TEAS Working Group
Internet Draft
Intended status: Best Current Practice

Ravi Singh
Juniper Networks
Rob Shakir
British Telecom
Vishnu Pavan Beeram
Juniper Networks
Tarek Saad
Cisco Systems

Expires: January 2, 2016

July 2, 2015

RSVP Setup Retry - BCP
draft-ravisingh-teas-rsvp-setup-retry-01

Status of this Memo

Copyright Notice

Abstract

This document discusses the best current practices associated with the implementation of RSVP setup-retry timer.

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 RFC-2119 [RFC2119].

Table of Contents

1. Introduction

In an RSVP-TE network with a very large number of LSPs, link/node failure(s) may produce a noticeable increase in RSVP-TE control traffic. As a result, RSVP-TE messages might get delayed by virtue of being stuck in a queue that is overwhelmed with messages to be sent or they might get lost forever. For example, a Path message intended to be sent by a transit router might be stuck in the output queue to be sent to the next-hop. Alternately, it might have got dropped on the receive side due to queue overflows. The same could happen for a Resv message in the reverse direction. Also, in the absence of reliable delivery of Path-Error messages [RFC2961], an error that gets generated at transit/egress for an LSP that is in the process of being setup may never make it to the ingress.

Lost/delayed RSVP-TE messages cause the following problems for an ingress router:

- In the absence of an error indication, how is an ingress to know that an LSP for which signaling was (re-)initiated and a Resv has not yet been received, is ever going to come up?
- In the absence of any indication, what action should the ingress take to support low-latency LSP-setup?

The above problems essentially boil-down to: how long should the ingress continue to wait before giving up on its attempt to bring up the LSP, and take some alternative course of action (e.g., try to bring up the LSP on an alternate path)?. To mitigate this problem, some implementations use a setup-retry timer mechanism. This document discusses the issues associated with a particular implementation of this timer and makes some specific recommendations to get around these issues.

2. Setup-Retry Timer

The setup-retry timer is usually a configurable timer which (in the absence of an error indication) goes off when an LSP with a given LSPID has not received the corresponding Resv in response to its Path during a pre-configured duration after its first Path had been sent.

Use of the setup-retry timer is based on the presumption that if signaling for a given LSP has not been completed within an "expected" duration, it is not going to be completed at all. The intent in the use of this timer is to expeditiously take some alternative course of action when an LSP has not yet completed its signaling within an "expected" duration of time.

3. Possible ill-effects due to implementation choices

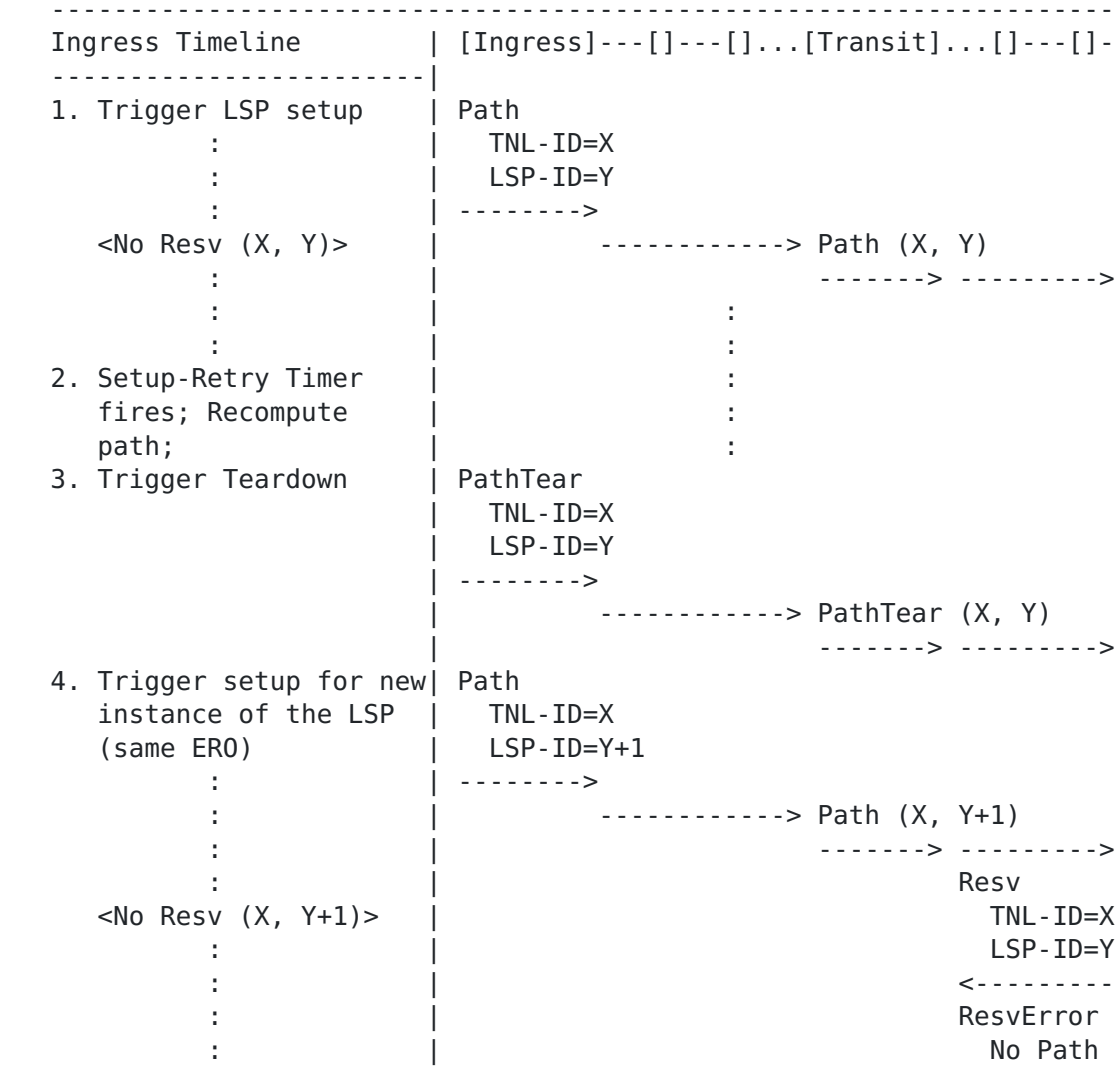
As mentioned in the previous section, the intent in the use of this timer is to take some alternative course of action when an LSP has not yet completed its signaling within an "expected" duration of time. One such course of action is for the ingress router to initiate tear-down for the previously in-the-process-of-being-signaled path via a PathTear; run CSPF; and use the outcome of this CSPF to signal the brand-new path for this tunnel with a different LSP-ID, typically, bumped up by 1. This section describes the problems caused by such course of action.

As mentioned in Section 1, in a network with a very large number of RSVP-TE LSPs, link/node failure(s) may produce a noticeable increase

in the volume of RSVP-TE control traffic, which in turn might cause a router to either drop RSVP-TE messages or alternately cause them to be sent excessively late.

As a result, the following problems can occur:
- LSP setup latency might be excessively high.
- Error messages that indicate failure in LSP setup might not make it to the ingress router.

A mix of the above problems can cause the setup-retry timer for a given LSP (at the ingress router) to fire repeatedly over a period of time. The situation being such the ingress gets stuck in a cycle as illustrated below for some/many LSPs:



In the above illustration, notice how the transit router never gets to completely process the "current" LSP-ID (see [RShakir] for more). The implementation recommendations made in this document will help avoid this snowball effect.

4. Causes of the above ill-effects

The implementation issues listed in section 3 end up causing an increase in the control plane load on a network whose control plane is already under stress. The foregoing is caused by unnecessarily doing the following even when there is no change in the computed path:

- Sending PathTears causes excessive and unjustifiable work on those downstream routers on the "previous ERO path" that had managed to bring the LSP UP. In other words, the slowness of a given transit router should not be the cause to penalize all other transit routers downstream of it, as doing so just increases the overall network stress.

- Sending Path for LSPID=Y+1 causes unnecessary work for all routers on the ERO path including those that were already running slow and were the real cause of the Resv for LSPDID=Y not having been received timely by the ingress.

5. Solution to the implementation issues

To eliminate causes of the ill-effects listed in the previous section and thus to eliminate the ill-effects, this document makes the following recommendations.

When the setup-retry timer fires:

If there is no change in the computed path (no error indication for that LSP has been received via a PathErr or a TE update indicating a failure),
- Do not send PathTear for LSPID=Y
- Just let the Path State get refreshed for LSPID=Y.

The recommended default behavior is to keep retrying until the path changes or the user intervenes. Implementations MAY choose to

provide the user with an option to override this default behavior and specify a policy to determine when to stop retrying.

Implementations SHOULD use the recommendations listed in this section to avoid getting stuck in a LSP signaling hysteresis.

6. Security Considerations

This document does not introduce any new security concerns.

7. IANA Considerations

None.

8. Normative References

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

[RShakir] Rob Shakir, "The next spring forward", http://rob.sh/files/the-next-spring-forward_rjs120314.pdf March 2014.

[RFC2961] Berger, L., "RSVP Refresh Overhead Reduction Extensions", RFC 2961, April 2001.

9. Acknowledgments

The authors would like to thank Yakov Rekhter and Raveendra Torvi for their inputs.

10. Authors' Addresses

Ravi Singh
Juniper Networks
Email: ravis@juniper.net

Rob Shakir
British Telecom
Email: rob.shakir@bt.com

Tarek Saad
Cisco Systems
Email: tsaad@cisco.com

Vishnu Pavan Beeram

Juniper Networks
Email: vbeeram@juniper.net

Contributors

Markus Jork
Juniper Networks
Email: mjork@juniper.net

Aman Kapoor
Juniper Networks
Email: amanka@juniper.net