I-D.ietf-ccamp-lsp-dppm] as well.

4. Terms used in this document

- Forward data path - the data path from the ingress to the egress. Instances of forward data path include the data path of a uni-directional LSP and data path from the ingress node to the egress node in a bidirectional LSP.

- Reverse data path - the data path from the egress to the ingress in a bidirectional LSP.

- Error free signal - data plane specific indication of availability of the data path. For example, for packet switched interfaces, the reception of the first error free packet from one side of the LSP to the other can be used as the error free signal. For SDH/SONET cross connects, the disappearance of alarm can be used as the error free signal. Through out this document, we will use the "error free signal" as a general term. An implementations must choose a proper data path signal that is specific to the data path technology being tested.

- Ingress/egress node - in this memo, an ingress/egress node means a measurement endpoint with both control plane and data plane features. Typically, the control plane part on an ingress/egress node interact with the control plane of the network under test. The data plane part of an ingress/egress node will generate data path signals and send the signal to the data plane of the network under test, or receive data path signals from the network under test.

5. A singleton Definition for RRFD

This part defines a metric for forward data path delay when an LSP is setup.

As described in [I-D.shiomoto-ccamp-switch-programming], the completion of the RSVP-TE signaling process does not necessarily mean that the cross connections along the LSP being setup are in place and ready to carry traffic. This metric defines the time difference between the reception of RESV message by the ingress node and the completion of the cross connection programming along the forward data path.

5.1. Motivation

RRFD is useful for several reasons:

- o For the reasons described in [I-D.shiomoto-ccamp-switch-programming], the data path may not be available instantly after the completion of the RSVP-TE signaling process. The delay itself is part of the implementation performance.

- o The completion of the signaling process may be used by application designers as indication of data path availability. The existence of this delay and the potential failure of cross connection programming, if not properly treated, will result in data loss or application failure. The typical value of this delay can thus help designers to improve the application model.

5.2. Metric Name

RRFD

5.3. Metric Parameters

- o ID0, the ingress LSR ID

- o ID1, the egress LSR ID

- o T, a time when the setup is attempted

5.4. Metric Units

Either a real number of milli-seconds or undefined.

5.5. Definition

For a real number dT, RRFD from ingress node ID0 to egress node ID1 at T is dT means that ingress node ID0 send a PATH message to egress node ID1 and the last bit of the corresponding RESV message is received by ingress node ID0 at T, and an error free signal is received by egress node ID1 by using a data plane specific test pattern at T+dT.

5.6. Discussion

The following issues are likely to come up in practice:

- o The accuracy of RRFD depends on the clock resolution of both the ingress node and egress node. Clock synchronization between the ingress node and egress node is required.

- o The accuracy of RRFD is also dependent on how the error free signal is received and may differ significantly when the underline data plane technology is different. For instance, for an LSP between a pair of Ethernet interfaces, the ingress node (sometimes the tester) may use a rate based method to verify the availability of the data path and use the reception of the first error free frame as the error free signal. In this case, the interval between two successive frames has a significant impact on accuracy. It is RECOMMENDED that the ingress node uses small intervals, under the condition that the injected traffic does not exceed the capacity of the forward data path. The value of the interval MUST be reported.

- o The accuracy of RRFD is also dependent on the time needed to propagate the error free signal from the ingress node to the egress node. A typical value of propagating the error free signal from the ingress node to the egress node under the same measurement setup MAY be reported. The methodology to obtain such values is outside the scope of this document.

- o It is possible that under some implementations, a node may program the cross connection before it sends PATH message further downstream and the data path may be available before a RESV message reaches the ingress node. In such cases, RRFD can be a negetive value. It is RECOMMENDED that PRFD measurement is carried out to further characterize the forward data path delay when a negetive RRFD value is observed.

- o If error free signal is received by the egress node before PATH message is sent, an error MUST be reported and the measurement SHOULD terminate.

- o If the corresponding RESV message is received, but no error free signal is received by the egress node within a reasonable period of time, RRFD MUST be treated as undefined. The value of the threshold MUST be reported.

- o If the LSP setup fails, RRFD MUST NOT be counted.

5.7. Methodologies

Generally the methodology would proceed as follows:

- o Make sure that the network has enough resource to set up the requested LSP.

- o Start the data path measurement and/or monitoring procedures on the ingress node and egress node. If error free signal is received by the egress node before PATH message is sent, report an error and terminate the mmeasurement.

- o At the ingress node, form the PATH message according to the LSP requirements and send the message towards the egress node.

- o Upon receiving the last bit of the corresponding RESV message, take the time stamp (T1) on the ingress node as soon as possible.

- o When an error free signal is observed on the egress node, take the time stamp (T2) as soon as possible. An estimate of RRFD (T2 - T1) can be computed.

- o If the corresponding RESV message arrives, but no error free signal is received within a reasonable period of time by the ingress node, RRFD is deemed to be undefined.

- o If the LSP setup fails, RRFD is not counted.

6. A singleton Definition for RSRD

This part defines a metric for reverse data path delay when an LSP is setup.

As described in [I-D.shiomoto-ccamp-switch-programming], the completion of the RSVP-TE signaling process does not necessarily mean that the cross connections along the LSP being setup are in place and ready to carry traffic. This metric defines the time difference between the completion of the signaling process and the completion of the cross connection programming along the reverse data path. This metric MAY be used together with RRFD to characterize the data path delay of a bidirectional LSP.

6.1. Motivation

RSRD is useful for several reasons:

o For the reasons described in [I-D.shiomoto-ccamp-switch-programming], the data path may not be available instantly after the completion of the RSVP-TE signaling process. The delay itself is part of the implementation performance.

o The completion of the signaling process may be used by application designers as indication of data path availability. The existence of this delay and the possible failure of cross connection programming, if not properly treated, will result in data loss or application failure. The typical value of this delay can thus help designers to improve the application model.

6.2. Metric Name

RSRD

6.3. Metric Parameters

o ID0, the ingress LSR ID

o ID1, the egress LSR ID

o T, a time when the setup is attempted

6.4. Metric Units

Either a real number of milli-seconds or undefined.

6.5. Definition

For a real number dT, RSRD from ingress node ID0 to egress node ID1 at T is dT means that ingress node ID0 send a PATH message to egress node ID1 and the last bit of the corresponding RESV message is sent by egress node ID1 at T, and an error free signal is received by the ingress node ID0 using a data plane specific test pattern at T+dT.

6.6. Discussion

The following issues are likely to come up in practice:

- o The accuracy of RSRD depends on the clock resolution of both the ingress node and egress node. And clock synchronization between the ingress node and egress node is required.

- o The accuracy of RSRD is also dependent on how the error free signal is received and may differ significantly when the underline data plane technology is different. For instance, for an LSP between a pair of Ethernet interfaces, the egress node (sometimes the tester) may use a rate based method to verify the availability of the data path and use the reception of the first error free frame as the error free signal. In this case, the interval between two successive frames has a significant impact on accuracy. It is RECOMMENDED that in this case the egress node uses small intervals, under the condition that the injected traffic does not exceed the capacity of the reverse data path. The value of the interval MUST be reported.

- o The accuracy of RSRD is also dependent on the time needed to propagate the error free signal from the egress node to the ingress node. A typical value of propagating the error free signal from the egress node to the ingress node under the same measurement setup MAY be reported. The methodology to obtain such values is outside the scope of this document.

- o If the corresponding RESV message is sent, but no error free signal is received by the ingress node within a reasonable period of time, RSRD MUST be treated as undefined. The value of the threshold MUST be reported.

- o If error free signal is received before PATH message is sent, an error MUST be reported and the measurement SHOULD terminate.

- o If the LSP setup fails, RSRD MUST NOT be counted.

6.7. Methodologies

Generally the methodology would proceed as follows:

- o Make sure that the network has enough resource to set up the requested LSPs.

- o Start the data path measurement and/or monitoring procedures on the ingress node and egress node. If error free signal is received by the ingress node before PATH message is sent, report an error and terminate the mmeasurement.

- o At the ingress node, form the PATH message according to the LSP requirements and send the message towards the egress node.

- o Upon sending the last bit of the corresponding RESV message, take the time stamp (T1) on the egress node as soon as possible.

- o When an error free signal is observed on the ingress node, take the time stamp (T2) as soon as possible. An estimate of RSRD (T2-T1) can be computed.

- o If the LSP setup fails, RSRD is not counted.

- o If no error free signal is received within a reasonable period of time by the ingress node, RSRD is deemed to be undefined.

7. A singleton Definition for PRFD

This part defines a metric for forward data path delay when an LSP is setup.

In an RSVP-TE implementation, when setting up an LSP, each node may choose to program the cross connection before it sends PATH message further downstream. In this case, the forward data path may become available before the signaling process completes, ie. before the RESV reaches the ingress node. This metric can be used to identify such implementation practice and give useful information to application designers.

7.1. Motivation

PRFD is useful for the following reasons:

- PRFD can be used to identify an RSVP-TE implementation practice, in which cross connections are programmed before PATH message is sent downtream.

- The value of PRFD may also help application designers to fine tune their application model.

7.2. Metric Name

PRFD

7.3. Metric Parameters

- ID0, the ingress LSR ID

- ID1, the egress LSR ID

- T, a time when the setup is attempted

7.4. Metric Units

Either a real number of milli-seconds or undefined.

7.5. Definition

For a real number dT, PRFD from ingress node ID0 to egress node ID1 at T is dT means that ingress node ID0 send a PATH message to egress node ID1 and the last bit of the PATH message is received by egress node ID1 at T, and an error free signal is received by the egress node ID1 using a data plane specific test pattern at T+dT.

7.6. Discussion

The following issues are likely to come up in practice:

- The accuracy of PRFD depends on the clock resolution of the egress node. And clock synchronization between the ingress node and egress node is not required.

- The accuracy of PRFD is also dependent on how the error free signal is received and may differ significantly when the underline data plane technology is different. For instance, for an LSP between a pair of Ethernet interfaces, the egress node (sometimes the tester) may use a rate based method to verify the availability of the data path and use the reception of the first error free frame as the error free signal. In this case, the interval between two successive frames has a significant impact on accuracy. It is RECOMMENDED that in this case the ingress node uses small intervals, under the condition that the injected traffic does not exceed the capacity of the forward data path. The value of the interval MUST be reported.

- The accuracy of PRFD is also dependent on the time needed to propagate the error free signal from the ingress node to the egress node. A typical value of propagating the error free signal from the ingress node to the egress node under the same measurement setup MAY be reported. The methodology to obtain such values is outside the scope of this document.

- If error free signal is received before PATH message is sent, an error MUST be reported and the measurement SHOULD terminate.

- If the LSP setup fails, PRFD MUST NOT be counted.

- This metric SHOULD be used together with RRFD. It is RECOMMENDED that PRFD measurement is carried out after a negetive RRFD value has already been observed.

7.7. Methodologies

Generally the methodology would proceed as follows:

- Make sure that the network has enough resource to set up the requested LSPs.

- Start the data path measurement and/or monitoring procedures on the ingress node and egress node. If error free signal is received by the egress node before PATH message is sent, report an error and terminate the mmeasurement.

- o At the ingress node, form the PATH message according to the LSP requirements and send the message towards the egress node.

- o Upon receiving the last bit of the PATH message, take the time stamp (T1) on the egress node as soon as possible.

- o When an error free signal is observed on the egress node, take the time stamp (T2) as soon as possible. An estimate of PRFD (T2-T1) can be computed.

- o If the LSP setup fails, PRFD is not counted.

- o If no error free signal is received within a reasonable period of time by the egress node, PRFD is deemed to be undefined.

8. A Definition for Samples of Data Path Delay

In Section Section 5, Section 6 and Section 7, we define the singleton metrics of data path delay. Now we define how to get one particular sample of such delay. Sampling is to select a particular potion of singleton values of the given parameters. Like in [RFC2330], we use Poisson sampling as an example.

8.1. Metric Name

Type <X> Data path delay sample, where X is either RRFD, RSRD or PRFD.

8.2. Metric Parameters

- o ID0, the ingress LSR ID
- o ID1, the egress LSR ID
- o T0, a time
- o Tf, a time
- o Lambda, a rate in the reciprocal seconds
- o Th, LSP holding time
- o Td, the maximum waiting time for successful LSP setup
- o Ts, the maximum waiting time for error free signal

8.3. Metric Units

A sequence of pairs; the elements of each pair are:

- o T, a time when setup is attempted
- o dT, either a real number of milli-seconds or undefined

8.4. Definition

Given T0, Tf, and lambda, compute a pseudo-random Poisson process beginning at or before T0, with average arrival rate lambda, and ending at or after Tf. Those time values greater than or equal to T0 and less than or equal to Tf are then selected. At each of the times in this process, we obtain the value of data path delay sample of type <X> at this time. The value of the sample is the sequence made up of the resulting <time, type <X> data path delay> pairs. If there are no such pairs, the sequence is of length zero and the sample is said to be empty.

8.5. Discussion

The following issues are likely to come up in practice:

- o The parameters lambda, Th and Td should be carefully chosen, as explained in the discussions for LSP setup delay.
- o The parameter Ts should be carefully chosen and MUST be reported along with the LSP forward/reverse data path delay sample.
- o Note that for online or passive measurements, the holding time of an LSP is determined by actual traffic, hence in this case Th is not an input parameter.

8.6. Methodologies

Generally the methodology would proceed as follows:

- o The selection of specific times, using the specified Poisson arrival process, and
- o Set up the LSP and obtain the value of type <X> data path delay
- o Release the LSP after Th, and wait for the next Poisson arrival process

8.7. Typical testing cases

8.7.1. With No LSP in the Network

8.7.1.1. Motivation

Data path delay with no LSP in the network is important because this reflects the inherent delay of a device implementation. The minimum value provides an indication of the delay that will likely be experienced when an LSP data path is configured under light traffic load.

8.7.1.2. Methodologies

Make sure that there is no LSP in the network, and proceed with the methodologies described in Section 8.6.

8.7.2. With a Number of LSPs in the Network

8.7.2.1. Motivation

Data path delay with a number of LSPs in the network is important because it reflects the performance of an operational network with considerable load. This delay may vary significantly as the number of existing LSPs varies. It can be used as a scalability metric of a device implementation.

8.7.2.2. Methodologies

Setup the required number of LSPs, and wait until the network reaches a stable state, and then proceed with the methodologies described in Section 8.6.

9. Some Statistics Definitions for Metrics to Report

Given the samples of the performance metric, we now offer several statistics of these samples to report. From these statistics, we can draw some useful conclusions of a GMPLS network. The value of these metrics is either a real number, or an undefined number of milliseconds. In the following discussion, we only consider the finite values.

9.1. The Minimum of Metric

The minimum of metric is the minimum of all the dT values in the sample. In computing this, undefined values SHOULD be treated as infinitely large. Note that this means that the minimum could thus be undefined if all the dT values are undefined. In addition, the metric minimum SHOULD be set to undefined if the sample is empty.

9.2. The Median of Metric

Metric median is the median of the dT values in the given sample. In computing the median, the undefined values MUST NOT be counted in.

9.3. The percentile of Metric

Given a metric and a percent X between 0% and 100%, the Xth percentile of all the dT values in the sample. In addition, the percentile is undefined if the sample is empty.

Example: suppose we take a sample and the results are: Stream1 = <<T1, 100 msec>, <T2, 110 msec>, <T3, undefined>, <T4, 90 msec>, <T5,500 msec>>. Then the 50th percentile would be 110 msec, since 90 msec and 100 msec are smaller, and 110 and 500 msec are larger (undefined values are not counted in).

9.4. The Failure Probability

In the process of LSP setup/release, it may fail for some reason. The failure probability is the ratio of the unsuccessful times to the total times.

10. Security Considerations

In the control plane, since the measurement endpoints must be conformant to signaling specifications and behave as normal signaling endpoints, it will not incur other security issues than normal LSP provisioning. However, the measurement parameters must be carefully selected so that the measurements inject trivial amounts of additional traffic into the networks they measure. If they inject "too much" traffic, they can skew the results of the measurement, and in extreme cases cause congestion and denial of service.

In the data plane, the measurement endpoint MUST use a signal that is consistent with what is specified in the control plane. For example, in a packet switched case, the traffic injected into the data plane MUST NOT exceed the specified rate in the corresponding LSP setup request. In a wavelength switched case, the measurement endpoint MUST use the specified or negotiated lambda with appropriate power.

The security considerations pertaining to the original RSVP protocol [RFC2205] and its TE extensions [RFC3209] also remain relevant.

11. IANA Considerations

This document makes no requests for IANA action.

12. Acknowledgements

We wish to thank Adrian Farrel and Lou Berger for their comments and helps.

This document contains ideas as well as text that have appeared in existing IETF documents. The authors wish to thank G. Almes, S. Kalidindi and M. Zekauskas.

We also wish to thank Weisheng Hu, Yaohui Jin and Wei Guo in the state key laboratory of advanced optical communication systems and networks for the valuable comments. We also wish to thank the support from NSFC and 863 program of China.

13. References

13.1. Normative References

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

[RFC2205] Braden, B., Zhang, L., Berson, S., Herzog, S., and S. Jamin, "Resource ReSerVation Protocol (RSVP) -- Version 1 Functional Specification", RFC 2205, September 1997.

[RFC2679] Almes, G., Kalidindi, S., and M. Zekauskas, "A One-way Delay Metric for IPPM", RFC 2679, September 1999.

[RFC2681] Almes, G., Kalidindi, S., and M. Zekauskas, "A Round-trip Delay Metric for IPPM", RFC 2681, September 1999.

[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.

13.2. Informative References

[I-D.ietf-ccamp-lsp-dppm] Sun, W., Zhang, G., Gao, J., Xie, G., Papneja, R., Gu, B., Wei, X., Otani, T., and R. Jing, "Label Switched Path (LSP) Dynamic Provisioning Performance Metrics in Generalized MPLS Networks", draft-ietf-ccamp-lsp-dppm-10 (work in progress), October 2009.

[I-D.shiomoto-ccamp-switch-programming] Shiomoto, K. and A. Farrel, "Advice on When It is Safe to Start Sending Data on Label Switched Paths Established Using RSVP-TE", draft-shiomoto-ccamp-switch-programming-01 (work in progress), October 2009.

[RFC2330] Paxson, V., Almes, G., Mahdavi, J., and M. Mathis, "Framework for IP Performance Metrics", RFC 2330, May 1998.

Authors' Addresses

Weiqiang Sun, Editor
Shanghai Jiao Tong University
800 Dongchuan Road
Shanghai 200240
China

Phone: +86 21 3420 5359
Email: sunwq@mit.edu

Guoying Zhang, Editor
China Academy of Telecommunication Research, MIIT, China.
No.11 YueTan South Street
Beijing 100045
China

Phone: +86 1068094272
Email: zhangguoying@mail.ritt.com.cn

Jianhua Gao
Huawei Technologies Co., LTD.
China

Phone: +86 755 28973237
Email: gjhhit@huawei.com

Guowu Xie
University of California, Riverside
900 University Ave.
Riverside, CA 92521
USA

Phone: +1 951 237 8825
Email: xieg@cs.ucr.edu

Rajiv Papneja
Isocore
12359 Sunrise Valley Drive, STE 100
Reston, VA 20190
USA

Phone: +1 703 860 9273
Email: rpapneja@isocore.com

Contributors

Bin Gu
IXIA
Oriental Kenzo Plaza 8M, 48 Dongzhimen Wai Street, Dongcheng District
Beijing 200240
China

Phone: +86 13611590766
Email: BGu@ixiacom.com

Xueqin Wei
Fiberhome Telecommunication Technology Co., Ltd.
Wuhan
China

Phone: +86 13871127882
Email: xqwei@fiberhome.com.cn

Tomohiro Otani
KDDI R&D Laboratories, Inc.
2-1-15 Ohara Kamifukuoka Saitama
356-8502
Japan

Phone: +81-49-278-7357
Email: otani@kddilabs.jp

Ruiquan Jing
China Telecom Beijing Research Institute
118 Xizhimenwai Avenue
Beijing 100035
China

Phone: +86-10-58552000
Email: jingrq@ctbri.com.cn