SPRING Working Group
Internet-Draft
Intended status: Informational
Expires: August 19, 2021

T. Saad
V. Beeram
C. Barth
Juniper Networks, Inc.
S. Sivabalan
Ciena Corporation.

February 15, 2021

Segment-Routing over Forwarding Adjacency Links
draft-saad-sr-fa-link-03

Abstract

Label Switched Paths (LSPs) set up in Multiprotocol Label Switching (MPLS) networks can be used to form Forwarding Adjacency (FA) links that carry traffic in those networks. An FA link can be assigned Traffic Engineering (TE) parameters that allow other LSR(s) to include it in their constrained path computation. FA link(s) can be also assigned Segment-Routing (SR) segments that enable the steering of traffic on to the associated FA link(s). The TE and SR attributes of an FA link can be advertised using known protocols that carry link state information. This document elaborates on the usage of FA link(s) and their attributes in SR enabled networks.

Table of Contents

1. Introduction 2
2. Terminology 3
3. Forwarding Adjacency Links 3
3.1. Creation and Management 4
3.2. Link Flooding 4
3.3. Underlay LSP(s) 5
3.4. State Changes 5
3.5. TE Parameters 5
3.6. Link Local and Remote Identifiers 6
4. Segment-Routing over FA Links 6
4.1. SR IGP Segments for FA 7
4.1.1. Parallel Adjacencies 7
4.2. SR BGP Segments for FA 7
4.3. Applicability to Interdomain 8
5. IANA Considerations 9
6. Security Considerations 9
7. Acknowledgement 9
8. Normative References 9
Authors' Addresses 10

1. Introduction

To improve scalability in Multi-Protocol Label Switching (MPLS) networks, it may be useful to create a hierarchy of LSPs as Forwarding Adjacencies (FA). The concept of FA link(s) and FA-LSP(s) was introduced in [RFC4206].

In Segment-Routing (SR), this is particularly useful for two main reasons.

First, it allows the stitching of sub-path(s) so as to realize an end-to-end SR path. Each sub-path can be represented by a FA link that is supported by one or more underlying LSP(s). The underlying LSP(s) that support an FA link can be setup using different technologies- including RSVP-TE, LDP, and SR. The sub-path(s), or FA link(s) in this case, can possibly interconnect multiple administrative domains, allowing each FA link within a domain to use a different technology to setup the underlying LSP(s).

Second, it allows shortening of a large SR Segment-List by compressing one or more slice(s) of the list into a corresponding FA TE link that each can be represented by a single segment- see Section 4. Effectively, it reduces the number of segments that an ingress router has to impose to realize an end-to-end path.

The FA links are treated as normal link(s) in the network and hence it can leverage existing link state protocol extensions to advertise properties associated with the FA link. For example, Traffic-Engineering (TE) link parameters and Segment-Routing (SR) segments parameters can be associated with the FA link and advertised throughout the network.

Once advertised in the network using a suitable protocols that support carrying link state information, such as OSPF, ISIS or BGP Link State (LS)), other LSR(s) in the network can use the FA TE link(s) as well as possibly other normal TE link(s) when performing path computation and/or when specifying the desired explicit path.

Though the concepts discussed in this document are specific to MPLS technology, these are also extensible to other dataplane technologies - e.g. SRv6.

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.

3. Forwarding Adjacency Links

FA Link(s) can be created and supported by underlying FA LSPs. The FA link is of type point-to-point. FA links may be represented as either unnumbered or numbered. The nodes connected by an FA link do not usually establish a routing adjacency over the FA link. When FA links are numbered with IPv4 addresses, the local and remote IPv4 addresses can come out of a /31 that is allocated by the LSR that originates the FA-LSP. For unnumbered FA link(s), other provisions may exist to exchange link identifier(s) between the endpoints of the FA.

3.1. Creation and Management

In general, the creation/termination of an FA link and its FA-LSP is driven either via configuration on the LSR at the head-end of the adjacency, or dynamically using suitable North Bound Interface (NBI) protocol, e.g. Netconf, gRPC, PCEP, etc.

The following FA-LSP attributes may be configured, including: bandwidth and resource colors, and other constraints. The path taken by the FA-LSP may be either computed by the LSR at the head-end of the FA-LSP, or externally by a PCE and furnished to the headend.

The attributes of the FA link can be inherited from the underlying LSP(s) that induced its creation. In general, for dynamically provisioned FAs, a policy-based mechanism may be needed to associate link attributes to those of the FA-LSPs.

When the FA link is supported by bidirectional FA LSP(s), a pair of FA link(s) are advertised from each endpoint of the FA. These are usually referred to as symmetrical link(s).

3.2. Link Flooding

Multiple protocols exist that can exchange link state information in the network. For example, when advertising TE link(s) and their attribute(s) using OSPF and ISIS protocols, the respective extensions are defined in [RFC3630] and [RFC5305]. Also, when exchanging such information in BGP protocol, extensions for BGP link state are defined in [RFC7752] and [RFC8571]. The same protocol encodings can be used to advertise FA(s) as TE link(s). As a result, the FA TE link(s) and other normal TE link(s) will appear in the TE link state database of any LSR in the network, and can be used for computing end-to-end TE path(s).

When IGP protocols are used to advertise link state information about FA links, the FA link(s) can appear in both the TE topology, as well as the IGP topology. The use of FA link in the IGP topology may result in undesirable routing loops. A router SHOULD leverage exisitng mechanisms to exclude the FA link from the IGP Shortest Path First (SPF) computations, and to restrict its use within the TE topology for traffic engineered paths computation.

For example, when using ISIS to carry FA link state information, [RFC5305] section 3 describes a way to restrict the link to the TE topology by setting the IGP link metric to maximum (2^24 - 1). Alternatively, when using OSPF, the FA link(s) can be advertised using TE Opaque LSA(s) only, and hence, strictly show up in the TE topology as described in [RFC3630] .

3.3. Underlay LSP(s)

The LSR that hosts an FA link can setup the underlying LSP(s) using different technologies - e.g. RSVP-TE, LDP, and SR.

The FA link can be supported by one or more underlay LSP(s) that terminate on the same remote endpoint. The underlay path(s) can be setup using different signaling technologies, e.g. using RSVP-TE, LDP, SR, etc. When multiple LSP(s) support the same FA link, the attributes of the FA link can be derived from the aggregate properties of each of the underlying LSP(s).

3.4. State Changes

The state of an FA TE link reflects the state of the underlying LSP path that supports it. The TE link is assumed operational and is advertised as long as the underlying LSP path is valid. When all underlying LSP paths are invalidated, the FA TE link advertisement is withdrawn.

3.5. TE Parameters

The TE metrics and TE attributes are used by path computation algorithms to select the TE link(s) that a TE path traverses. When advertising an FA link in OSPF or ISIS, or BGP-LS, the following TE parameters are defined:

TE Path metrics: the FA link advertisement can include information about TE, IGP, and other performance metrics (e.g. delay, and loss). The FA link TE metrics, in this case, can be derived from the underlying path(s) that support the FA link by producing the path accumulative metrics. When multiple LSP(s) support the same FA link, then the higher accumulative metric amongst the LSP(s) is inherited by the FA link.

Resource Class/Color: An FA link can be assigned (e.g. via configuration) a specific set of admin-groups. Alternatively, in some cases, this can be derived from the underlying path affinity - for example, the underlying path strictly includes a specific admin-group.

SRLGs: An FA advertisement could contain the information about the Shared Risk Link Groups (SRLG) for the path taken by the FA LSP associated with that FA. This information may be used for path calculation by other LSRs. The information carried is the union of the SRLGs of the underlying TE links that make up the FA LSP path. It is possible that the underlying path information might change over time, via configuration updates or dynamic route

modifications, resulting in the change of the union of SRLGs for the FA link. If multiple LSP(s) support the same FA link, then it is expected all LSP(s) have the same SRLG union - note, that the exact paths need not be the same.

It is worth noting, that topology changes in the network may affect the FA link underlying LSP path(s), and hence, can dynamically change the TE metrics and TE attributes of the FA links.

3.6. Link Local and Remote Identifiers

It is possible for the FA link to be numbered or unnumbered. [RFC4206] describes a procedure for identifying a numbered FA TE link using IPv4 addresses.

For unnumbered FA link(s), the assignment and handling of the local and remote link identifiers is specified in [RFC3477]. The LSR at each end of the unnumbered FA link assigns an identifier to that link. This identifier is a non-zero 32-bit number that is unique within the scope of the LSR that assigns it. There is no a priori relationship between the identifiers assigned to a link by the LSRs at each end of that link.

The FA link is a unidirectional and point-to-point link. Hence, the combination of link local identifier and advertising node can uniquely identify the link in the TED. In some cases, however, it is desirable to associate the forward and reverse FA links in the TED. In this case, the combination of link local and remote identifier can identify the pair of forward and reverse FA link(s). The LSRs at the two end points of an unnumbered link can exchange with each other the identifiers they assign to the link. Exchanging the identifiers may be accomplished by configuration, or by means of protocol extensions. For example, when the FA link is established over RSVP-TE FA LSP(s), then RSVP extensions have been introduced to exchange the FA link identifier in [RFC3477]. Other protocol extensions pertaining to specific link state protocols, and LSP setup technologies will be discussed in a separate document.

If the link remote identifier is unknown, the value advertised is set to 0 [RFC5307].

4. Segment-Routing over FA Links

The Segment Routing (SR) architecture [RFC4206] describes that an IGP adjacency can be formed over a FA link - in which the remote node of an IGP adjacency is a non-adjacent IGP neighbor.

In Segment-Routing (SR), the adjacency that is established over a link can be assigned an SR Segment [RFC8402]. For example, the Adj-SID allows to strictly steer traffic on to the specific adjacency that is associated with the Adj-SID.

4.1. SR IGP Segments for FA

Extensions have been defined to ISIS [RFC8667] and OSPF [RFC8665] in order to advertise the the Adjacency-SID associated with a specific IGP adjacency. The same extensions apply to adjacencies over FA link. A node can bind an Adj-SID to an FA data-link. The Adj-SID dictates the forwarding of packets through the specific FA link or FA link(s) identified by the Adj-SID, regardless of its IGP/SPF cost.

When the FA link Adj-SID is supported by a single underlying LSP that is associated with a binding label or SID, the same binding label can be used for the FA link Adj-SID. For example, if the FA link is supported by an SR Policy that is assigned a Binding SID B, the Adj-SID of the FA link can be assigned the same Binding SID B.

When the FA link Adj-SID is supported by multiple underlying LSP(s) or SR Policies - each having its own Binding label or SID, an independent FA link Adj-SID is allocated and bound to the multiple underlying LSP(s).

4.1.1. Parallel Adjacencies

Adj-SIDs can also be used in order to represent a set of parallel FA link(s) between two endpoints.

When parallel FA links are associated with the same Adj-SID, a "weight" factor can be assigned to each link and advertised with the Adj-SID advertised with each FA link. The weight informs the ingress (or an SDN/orchestration system) about the load-balancing factor over the parallel adjacencies.

4.2. SR BGP Segments for FA

BGP segments are allocated and distributed by BGP. The SR architecture [RFC8402] defines three types of BGP segments for Egress Peer Engineering (EPE): PeerNode SID, PeerAdj SID, and PeerSet SID.

The applicability of each of the three types to FA links is discussed below:

o PeerNode SID: a BGP PeerNode segment/SID is a local segment. At the BGP node advertising, the forwarding semantics are:

* SR operation: NEXT.

* Next-Hop: forward over any FA link associated with the segment that terminates on remote endpoint.

o PeerAdj SID: a BGP PeerAdj segment/SID is a local segment. At the BGP node advertising it, the forwarding semantics are:

* SR operation: NEXT.

* Next-Hop: forward over the specific FA link to the remote endpoint to which the segment is related.

o PeerSet SID: a BGP PeerSet segment/SID is a local segment. At the BGP node advertising it, the semantics are:

* SR operation: NEXT.

* Next-Hop: load-balance across any of the FA links to any remote endpoint in the related set. The group definition is a policy set by the operator.

4.3. Applicability to Interdomain

In order to determine the potential to establish a TE path through a series of interconnected domains or multi-domain network, it is necessary to have available a certain amount of TE information about each network domain. This need not be the full set of TE information available within each network but does need to express the potential of providing such TE connectivity.

Topology abstraction is described in [RFC7926]. Abstraction allows applying a policy to the available TE information within a domain so to produce selective information that represents the potential ability to connect across the domain. Thus, abstraction does not necessarily offer all possible connectivity options, but presents a general view of potential connectivity according to the policies that determine how the domain's administrator wants to allow the domain resources to be used.

Hence, the domain may be constructed as a mesh of border node to border node TE FA links. When computing a path for an LSP that crosses the domain, a computation point can see which domain entry points can be connected to which others, and with what TE attributes.

5. IANA Considerations

This document has no IANA actions.

6. Security Considerations

TBD.

7. Acknowledgement

The authors would like to thank Peter Psenak for reviewing and providing valuable feedback on this document.

8. Normative References

[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>.

[RFC3477] Kompella, K. and Y. Rekhter, "Signalling Unnumbered Links in Resource ReSerVation Protocol - Traffic Engineering (RSVP-TE)", RFC 3477, DOI 10.17487/RFC3477, January 2003, <https://www.rfc-editor.org/info/rfc3477>.

[RFC3630] Katz, D., Kompella, K., and D. Yeung, "Traffic Engineering (TE) Extensions to OSPF Version 2", RFC 3630, DOI 10.17487/RFC3630, September 2003, <https://www.rfc-editor.org/info/rfc3630>.

[RFC4206] Kompella, K. and Y. Rekhter, "Label Switched Paths (LSP) Hierarchy with Generalized Multi-Protocol Label Switching (GMPLS) Traffic Engineering (TE)", RFC 4206, DOI 10.17487/RFC4206, October 2005, <https://www.rfc-editor.org/info/rfc4206>.

[RFC5305] Li, T. and H. Smit, "IS-IS Extensions for Traffic Engineering", RFC 5305, DOI 10.17487/RFC5305, October 2008, <https://www.rfc-editor.org/info/rfc5305>.

[RFC5307] Kompella, K., Ed. and Y. Rekhter, Ed., "IS-IS Extensions in Support of Generalized Multi-Protocol Label Switching (GMPLS)", RFC 5307, DOI 10.17487/RFC5307, October 2008, <https://www.rfc-editor.org/info/rfc5307>.

[RFC7752] Gredler, H., Ed., Medved, J., Previdi, S., Farrel, A., and S. Ray, "North-Bound Distribution of Link-State and Traffic Engineering (TE) Information Using BGP", RFC 7752, DOI 10.17487/RFC7752, March 2016, <https://www.rfc-editor.org/info/rfc7752>.

[RFC7926] Farrel, A., Ed., Drake, J., Bitar, N., Swallow, G., Ceccarelli, D., and X. Zhang, "Problem Statement and Architecture for Information Exchange between Interconnected Traffic-Engineered Networks", BCP 206, RFC 7926, DOI 10.17487/RFC7926, July 2016, <https://www.rfc-editor.org/info/rfc7926>.

[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>.

[RFC8402] Filsfils, C., Ed., Previdi, S., Ed., Ginsberg, L., Decraene, B., Litkowski, S., and R. Shakir, "Segment Routing Architecture", RFC 8402, DOI 10.17487/RFC8402, July 2018, <https://www.rfc-editor.org/info/rfc8402>.

[RFC8571] Ginsberg, L., Ed., Previdi, S., Wu, Q., Tantsura, J., and C. Filsfils, "BGP - Link State (BGP-LS) Advertisement of IGP Traffic Engineering Performance Metric Extensions", RFC 8571, DOI 10.17487/RFC8571, March 2019, <https://www.rfc-editor.org/info/rfc8571>.

[RFC8665] Psenak, P., Ed., Previdi, S., Ed., Filsfils, C., Gredler, H., Shakir, R., Henderickx, W., and J. Tantsura, "OSPF Extensions for Segment Routing", RFC 8665, DOI 10.17487/RFC8665, December 2019, <https://www.rfc-editor.org/info/rfc8665>.

[RFC8667] Previdi, S., Ed., Ginsberg, L., Ed., Filsfils, C., Bashandy, A., Gredler, H., and B. Decraene, "IS-IS Extensions for Segment Routing", RFC 8667, DOI 10.17487/RFC8667, December 2019, <https://www.rfc-editor.org/info/rfc8667>.

Authors' Addresses

Tarek Saad
Juniper Networks, Inc.

Email: tsaad@juniper.net

Vishnu Pavan Beeram
Juniper Networks, Inc.

Email: vbeeram@juniper.net

Colby Barth
Juniper Networks, Inc.

Email: cbarth@juniper.net

Siva Sivabalan
Ciena Corporation.

Email: ssivabal@ciena.com