LSR Working Group
Internet-Draft
Intended status: Standards Track
Expires: August 18, 2019

U. Chunduri
R. Li
Huawei USA
R. White
Juniper Networks
J. Tantsura
Apstra Inc.
L. Contreras
Telefonica
Y. Qu
Huawei USA

February 14, 2019

Preferred Path Routing (PPR) in IS-IS
draft-chunduri-lsr-isis-preferred-path-routing-02

Abstract

This document specifies a Preferred Path Routing (PPR), a routing protocol mechanism to simplify the path description of data plane traffic in Segment Routing (SR) deployments. PPR aims to mitigate the MTU and data plane processing issues that may result from SR packet overheads; and also supports traffic measurement, accounting statistics and further attribute extensions along the paths. Preferred Path Routing is achieved through the addition of descriptions to IS-IS advertised prefixes, and mapping those to a PPR data-plane identifier.

Requirements Language

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 [RFC2119], RFC8174 [RFC8174] when, and only when they appear in all capitals, as shown here.

Status of This Memo

Copyright Notice

Table of Contents

1. Introduction

In a network implementing Segment Routing (SR), packets are steered through the network using Segment Identifiers (SIDs) carried in the packet header. Each SID uniquely identifies a segment as defined in [I-D.ietf-spring-segment-routing]. SR capabilities are defined for MPLS and IPv6 data planes called SR-MPLS and SRv6 respectively.

In SR-MPLS, each segment is encoded as a label, and an ordered list of segments are encoded as a stack of labels. This stack of labels is carried as part of the packet header. In SRv6, a segment is encoded as an IPv6 address, within a new type of IPv6 hop-by-hop routing header/extension header (EH) called SRH [I-D.ietf-6man-segment-routing-header]; an ordered list of IPv6 addresses/segments are encoded in SRH.

Section 1.2 and Section 1.3 describe performance, hardware capabilities and various associated issues which may result in SR deployments. These motivate the proposed solution, Preferred Path Routing, which is specified in Section 2.

1.1. Acronyms

EL - Entropy Label

ELI - Entropy Label Indicator

LSP - IS-IS Link State PDU

MPLS - Multi Protocol Label Switching

MSD - Maximum SID Depth

MTU - Maximum Transferrable Unit

PPR - Preferred Path Routing/Route

PPR-ID - Preferred Path Route Identifier, a data plane identifier

SID - Segment Identifier

SPF - Shortest Path First

SR-MPLS - Segment Routing with MPLS data plane

SRH   - Segment Routing Header - IPv6 routing Extension headr

SRv6   - Segment Routing with Ipv6 data plane with SRH

TE   - Traffic Engineering

1.2. Challenges with Increased SID Depth

SR label stacks carried in the packet header create challenges in the design and deployment of networks and networking equipment. Following examples illustrates the need for increased SID depth in various use cases:

(a). Consider the following network where SR-MPLS data plane is in use and with same SRGB (5000-6000) on all nodes i.e., A1 to A7 and B1 to B7 for illustration:

```
SID:10   SID:20   SID:30   SID:40  SID:50 SID:300(Ax)  SID:60    SID:70
A1--------A2-------A3-------A4-------A5===============A6-- ----------A7
|          \               /  \5     5/ \   SID:310(Ay) \            /
| 5         \ 10        10/    +-A10-+   \               \10      10/
|            \  SID:80   /     |SID:100   \               \        /
A11 SID:111   \A8-----A9/      |           \  40           \      /
|             /  SID:90 \      +-----+      +---+           \    /
| 5          /10         \10     5    \          \           \  /
|           /SID:125(B2x) \            \          \           \/
B1-------B2==============B3----B4------B5-------=B6----------B7
            SID:127(B2y)
SID:110 SID:120      SID:130 SID:140   SID:150  SID:160  SID:170

  === = Path with Parallel Adjecencies and ADJ-SIDs
  --- = Shortest Path Nodal SID
```

Figure 1: SR-MPLS Network

Global ADJ-SIDs are provisioned between A5-A6 and B2-B3 (with parallel adjecencies). All other SIDs shown are nodal SID indices.

All metrics of the links are set to 1, unless marked otherwise.

Shortest Path from A1 to A7: A2-A3-A4-A5-A6-A7

Path-x: From A1 to A7 - A2-A8-B2-B2x-A9-A10-Ax-A7; Pushed Label Stack @A1: 5020:5080:5120:5125:5090:5100:5300:5070 (where B2x is a local ADJ-SID and Ax is a global ADJ-SID).

In this example, the traffic engineered path is represented with a combination of Adjacency and Node SIDs with a stack of 8 labels. However, this value can be larger, if the use of entropy label [RFC6790] is desired and based on the Readable Label Depth (Section 1.3) capabilities of each node and additional labels required to insert ELI/EL at appropriate places.

Though above network is shown with SR-MPLS data plane, if the network were to use SR-IPv6 data plane, path size would be increased even more because of the size of the IPv6 SID (16 bytes) in SRH.

(b). Apart from the TE case above, when deploying [I-D.ietf-mpls-sfc] or [I-D.xuclad-spring-sr-service-chaining], with the inclusion of services, or non-topological segments on the label stack, can also make the size of the stack much larger.

(c). Some SR-MPLS deployments need accounting statistics for path monitoring and traffic re-optimizations. [I-D.hegde-spring-traffic-accounting-for-sr-paths] and [I-D.cheng-spring-mpls-path-segment] propose solutions with various forms of path segments (either with special labels or PATH segment encoded at the bottom of the stack respectively). However, these proposals further increases the depth of SID stack, when it is compounded with MSD/RLDs of various nodes in the path.

Overall the additional path overhead in various SR deployments may cause the following issues:

a. HW Capabilities: Not all nodes in the path can support the ability to push or read label stack (with additional non-topological and special labels) needed [I-D.ietf-isis-segment-routing-msd] to satisfy user/operator requirements. Alternate paths, which meet these user/operator requirements may not be available.

b. Line Rate: Potential performance issues in deployments, which use SRH data plane with the increased size of the SRH with 16 byte SIDs.

c. MTU: Larger SID stacks on the data packet can cause potential MTU/fragmentation issues (SRH).

d. Header Tax: Some deployments, such as 5G, require minimal packet overhead in order to conserve network resources. Carrying 40 or 50 octets of data in a packet with hundreds of octet of header would be an unacceptable use of available bandwidth (SRH).

With the solution proposed in this document (Section 5) and Section 4), for Path-x in the example network Figure 1 above, SID stack would be reduced from 8 SIDs to a single SID.

1.3. Mitigation with MSD

The number of SIDs in the stack a node can impose is referred as Maximum SID Depth (MSD) capability [I-D.ietf-isis-segment-routing-msd], which must be taken into consideration when computing a path to transport a data packet in a network implementing segment routing. [I-D.ietf-isis-mpls-elc] defines another MSD type, Readable Label Depth (RLD) that is used by a head-end to insert Entropy Label pair (ELI/EL) at appropriate depth, so it could be read by transit nodes. There are situations where the source routed path can be excessive as path represented by SR SIDs need to describe all the nodes and ELI/EL based on the readability of the nodes in that path. [I-D.ietf-isis-segment-routing-msd] defines one registry element applicable for MPLS data plane and this registry can be used for IPv6 data plane with SRH.

MSDs (and RLD type) capabilities advertisement help mitigate the problem for a central entity to create the right source routed path per application/operator requirements. However the availability of actual paths meeting these requirements are still limited by the underlying hardware and their MSD capabilities in the data path.

2. Preferred Path Routing (PPR)

PPR mitigates the issues described in Section 1.2, while continuing to allow the direction of traffic along an engineered path through the network by replacing the label stack with a PPR-ID. The PPR-ID can either be a single label or a native destination address. To facilitate the use of a single label to describe an entire path, a new TLV is added to IS-IS, as described below in Section 3.

A PPR could be an SR path, a traffic engineered path computed based on some constraints, an explicitly provisioned Fast Re-Route (FRR) path or a service chained path. A PPR can be signaled by any node, computed by a central controller, or manually configured by an operator. PPR extends the source routing and path steering capabilities to native IP (IPv4 and IPv6) data planes without hardware upgrades; see Section 5.

2.1. PPR-ID and PPR Path Description

The PPR-ID describes a path through the network. For SR- MPLS this is an MPLS Label/SID and for SRv6 this is an IPv6-SID. For native IP data planes this is either IPv4 or IPv6 address/prefix.

The path identified by the PPR-ID is described as a set of Path Description Elements (PDEs), each of which represents a segment of the path. Each node determines its location in the path as described, and forwards to the next segment/hop or label of the path description (see the Forwarding Procedure Example later in this document).

These PPR-PDEs as defined in Section 3.3, like SR SIDs, can represent topological elements like links/nodes, backup nodes, as well as non-topological elements such as a service, function, or context on a particular node. PPR-PDE optionally, can also have more information as described with in their Sub-TLVs.

A PPR path can be described as a Strict-PPR or a Loose-PPR. In a Strict-PPR all nodes/links on the path are described with SR SIDs for SR data planes or IPv4/IPV6 addresses for native IP data planes. In a Loose-PPR only some of the nodes/links from source to destination are described. More specifics and restrictions around Strict/Loose PPRs are described in respective data planes in Section 5. Each PDE is described as either an MPLS label towards the next hop in MPLS enabled networks, or as an IP next hop, in the case of either "plain"/"native" IP or SRv6 enabled networks. A PPR path is related to a set of PDEs using the following TLVs.

3. PPR Related TLVs

This section describes the encoding of PPR TLV. This TLV can be seen as having 4 logical sections viz., encoding of the PPR-Prefix (IS-IS Prefix), encoding of PPR-ID, encoding of path description with an ordered PDE Sub-TLVs and a set of optional PPR attribute Sub-TLVs, which can be used to describe one or more parameters of the path. Multiple instances of this TLV MAY be advertised in IS-IS LSPs with different PPR-ID Type and with corresponding PDE Sub-TLVS. The PPR TLV has Type TBD (suggested value xxx), and has the following 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
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|      Type     |     Length    |  PPR-Flags                    |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|          PPR-Prefix Sub-TLV (variable size)                   |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|          PPR-ID Sub-TLV (variable size)                       |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|          PPR-PDE Sub-TLVs (variable)                          |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|          PPR-Attribute Sub-TLVs (variable)                    |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 2: PPR TLV Format

- o Type: TBD (IANA) from IS-IS top level TLV registry.
- o Length: Total length of the value field in bytes.
- o PPR-Flags: 2 Octet bit-field of flags for this TLV; described below.
- o PPR-Prefix: A variable size sub-TLV representing the destination of the path being described. This is defined in Section 3.1.
- o PPR-ID: A variable size Sub-TLV representing the data plane or forwarding identifier of the PPR. Defined in Section 3.2.
- o PPR-PDEs: Variable number of ordered PDE Sub-TLVs which represents the path. This is defined in Section 3.3.
- o PPR-Attributes: Variable number of PPR-Attribute Sub-TLVs which represent the path attributes. These are defined in Section 3.4.

The Flags field has the following flag bits defined:

PPR TLV Flags Format

```
 0 1 2 3 4 5 6 7              15
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|S|D|A|L|Reserved | Fragment-ID |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

1. S: If set, the PPR TLV MUST be flooded across the entire routing domain. If the S flag is not set, the PPR TLV MUST NOT be leaked

between IS-IS levels. This bit MUST NOT be altered during the TLV leaking

2. D: When the PPR TLV is leaked from IS-IS level-2 to level-1, the D bit MUST be set. Otherwise, this bit MUST be clear. PPR TLVs with the D bit set MUST NOT be leaked from level-1 to level-2. This is to prevent TLV looping across levels.

3. A: The originator of the PPR TLV MUST set the A bit in order to signal that the prefixes and PPR-IDs advertised in the PPR TLV are directly connected to the originators. If this bit is not set, this allows any other node in the network advertise this TLV on behalf of the originating node of the PPR-Prefix. If PPR TLV is leaked to other areas/levels the A-flag MUST be cleared. In case if the originating node of the prefix must be disambiguated for any reason including, if it is a Multi Homed Prefix (MHP) or leaked to a different IS-IS level or because [RFC7794] X-Flag is set, then PPR-Attribute Sub-TLV Source Router ID SHOULD be included.

4. L: L bit MUST be set if a path has only one fragment or if it is the last Fragment of the path. PPR-ID value for all fragments of the same path MUST be same.

5. Reserved: For future use; MUST be set to 0 on transmit and ignored on receive.

6. Fragment-ID: This is a 7-bit Identifier value (0-127) of the fragment. If fragments are not needed to represent the complete path, L bit MUST be set and this value MUST be set to 0.

The following sub-TLVs draw from a new registry for sub-TLV numbers; this registry is to be created by IANA, and administered using the first come first serve process.

3.1. PPR-Prefix Sub-TLV

The structure of PPR-Prefix 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
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
| Type          |  Length       | MT-ID                         |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
| Prefix Length |  Mask Length  |  IS-IS Prefix                 |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
//           IS-IS Prefix (continued, variable)                //
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|            PPR-Prefix  Sub-TLVs (variable)                    |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 3: PPR-Prefix Sub-TLV Format

- Type: TBD (IANA to assign from sub-TLV registry described above).

- Length: Total length of the value field in bytes.

- MT-ID: The multi-topology identifier defined in [RFC5120]; the 4 most significant bits MUST be set to 0 on transmit and ignored on receive. The remaining 12-bit field contains the MT-ID.

- Prefix Length: The length of the prefix in bytes. For IPv4 it MUST be 4 and IPv6 it MUST be 16 bytes.

- Mask Length: The length of the prefix in bits. Only the most significant octets of the Prefix are encoded.

- IS-IS Prefix: The IS-IS prefix at the tail-end of the advertised PPR. This corresponds to a routable prefix of the originating node and it MAY have one of the [RFC7794] flags set (X-Flag/R-Flag/N-Flag). Value of this field MUST be 4 octets for IPv4 Prefix and MUST be 16 octets for IPv6 Prefix. Encoding is similar to TLV 135 [RFC5305] and TLV 236 [RFC5308] or MT-Capable [RFC5120] IPv4 (TLV 235) and IPv6 Prefixes (TLV 237) respectively.

- PPR-Prefix Sub-TLVs - TBD. These have 1 octet type, 1 octet length and value field is defined per the type field.

3.2. PPR-ID Sub-TLV

This is the actual data plane identifier in the packet header and could be of any data plane as defined in PPR-ID Type field. Both PPR-Prefix and PPR-ID MUST belong to a same node in the network.

```
 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
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
| Type          |  Length       |PPR-ID Flags                   |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
| PPR-ID Type   | PPR-ID Length |PPR-ID Mask Len| Algo          |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
//                 PPR-ID (variable size)                      //
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 4: PPR-ID Sub-TLV Format

o Type: TBD (IANA to assign from sub-TLV registry described above).

o Length: Total length of the value field in bytes.

o

   * PPR-ID Flags: 2 Octet field for PPR-ID flags:

o

PPR-ID Flags Format

```
 0 1 2 3 4 5 6 7..             15
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|L|A|Reserved                   |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

o

1. L: If set, the PPR path is a Loose-PPR. If the this flag is unset, then the PPR path is a Strict-PPR. A Strict-PPR lists every single node or adjacency in the path description from source to the destination.

2. A: If set, all non-PPR path nodes in the IS-IS area/domain MUST add a FIB entry for the PPR-ID with NH set to the shortest path NH for the prefix being advertised. The use of this is TBD. By default this flag MUST be unset.

3. Reserved: For future use; MUST be set to 0 on transmit and ignored on receive.

o

* PPR-ID Type: Data plane type of PPR-ID. This is a new registry (TBD IANA - Suggested values as below) for this Sub-TLV and the defined types are as follows:

o

A. Type: 1 MPLS SID/Label

B. Type: 2 Native IPv4 Address/Prefix

C. Type: 3 Native IPv6 Address/Prefix

D. Type: 4 IPv6 SID in SRv6 with SRH

o PPR-ID Length: Length of the PPR-ID field in octets and this depends on the PPR-ID type. See PPR-ID below for the length of this field and other considerations.

o PPR-ID Mask Length: It is applicable for only for PPR-ID Type 2, 3 and 4. For Type 1 this value MUST be set to zero. It contains the length of the PPR-ID Prefix in bits. Only the most significant octets of the Prefix are encoded. This is needed, if PPR-ID followed is an IPv4/IPv6 Prefix instead of 4/16 octet Address respectively.

o Algo: 1 octet value represents the SPF algorithm. Algorithm registry is as defined in [I-D.ietf-isis-segment-routing-extensions].

o PPR-ID: This is the Preferred Path forwarding identifier that would be on the data packet. The value of this field is variable and it depends on the PPR-ID Type - for Type 1, this is and MPLS SID/Label. For Type 2 this is a 4 byte IPv4 address. For Type 3 this is a 16 byte IPv6 address. For Type 2 and Type 3 encoding is similar to "IS-IS Prefix" as specified in Section 3.1. For Type 4, it is a 16 byte IPv6 SID.

For PPR-ID Type 2, 3 or 4, if the PPR-ID Len is set to non-zero value, then the PPR-ID MUST NOT be advertised as a routable prefix in TLV 135, TLV 235, TLV 236 and TLV 237. Also PPR-ID MUST belong to the node where Prefix is advertised from. PPR-ID Len = 0 case is a special case and is discussed in Section 4.1.

3.3. PPR-PDE Sub-TLV

This Sub-TLV represents the PPR Path Description Element (PDE). PPR-PDEs are used to describe the path in the form of set of contiguous and ordered Sub-TLVs, where first Sub-TLV represents (the top of the

stack in MPLS data plane or) first node/segment of the path. These set of ordered Sub-TLVs can have both topological elements and non-topological elements (e.g., service segments).

```
 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
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
| Type          |  Length       | PPR-PDE Type  | Reserved      |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
| PDE-ID Type   |  PDE-ID Len   | PPR-PDE Flags                 |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
//               PDE-ID Value (continued, variable size)       //
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                  PPR-PDE  Sub-TLVs (variable)                 |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 5: PPR-PDE Sub-TLV Format

o Type: TBD (See IANA for suggested value) from IS-IS PPR TLV Section 3 Sub-TLV registry.

o Length: Total length of the value field in bytes.

o PPR-PDE Type: A new registry (TBD IANA) for this Sub-TLV and the defined types are as follows:

a. Type: 1 Topological

b. Type: 2 Non-Topological

o PDE-ID Type: 1 Octet PDE-forwarding IDentifier Type. A new registry (TBD IANA) for this Sub-TLV and the defined types and corresponding PDE-ID Len, PDE-ID Value are as follows:

a. Type 1: SID/Label type as defined in [I-D.ietf-isis-segment-routing-extensions]. PDE-ID Len and PDE-ID Value fields are per Section 2.3 of the referenced document.

b. Type 2: SR-MPLS Prefix SID. PDE-ID Len and PDE-ID Value are same as Type 1.

c. Type 3: SR-MPLS Adjacency SID. PDE-ID Len and PDE-ID Value are same as Type 1.

d. Type 4: IPv4 Address. PDE-ID Len is 4 bytes and PDE-ID Value is 4 bytes IPv4 address encoded similar to IPv4 Prefix described in Section 3.1.

e. Type 5: IPv6 Address. PDE-ID Len is 16 bytes and PDE-ID Value is 16 bytes IPv6 address encoded similar to IPv6 Prefix described in Section 3.1.

f. Type 6: SRv6 Node SID as defined in [I-D.bashandy-isis-srv6-extensions]. PDE-ID Len and PDE-ID Value are as defined in SRv6 SID from the refrenced draft.

g. Type 7: SRv6 Adjacency-SID. PDE-ID Len and PDE-ID Values are similar to SRv6 Node SID above.

o PPR-PDE Flags: 2 Octet bit-field of flags; described below:

PPR-PDE Flags Format

```
 0 1 2 3 4 5 6 7 ..            15
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|L|D|Reserved                   |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

1. L: Loose Bit. Indicates the type of next "Topological PDE-ID" in the path description and overrides the L bit in Section 3.2. If set, the next PDE is Loose. If this flag is unset, the next Topological PDE is Strict Type.

2. D: Destination bit. By default this bit MUST be unset. This bit MUST be set only for PPR-PDE Type is 1 i.e., Topological and this PDE represents the node PPR-Prefix Section 3.1 belongs to, if there is no sub-sub-TLV to override PPR-Prefix and PPR-ID values.

3. Reserved: 1 Octet reserved bits for future use. Reserved bits MUST be reset on transmission and ignored on receive.

o PPR-PDE Sub-TLVs: TBD. These have 1 octet type, 1 octet length and value field is defined per the type field.

3.4. PPR-Attributes Sub-TLV

PPR-Attribute Sub-TLVs describe the attributes of the path. The following sub-TLVs draw from a new registry for sub-TLV numbers; this registry is to be created by IANA, and administered using the first come first serve process:

o Type 1 (Suggested Value - IANA TBD): Packet Traffic accounting Sub-TLV. Length 0 and no value field. Specifies to create a counter to count number of packets forwarded on this PPR-ID on each node in the path description.

- o Type 2 (Suggested Value - IANA TBD): Traffic statistics in Bytes Sub-TLV. Length 0 and no value field. Specifies to create a counter to count number of bytes forwarded on this PPR-ID specified in the network header (e.g. IPv4, IPv6) on each node in the path description.

- o Type 3 (Suggested Value - IANA TBD): PPR-Prefix originating node's IPv4 Router ID Sub-TLV. Length and Value field are as specified in [RFC7794].

- o Type 4 (Suggested Value - IANA TBD): PPR-Prefix originating node's IPv6 Router ID Sub-TLV. Length and Value field are as specified in [RFC7794].

- o Type 5 (Suggested Value - IANA TBD): PPR-Metric Sub-TLV. Length 4 bytes, and Value is metric of this path represented through the PPR-ID. Different nodes can advertise the same PPR-ID for the same Prefix with a different set of PPR-PDE Sub-TLVs and the receiving node MUST consider the lowest metric value (TBD more, on what happens when metric is same for two different set of PPR-PDE Sub-TLVs).

4. PPR Processing Procedure Example

As specified in Section 2, a PPR can be a TE path, locally provisioned by the operator or by a controller. Consider the following IS-IS network to describe the operation of PPR TLV as defined in Section 3:

```
                  1
               _______
              /   1   \
          +---R2-------R3---+
         /    \_______/      \
        /         1           \
     1 /                       \ 1
      /      1__R13__1          \
     /       /       \           \
   R1------R6        R7-----------R4
     \ 2     \__R14__/       2   /\
      \      2       2          /  \
     3 \                       / 3  \1
        \                 4   /      \
         +----R8------R9-----R10------R12
                       \          1   /
                      1 \            / 1
                         +----R11---+
```

Figure 6: IS-IS Network

In the (Figure 6) shown, consider node R1 as an ingress node, or a head-end node, and the node R4 MAY be an egress node or another head-end node. The numbers shown on links between nodes R1-R14 indicate the bi-directional IS-IS metric as provisioned. R1 may be configured to receive TE source routed path information from a central entity (PCE [RFC5440], Netconf [RFC6241] or a Controller) that comprises of PPR information which relates to sources that are attached to R1. It is also possible to have a PPR provisioned locally by the operator for non-TE needs (FRR or for chaining certain services).

The PPR TLV (as specified in Section 3) is encoded as an ordered list of PPR-PDEs from source to a destination node in the network and is represented with a PPR-ID Section 3.2. The PPR TLV includes PPR-PDE Sub-TLVS Section 3.3, which represent both topological and non-topological elements and specifies the actual path towards a PPR-Prefix at R4.

- o The shortest path towards R4 from R1 are through the following sequence of nodes: R1-R2-R3-R4 based on the provisioned metrics.

- o The central entity can define a few PPRs from R1 to R4 that deviate from the shortest path based on other network characteristic requirements as requested by an application or service. For example, the network characteristics or performance requirements may include bandwidth, jitter, latency, throughput, error rate, etc.

o A first PPR may be identified by PPR-ID = 1 (value) and may include the path of R1-R6-R7-R4 for a Prefix advertised by R4. This is an example for a Loose-PPR and 'L' bit MUST be set on Section 3.2.

o A second PPR may be identified by PPR-ID = 2 (value) and may include the path of R1-R8-R9-R10-R4. This is an example for a Strict-PPR and 'L' bit MUST be unset on Section 3.2. Though this example shows PPR with all nodal SIDs, it is possible to have a PPR with combination of node and adjacency SIDs (local or global) or with PPR-PDE Type set to Non-Topological as defined in Section 3.3 elements along with these.

4.1. PPR TLV Processing

The first topological sub-object or PDE (Section 3.3) relative to the beginning of PPR Path contains the information about the first node (e.g. in SR-MPLS it's the topmost label). The last topological sub-object or PDE contains information about the last node (e.g. in SR-MPLS it's the bottommost label).

Each receiving node, determines whether an advertised PPR includes information regarding the receiving node. Before processing any further, validation MUST be done to see if any PPR topological PDE is seen more than once (possible loop), if yes, this PPR TLV MUST be ignored. Processing of PPR TLVs can be done, during the end of the SPF computation (for MTID that is advertised in this TLV) and for the each prefix described through PPR TLV. For example, node R9 receives the PPR information, and ignores the PPR-ID=1 (Section 4) because this PPR TLV does not include node R9 in the path description/ordered PPR-PDE list.

However, node R9 may determine that the second PPR identified by PPR-ID = 2 does include the node R9 in its PDE list. Therefore, node R9 updates the local forwarding database to include an entry for the destination address of R4 indicates, that when a data packet comprising a PPR-ID of 2 is received, forward the data packet to node R10 instead of R11. This is even though from R9 the shortest path to reach R4 via R11 (Cost 3: R9-R11-R12-R4) it chooses the nexthop to R10 to reach R4 as specified in the PPR path description. Same process happens to all nodes or all topological PDEs as described in the PPR TLV.

In summary, the receiving node checks first, if this node is on the path by checking the node's topological elements (with PPR-PDE Type set to Topological) in the path list. If yes, it adds/adjusts the shortest path nexthop computed towards PPR Prefix to the shortest path nexthop towards the next topological PDE in the PPR's Path.

For PPR-ID (Section 3.2) Type 2, 3 or 4, if the PPR-ID Len is set to 0, then Prefix would also become the PPR-ID (a special case). This can be used for some situations, where certain optimizations are required in the network. For this, path described in the PPR TLV SHOULD be completely dis-joint from the shortest path route to the prefix. If the disjoint-ness property is not maintained then the traffic MAY not be using the PPR path, as and when it encounters any node which is not in the path description.

4.2. Path Fragments

A complete PPR path may not fit into maximum allowable size of the IS-IS TLV. To overcome this a 7 bit Fragment-ID field is defined in Section 3 . With this, a single PPR path is represented via one or more fragmented PPR path TLVs, with all having the same PPR-ID. Each fragment carries the PPR-ID as well as a numeric Fragment-ID from 0 to (N-1), when N fragments are needed to describe the PPR Graph (where N>1). In this case Fragment (N-1) MUST set the L bit to indicate it is the last fragment. If Fragment-ID is non zero in the TLV, then it MUST not carry PPR-Prefix Sub-TLV. The optional PPR Attribute Sub-TLVs which describe the path overall MUST be included in the last fragment only (i.e., when the L bit is set).

5. PPR Data Plane aspects

Data plane for PPR-ID is selected by the entity (e.g., a controller, locally provisioned by operator), which selects a particular PPR in the network. Section 3.2 defines various data plane identifier types and a corresponding data plane identifier is selected by the entity which selects the PPR. Other data planes other than described below can also use this TLV to describe the PPR. Further details TBD.

5.1. SR-MPLS with PPR

If PPR-ID Type is 1, then the PPR belongs to SR-MPLS data plane and the complete PPR stack is represented with a unique SR SID/Label and this gets programmed on the data plane of each node, with the appropriate nexthop computed as specified in Section 4. PPR-ID here is a label/index from the SRGB (like another node SID or global ADJ-SID). PPR path description here is a set of ordered SIDs represented with PPR-PDE (Section 3.2) Sub-TLVs. Non-Topological segments also programmed in the forwarding to enable specific function/service, when the data packet hits with corresponding PPR-ID.

Based on 'L' flag in PPR-ID Flags (Section 3.2), for SR-MPLS data plane either 1 label or 2 labels need to be provisioned on individual nodes on the path description. For the example network in Section 4, for PPR-ID=1, which is a loose path, node R6 programs the bottom

label as PPR-ID and the top label as the next topological PPR-PDE in the path, which is a node SID of R7. The NH computed at R6 would be the shortest path towards R7 i.e., the interface towards R13. If 'L' flag is unset only PPR-ID is programmed on the data plane with NH set to the shortest path towards next topological PPR-PDE.

5.2. SRv6 with PPR

If PPR-ID Type is 4, the PPR belongs to SRv6 with SRH data plane and the complete PPR stack is represented with IPv6 SIDs and this gets programmed on the data plane with the appropriate nexthop computed as specified in Section 4. PPR-ID here is a SRv6 SID. PPR path description here is a set of ordered SID TLVs similar to as specified in Section 5.1. One way PPR-ID would be used in this case is by setting the same as the destination IPv6 address and SL field in SRH would be set to 0; however SRH [I-D.ietf-6man-segment-routing-header] can contain any other TLVs and non-topological SIDs as needed.

5.3. PPR Native IP Data Planes

If PPR-ID Type is 2 then source routing and packet steering can be done in IPv4 data plane (PPR-IPv4), along the path as described in PPR Path description. This is achieved by setting the destination IP address as PPR-ID, which is an IPv4 address in the data packet (tunneled/encapsulated). There is no data plane change or upgrade needed to support this. However this is applicable to only Strict-PPR paths ('L' bit as specified in Section 3.2 MUST be unset).

Similarly for PPR-ID Type is 3, then source routing and packet steering can be done in IPv6 data plane (PPR-IPv6), along the path as described in PPR Path description. Whatever specified above for IPv4 applies here too, except that destination IP address of the data packet is IPv6 Address (PPR-ID). This doesn't require any IPv6 extension headers (EH), if there is no metadata/TLVs need to be carried in the data packet.

6. PPR Scaling Considerations

In a network of N nodes O(N^2) total (unidirectional) paths are necessary to establish any-to-any connectivity, and multiple (k) such path sets may be desirable if multiple path policies are to be supported (lowest latency, highest throughput etc.).

In many solutions and topologies, N may be small enough and/or only a small set of paths need to be preferred paths, for example for high value traffic (DetNet, some of the defined 5G slices), and then the technology specified in this document can support these deployments.

However, to address the scale needed when a larger number of PPR paths are required, the PPR TREE structure can be used [I-D.draft-ce-ppr-graph-00]. Each PPR Tree uses one label/SID and defines paths from any set of nodes to one destination, thus reduces the number of entries needed in SRGB at each node (for SR-MPLS data plane Section 5.1). These paths form a tree rooted in the destination. In other word, PPR Tree identifiers are destination identifiers and PPR Treed are path engineered destination routes (like IP routes) and it scaling simplifies to linear in N i.e., O(k*N).

7. PPR Traffic Accounting

Section 3.4 defines few PPR-Attributes to indicate creation of traffic accounting statistics in each node of the PPR path description. Presence of "Packet Traffic Accounting" and "Traffic Statistics" Sub-TLVs instruct to provision the hardware, to account for the respective traffic statistics. Traffic accounting should happen, when the actual data traffic hits for the PPR-ID in the forwarding plane. This allows more granular and dynamic enablement of traffic statistics for only certain PPRs as needed.

This approach, thus is more safe and secure than any mechanism that involves creation of the state in the nodes with the data traffic itself. This is because, creation and deletion of the traffic accounting state for PPRs happen through IS-IS LSP processing and IS-IS protocol control plane security Section 10 options are applicable to this TLV.

How the traffic accounting is distributed to a central entity is out of scope of this document. One can use any method (e.g. gRPC) to extract the PPR-ID traffic stats from various nodes along the path.

8. Acknowledgements

Thanks to Alex Clemm, Lin Han, Toerless Eckert, Stewart Bryant and Kiran Makhijani for initial discussions on this topic. Thanks to Kevin Smith and Stephen Johnson for various deployment scenarios applicability from ETSI WGs perspective. Authors also acknowledge Alexander Vainshtein for detailed discussions and few suggestions on this topic.

Earlier versions of draft-ietf-isis-segment-routing-extensions have a mechanism to advertise EROs through Binding SID.

9. IANA Considerations

This document requests the following new TLV in IANA IS-IS TLV code-point registry.

TLV #	Name
TBD	PPR TLV

This document requests IANA to create a new Sub-TLV registry for PPR TLV Section 3 with the following initial entries (suggested values):

Sub-TLV #	Sub-TLV Name
1	PPR-Prefix (Section 3.1)
2	PPR-ID (Section 3.2)
3	PPR-PDE (Section 3.3)

This document also requests IANA to create a new Sub-TLV registry for PPR Path attributes with the following initial entries (suggested values):

Sub-TLV #	Sub-TLV Name
1	Packet Traffic Accounting (Section 3.4)
2	Traffic Statistics (Section 3.4)
3	PPR-Prefix Source IPv4 Router ID (Section 3.4)
4	PPR-Prefix Source IPv6 Router ID (Section 3.4)
5	PPR-Metric (Section 3.4)

This document requests additional IANA registries in an IANA managed registry "Interior Gateway Protocol (IGP) Parameters" for various PPR TLV parameters. The registration procedure is based on the "Expert Review" as defined in [RFC8126]. The suggested registry names are:

- "PPR-Type" - Types are an unsigned 8 bit numbers. Values are as defined in Section 3 of this document.

- "PPR-Flags" - 1 Octet. Bits as described in Section 3 of this document.

- "PPR-ID Type" - Types are an unsigned 8 bit numbers. Values are as defined in Section 3.2 of this document.

- "PPR-ID Flags" - 1 Octet. Bits as described in Section 3.2 of this document.

- "PPR-PDE Type" - Types are an unsigned 8 bit numbers. Values are as defined in Section 3.3 of this document.

- "PPR-PDE Flags" - 1 Octet. Bits as described in Section 3.3 of this document.

- "PDE-ID Type" - Types are an unsigned 8 bit numbers. Values are as defined in Section 3.3 of this document.

10. Security Considerations

Security concerns for IS-IS are addressed in [RFC5304] and [RFC5310]. Further security analysis for IS-IS protocol is done in [RFC7645] with detailed analysis of various security threats and why [RFC5304] should not be used in the deployments. Advertisement of the additional information defined in this document introduces no new security concerns in IS-IS protocol. However as this extension is related to SR-MPLS and SRH data planes as defined in [I-D.ietf-spring-segment-routing], those particular data plane security considerations does apply here.

11. References

11.1. Normative References

[I-D.ietf-isis-segment-routing-msd]
Tantsura, J., Chunduri, U., Aldrin, S., and L. Ginsberg, "Signaling MSD (Maximum SID Depth) using IS-IS", draft-ietf-isis-segment-routing-msd-19 (work in progress), October 2018.

[I-D.ietf-spring-segment-routing]
Filsfils, C., Previdi, S., Ginsberg, L., Decraene, B., Litkowski, S., and R. Shakir, "Segment Routing Architecture", draft-ietf-spring-segment-routing-15 (work in progress), January 2018.

[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>.

11.2. Informative References

[I-D.bashandy-isis-srv6-extensions]
Psenak, P., Filsfils, C., Bashandy, A., Decraene, B., and Z. Hu, "IS-IS Extensions to Support Routing over IPv6 Dataplane", draft-bashandy-isis-srv6-extensions-04 (work in progress), October 2018.

[I-D.cheng-spring-mpls-path-segment]
Cheng, W., Wang, L., Li, H., Chen, M., Gandhi, R., Zigler, R., and S. Zhan, "Path Segment in MPLS Based Segment Routing Network", draft-cheng-spring-mpls-path-segment-03 (work in progress), October 2018.

[I-D.hegde-spring-traffic-accounting-for-sr-paths]
Hegde, S., "Traffic Accounting for MPLS Segment Routing Paths", draft-hegde-spring-traffic-accounting-for-sr-paths-02 (work in progress), October 2018.

[I-D.ietf-6man-segment-routing-header]
Filsfils, C., Previdi, S., Leddy, J., Matsushima, S., and d. daniel.voyer@bell.ca, "IPv6 Segment Routing Header (SRH)", draft-ietf-6man-segment-routing-header-16 (work in progress), February 2019.

[I-D.ietf-isis-mpls-elc]
Xu, X., Kini, S., Sivabalan, S., Filsfils, C., and S. Litkowski, "Signaling Entropy Label Capability and Entropy Readable Label Depth Using IS-IS", draft-ietf-isis-mpls-elc-06 (work in progress), September 2018.

[I-D.ietf-isis-segment-routing-extensions]
Previdi, S., Ginsberg, L., Filsfils, C., Bashandy, A., Gredler, H., and B. Decraene, "IS-IS Extensions for Segment Routing", draft-ietf-isis-segment-routing-extensions-22 (work in progress), December 2018.

[I-D.ietf-mpls-sfc]
Farrel, A., Bryant, S., and J. Drake, "An MPLS-Based Forwarding Plane for Service Function Chaining", draft-ietf-mpls-sfc-05 (work in progress), February 2019.

[I-D.xuclad-spring-sr-service-chaining]
Clad, F., Xu, X., Filsfils, C., daniel.bernier@bell.ca, d., Li, C., Decraene, B., Ma, S., Yadlapalli, C., Henderickx, W., and S. Salsano, "Segment Routing for Service Chaining", draft-xuclad-spring-sr-service-chaining-01 (work in progress), March 2018.

[RFC5120] Przygienda, T., Shen, N., and N. Sheth, "M-ISIS: Multi Topology (MT) Routing in Intermediate System to Intermediate Systems (IS-ISs)", RFC 5120, DOI 10.17487/RFC5120, February 2008, <https://www.rfc-editor.org/info/rfc5120>.

[RFC5304] Li, T. and R. Atkinson, "IS-IS Cryptographic Authentication", RFC 5304, DOI 10.17487/RFC5304, October 2008, <https://www.rfc-editor.org/info/rfc5304>.

[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>.

[RFC5308] Hopps, C., "Routing IPv6 with IS-IS", RFC 5308, DOI 10.17487/RFC5308, October 2008, <https://www.rfc-editor.org/info/rfc5308>.

[RFC5310] Bhatia, M., Manral, V., Li, T., Atkinson, R., White, R., and M. Fanto, "IS-IS Generic Cryptographic Authentication", RFC 5310, DOI 10.17487/RFC5310, February 2009, <https://www.rfc-editor.org/info/rfc5310>.

[RFC5440] Vasseur, JP., Ed. and JL. Le Roux, Ed., "Path Computation Element (PCE) Communication Protocol (PCEP)", RFC 5440, DOI 10.17487/RFC5440, March 2009, <https://www.rfc-editor.org/info/rfc5440>.

[RFC6241] Enns, R., Ed., Bjorklund, M., Ed., Schoenwaelder, J., Ed., and A. Bierman, Ed., "Network Configuration Protocol (NETCONF)", RFC 6241, DOI 10.17487/RFC6241, June 2011, <https://www.rfc-editor.org/info/rfc6241>.

[RFC6790] Kompella, K., Drake, J., Amante, S., Henderickx, W., and L. Yong, "The Use of Entropy Labels in MPLS Forwarding", RFC 6790, DOI 10.17487/RFC6790, November 2012, <https://www.rfc-editor.org/info/rfc6790>.

[RFC7645] Chunduri, U., Tian, A., and W. Lu, "The Keying and Authentication for Routing Protocol (KARP) IS-IS Security Analysis", RFC 7645, DOI 10.17487/RFC7645, September 2015, <https://www.rfc-editor.org/info/rfc7645>.

[RFC7794] Ginsberg, L., Ed., Decraene, B., Previdi, S., Xu, X., and U. Chunduri, "IS-IS Prefix Attributes for Extended IPv4 and IPv6 Reachability", RFC 7794, DOI 10.17487/RFC7794, March 2016, <https://www.rfc-editor.org/info/rfc7794>.

[RFC8126] Cotton, M., Leiba, B., and T. Narten, "Guidelines for Writing an IANA Considerations Section in RFCs", BCP 26, RFC 8126, DOI 10.17487/RFC8126, June 2017, <https://www.rfc-editor.org/info/rfc8126>.

[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>.

Authors' Addresses

Uma Chunduri
Huawei USA
2330 Central Expressway
Santa Clara, CA 95050
USA

Email: uma.chunduri@huawei.com

Richard Li
Huawei USA
2330 Central Expressway
Santa Clara, CA 95050
USA

Email: renwei.li@huawei.com

Russ White
Juniper Networks
Oak Island, NC 28465
USA

Email: russ@riw.us

Jeff Tantsura
Apstra Inc.
333 Middlefield Road
Menlo Park, CA 94025
USA

Email: jefftant.ietf@gmail.com

Luis M. Contreras
Telefonica
Sur-3 building, 3rd floor
Madrid 28050
Spain

Email: luismiguel.contrerasmurillo@telefonica.com

Yingzhen Qu
Huawei USA
2330 Central Expressway
Santa Clara, CA 95050
USA

Email: yingzhen.qu@huawei.com