Workgroup: TEAS Working Group
Internet-Draft:
draft-saad-teas-rsvpte-ip-tunnels-02
Published: 12 April 2022
Intended Status: Standards Track
Expires: 14 October 2022

Authors: T. Saad   V.P. Beeram
         Juniper Networks   Juniper Networks

IP RSVP-TE: Extensions to RSVP for P2P IP-TE LSP Tunnels

Abstract

This document describes the use of RSVP (Resource Reservation Protocol), including all the necessary extensions, to establish Point-to-Point (P2P) Traffic Engineered IP (IP-TE) Label Switched Path (LSP) tunnel(s) for use in native IP forwarding networks.

This document proposes specific extensions to the RSVP protocol to allow the establishment of explicitly routed IP paths using RSVP as the signaling protocol. The result is the instantiation of an IP Path which can be automatically routed away from network failures, congestion, and bottlenecks.

Status of This Memo

Copyright Notice

Table of Contents

1. Introduction

In native IP networks, each router runs a routing protocol to determine the best next-hop(s) to a specific destination. The best next-hop(s) are usually determined by favoring those that run along the shortest path to the destination. When data flows across the network, it is routed hop-by-hop and follows the selected path by each hop towards that destination on each hop.

It is sometimes desirable for an ingress router to be able to steer traffic towards a destination along a pre-determined or pre-computed path that may follow a path other than the default shortest path. For example, some flows mayrequire to be forwarded along the least latency path. Others, may desire to be routed with bandwidth guarantees along the selected path, or along a path that honors certain resource affinities or Shared Risk Link Group (SRLG) memberships.

A solution to such use-cases entails: 1) router(s) in the network to be able to maintain and disseminate per link state information, 2) ingress routers or an external Path Computation Engine (PCE) to be able to perform a stateful path computation for feasible path(s) on top of the network topology, and 3) for ingress router(s) to be able to steer or tunnel the traffic along the established path towards the destination.

Mechanisms have been defined to achieve this with RSVP extensions for Traffic Engineered Multiprotocol Label Switching (MPLS-TE) networks as described in [RFC3209]. This document proposes extensions to the existing mechanisms for achieving this in networks that rely on native IP for their forwarding.

This document covers the necessary extensions for establishing Point-to-Point (P2P) Traffic-Engineered IP (IP-TE) Label Switched Path (LSP) Tunnels. The equivalent extensions needed for setting up multicast IP-TE LSPs are currently out of the scope of this document.

2. Terminology

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "NOT RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in BCP 14 [RFC2119] [RFC8174] when, and only when, they appear in all capitals, as shown here.

2.1. Acronyms

The reader is assumed to be familiar with the terminology used in [RFC2205] and [RFC3209].

IP-TE LSP (Traffic Engineered IP Label Switched Path):

The path created by programming of an IP route along the explicitly specified or dynamically computed sequence of router hops, allowing an IP packet to be forwarded from one hop to another along the established path.

IP-TE LSP Tunnel:

An IP-TE LSP which is used to tunnel traffic over the pre-established IP path.

Traffic Engineered IP Tunnel (IP-TE Tunnel):

A set of one or more IP-TE LSP Tunnels which carries a traffic trunk.

3. Overview of IP LSP Tunnels

IP-TE LSP tunnels are established over a native IP forwarding network. In many cases, IP-TE LSP(s) are explicitly routed from an ingress router. The explicit route used to establish an IP-TE LSP may be locally computed at the ingress router, or externally computed by an entity such as a Path Computation Element (PCE) [RFC4655].

To support the setup of IP-TE LSP tunnel(s), the egress routers reserve one or more local IP prefixes or Egress Address Block(s) (EABs) that are dedicated for RSVP to establish IP-TE LSP(s) tunnels.

The EAB(s) addresses at the egress router may be managed by the RSVP protocol and are not required to be exchanged by any other routing protocol.

It is possible in some cases, where the IP-TE LSP(s) are contained within a single administrative domain boundary, for EAB(s) to be allocated from the private IP address space as defined in [RFC1918] or from the unique-local space as defined in [RFC4193] and [RFC6890].

Also useful in some applications for sets of IP-TE LSP tunnels to be associated together to facilitate reroute operations or to spread a traffic trunk over multiple IP-TE LSP tunnel paths. For traffic engineering applications to IP-TE LSP tunnel(s), such sets are called traffic engineered tunnels (TE IP tunnels).

3.1. Creation and Management

An IP-TE LSP tunnel is unidirectional in nature. To create an IP-TE LSP tunnel, the ingress router of the IP-TE LSP tunnel creates an RSVP Path message with a session type of LSP_TUNNEL_IPv4 or LSP_TUNNEL_IPv6 and follows the procedures outlined in [RFC3473] to insert a Generalized Label Request object into the Path message. The Generalized Label Request object indicates that an IP address binding is requested to the IP-TE LSP tunnel. The binding of an EAB address to an IP-TE LSP tunnel happens at the egress router and is signaled using an RSVP Resv message sent from the egress router.

The ingress router uses a pre-computed explicit path to populate the EXPLICIT_ROUTE object that is added the RSVP Path message. The explicitly routed path can be administratively specified, or automatically computed by a suitable entity based on QoS and policy requirements, taking into consideration the prevailing network state. In addition, RSVP-TE signaling [RFC3209] allows for the specification of an explicit path as a sequence of strict and loose routes. Such combination of abstract nodes, and strict and loose routes significantly enhances the flexibility of path definitions.

The ingress MAY also add a RECORD_ROUTE object to the RSVP Path message in order to receive information about the actual route traversed by the IP-TE LSP tunnel. The RECORD_ROUTE object MAY also be used by the egress router to determine whether Shared Forwarding as described in Section 3.7 is possible amongst different IP-TE LSP tunnel(s).

3.2. Path Maintenance

If the ingress router discovers a better path, after an IP-TE LSP tunnel has been successfully established, it can dynamically reroute the session by changing the EXPLICIT_ROUTE object. If problems are encountered with the EXPLICIT_ROUTE object, either because it causes a routing loop or because some intermediate routers do not support it, the ingress is notified.

Make-before-break procedures can also be employed to modify the characteristics of an IP-TE LSP tunnel. As described in [RFC3209], the LSP ID in the Sender Template object is updated in the new RSVP Path message that is signaled. As usual, the combination of the LSP_TUNNEL SESSION object and the SE reservation style naturally accommodates smooth transitions in bandwidth and routing.

For example, to trigger a bandwidth increase, a new RSVP Path Message with a new LSP_ID can be used to attempt a larger bandwidth reservation while the current LSP_ID continues to be refreshed to ensure that the reservation is not lost if the larger reservation fails.

3.3. Signaling Extensions

This section describes RSVP signaling extensions and modifications to existing RSVP objects that are carried in RSVP Path or Resv messages and are required to establish IP-TE LSP tunnel(s).

3.3.1. RSVP Path message

To signal an IP-TE LSP tunnel, the Generalized Label Request object is carried in the RSVP Path message and used to request an IP address binding to the IP-TE LSP tunnel.

The Generalized Label Request is defined in [RFC3471] and has the below format:

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
| LSP Enc. Type |Switching Type |             G-PID             |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

To request an IPv4 or IPv6 binding to an IP-TE LSP tunnel, the Generalized Label object carries the following specifics:

1. the LSP encoding type is set to Packet (1) [RFC3471].

2. the LSP switching type is set to "IPv4-TE" (TBD1), or IPv6-TE (TBD2)

3. the Generalized Payload Identifier (G-PID) MAY be set to All (0) or in some cases to the specific payload type if known, e.g. Ethernet (33) [RFC3471].

3.4. RSVP Resv Label Object

The egress is responsible to bind an IP EAB address to an IP-TE LSP tunnel.

Once the egress router receives the RSVP Path message with the Generalized Label Request object containing the parameters described in Section 3.3.1, the egress router determines and binds an EAB address to the newly established IP-TE LSP tunnel. Note, subject to a local policy and additional path check(s), the egress MAY assign an already in used EAB address to the newly established IP-TE LSP tunnel.

The RSVP Resv message that is created by the egress router uses the Generalized Label defined in [RFC3471] to carry the EAB address that is bound to newly established IP-TE LSP tunnel.

The RSVP Generalized Label object has the following format:

LABEL class = 16, C_Type = 2

The information carried in a Generalized Label is:

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                             Label                             |
|                              ...                              |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Label (Variable Length):

Carries label information. The interpretation of this field depends the parameters signaled in the Generalized Label Request.

3.5. EAB address Handling

The RSVP Resv message that is created by the egress router is forwarded upstream along the signaling path towards the ingress router. Each router starting from the egress will perform the following steps when binding the EAB address to the IP-TE LSP tunnel.

3.5.1. Egress Router

The egress router manages the EAB addresses for the use of establishing IP LSP tunnel(s).

The egress router MAY assign unique EAB address to newly established IP-TE LSP tunnel(s) and MAY free an existing EAB address upon destroying a previously established IP-TE LSP tunnel. Note that an egress router MAY hold on to an EAB when the IP-TE LSP is being destroyed if it determines other IP-TE LSP(s) are sharing it.

Once an EAB address is allocated and bound to a new IP-TE LSP tunnel, the egress router programs the address in its forwarding table as local address - hence, resulting in decapsulation of the outer IP header on any packet arriving over the IP-TE LSP tunnel and hence yielding the original IP datagram that was tunneled over the IP LSP tunnel,

3.5.2. Ingress and Transit Router

A transit or an ingress router extracts the EAB address that the egress router binds to the IP-TE LSP tunnel from the Generalized Label object contained in the RSVP Resv message that is propagated upstream as described in Section 3.4. The transit or ingress router uses the EAB address to program an IP route in the Routing Information Base (RIB) and uses the previously signaled EXPLICIT_ROUTE object to derive the next-hop information associated with the EAB route at that hop.

An advantage of using RSVP to establish IP-TE LSP tunnels is that it enables the allocation of resources along the path. For example, bandwidth can be allocated to each IP-TE LSP tunnel using standard RSVP reservations as described in [RFC3209].

3.6. Protection

Fast Reroute (FRR) procedures that are defined in [RFC4090] describe the mechanisms for router along the LSP path to act as a Point of

Local Repair (PLR) and reroute traffic and signaling of a protected RSVP-TE LSP onto a pre-established bypass tunnel in the event of a protected TE link or node failure.

Similar mechanisms can be employed for protecting IP-TE LSP tunnel(s) in IP network(s). An ingress or transit router acting as potential PLR can pre-establish bypass tunnel(s) that protect the primary IP-TE LSP tunnel against the protected link or downstream node failure.

Upon failure of the protected link, the traffic arriving over the protected IP-TE LSP on the PLR is automatically tunneled over the pre-established bypass IP-TE LSP tunnel and packets are forwarded towards the Merge Point (MP) router. At the MP router, the incoming IP packets are decapsulated exposing the original IP header of the protected IP-TE LSP tunnel. The packets are forwarded downstream of the MP router along the

3.7. Shared Forwarding

One capability of the IP data plane is its ability to reuse the IP forwarding entry when setting up IP-TE LSP(s) from multiple sources and that share a common destination. This capability MAY be preserved provided certain requirements are met. We refer to this capability as "Shared Forwarding". Shared Forwarding is a local policy local to egress router responsible for binding an EAB address to the signaled IP-TE LSP tunnel.

The Shared Forwarding function allows the reduction of forwarding entries on any transit router RIB. The Shared forwarding paths are identical in function to independently routed Multi-point to Point (MP2P) paths that share part of their path(s) from the intersecting router and towards the egress router.

If the egress router policy allows for Shared Forwarding, and upon signaling a new IP-TE LSP tunnel, the egress inspects the recorded path (extracted from the RECORD_ROUTE object). If the egress router determines that the newly signaled IP-TE LSP path intersects and merges with other IP-TE LSP from the intersection point to the egress, and if Shared Forwarding is enabled, it MUST assign the same EAB address bound to the existing IP-TE LSP tunnel.

Note, forwarding memory savings from Shared Forwarding can be quite dramatic in some topologies where a high degree of meshing is required.

3.8. Error Conditions

This section will be updated in future revisions of this document.

4. Next Steps

The authors of this document are following up with the DetNet Working Group on ways to leverage this solution to signal and establish a TE IP path for a DetNet IP flow. The DetNet IP data plane uses "6-tuple" based flow identification as described in [I-D.ietf-detnet-ip].

A new revision of this document will be posted to describe the extensions required to signal the necessary flow identification so it can be programmed on all hops of the IP Path.

5. IANA Considerations

This section will be updated in future revisions of this document.

6. Security Considerations

This section will be updated in future revisions of this document.

7. Acknowledgement

The authors would like to thank Igor Bryskin for providing valuable feedback to this document.

8. Contributors

Raveendra Torvi
Juniper Networks

Email: rtorvi@juniper.net

9. References

9.1. Normative References

[I-D.ietf-detnet-ip] Varga, B., Farkas, J., Berger, L., Fedyk, D., and S. Bryant, "Deterministic Networking (DetNet) Data Plane: IP", Work in Progress, Internet-Draft, draft-ietf-detnet-ip-07, 3 July 2020, <https://www.ietf.org/archive/id/draft-ietf-detnet-ip-07.txt>.

[RFC1918] Rekhter, Y., Moskowitz, B., Karrenberg, D., de Groot, G. J., and E. Lear, "Address Allocation for Private Internets", BCP 5, RFC 1918, DOI 10.17487/RFC1918, February 1996, <https://www.rfc-editor.org/info/rfc1918>.

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, DOI 10.17487/

RFC2119, March 1997, <https://www.rfc-editor.org/info/rfc2119>.

[RFC2205] Braden, R., Ed., Zhang, L., Berson, S., Herzog, S., and S. Jamin, "Resource ReSerVation Protocol (RSVP) -- Version 1 Functional Specification", RFC 2205, DOI 10.17487/RFC2205, September 1997, <https://www.rfc-editor.org/info/rfc2205>.

[RFC3209] Awduche, D., Berger, L., Gan, D., Li, T., Srinivasan, V., and G. Swallow, "RSVP-TE: Extensions to RSVP for LSP Tunnels", RFC 3209, DOI 10.17487/RFC3209, December 2001, <https://www.rfc-editor.org/info/rfc3209>.

[RFC3471] Berger, L., Ed., "Generalized Multi-Protocol Label Switching (GMPLS) Signaling Functional Description", RFC 3471, DOI 10.17487/RFC3471, January 2003, <https://www.rfc-editor.org/info/rfc3471>.

[RFC3473] Berger, L., Ed., "Generalized Multi-Protocol Label Switching (GMPLS) Signaling Resource ReserVation Protocol-Traffic Engineering (RSVP-TE) Extensions", RFC 3473, DOI 10.17487/RFC3473, January 2003, <https://www.rfc-editor.org/info/rfc3473>.

[RFC4090] Pan, P., Ed., Swallow, G., Ed., and A. Atlas, Ed., "Fast Reroute Extensions to RSVP-TE for LSP Tunnels", RFC 4090, DOI 10.17487/RFC4090, May 2005, <https://www.rfc-editor.org/info/rfc4090>.

[RFC4193] Hinden, R. and B. Haberman, "Unique Local IPv6 Unicast Addresses", RFC 4193, DOI 10.17487/RFC4193, October 2005, <https://www.rfc-editor.org/info/rfc4193>.

[RFC8174] Leiba, B., "Ambiguity of Uppercase vs Lowercase in RFC 2119 Key Words", BCP 14, RFC 8174, DOI 10.17487/RFC8174, May 2017, <https://www.rfc-editor.org/info/rfc8174>.

9.2. Informative References

[RFC4655] Farrel, A., Vasseur, J.-P., and J. Ash, "A Path Computation Element (PCE)-Based Architecture", RFC 4655, DOI 10.17487/RFC4655, August 2006, <https://www.rfc-editor.org/info/rfc4655>.

[RFC6890] Cotton, M., Vegoda, L., Bonica, R., Ed., and B. Haberman, "Special-Purpose IP Address Registries", BCP 153, RFC 6890, DOI 10.17487/RFC6890, April 2013, <https://www.rfc-editor.org/info/rfc6890>.

Authors' Addresses

Tarek Saad
Juniper Networks

Email: tsaad@juniper.net

Vishnu Pavan Beeram
Juniper Networks

Email: vbeeram@juniper.net