Network Working Group   Z. Hu
Internet-Draft   Huawei Technologies
Intended status: Standards Track   H. Chen
Expires: February 3, 2021   Futurewei
                                                              J. Yao
                                                  Huawei Technologies
                                                           C. Bowers
                                                     Juniper Networks
                                                              Y. Zhu
                                                       China Telecom
                                                      August 2, 2020

SR-TE Path Midpoint Protection
draft-hu-spring-segment-routing-proxy-forwarding-11

Abstract

   Segment Routing Traffic Engineering (SR-TE) supports the creation of
   explicit paths using segment lists containing adjacency-SIDs, node-
   SIDs, anycast-SIDs, and binding-SIDs. When the segment list defining
   an SR-TE path contains a node-SID, and the node fails, the network
   may no longer be able to properly forward traffic on that SR-TE path.
   [I-D.ietf-rtgwg-segment-routing-ti-lfa] describes a mechanism that
   allows local repair actions on the direct neighbors of the failed
   node to temporarily route traffic to the node immediately following
   the failed node on the SR-TE path segment list. However, once the
   IGP shortest paths have converged, the local repair mechanism is no
   longer sufficient to continue forwarding traffic using the original
   segment list of the SR-TE path, since the non-neighbors of the failed
   node will no longer have a route to reach the failed node. This
   document describes a mechanism that allows traffic to continue to be
   forwarded on an SR-TE path for an extended period of time after the
   failure of a node used in the path's segment list.

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 RFC 2119 [RFC2119].

Status of This Memo

   This Internet-Draft is submitted in full conformance with the
   provisions of BCP 78 and BCP 79.

   Internet-Drafts are working documents of the Internet Engineering
   Task Force (IETF). Note that other groups may also distribute

working documents as Internet-Drafts. The list of current Internet-
Drafts is at https://datatracker.ietf.org/drafts/current/.

Internet-Drafts are draft documents valid for a maximum of six months
and may be updated, replaced, or obsoleted by other documents at any
time. It is inappropriate to use Internet-Drafts as reference
material or to cite them other than as "work in progress."

This Internet-Draft will expire on February 3, 2021.

Copyright Notice

Table of Contents

1. Introduction

   Segment Routing Traffic Engineering (SR-TE) is a technology that
   implements traffic engineering using Segment Routing. SR-TE supports
   the creation of explicit paths using adjacency-SIDs, node-SIDs,
   anycast-SIDs, and binding-SIDs. A node-SID in the segment list
   defining an SR-TE path indicates a loose hop that the SR-TE path
   should pass through. When a particular node fails, it would be
   useful to be able to continue to send traffic on an SR-TE path that
   uses the node-SID of the failed node for an extended period of time,
   without having to immediately modify the segment list used at the
   ingress to the SR-TE path.

   The first step to achieve this objective is to make the rest of the
   routers in the network continue to forward traffic using the node-SID
   of the failed node. If we don't do anything special, once the IGP
   converges to take into account the failed node, a given router will
   no longer maintain a route corresponding to the node-SID. Any
   traffic that arrives at the router with the node-SID of the failed
   node as the active segment will be dropped. This document addresses
   this problem by having each neighbor of the failed node advertise its
   SR proxy forwarding capability. This indicates that the neighbor
   (the Proxy Forwarder) will forward traffic on behalf of the failed
   node. A router receiving the SR Proxy Forwarding capability from
   neighbors of a failed node will send traffic using the node-SID of
   the failed node to the nearest Proxy Forwarder.

   Once the affected traffic reaches a Proxy Forwarder, the Proxy
   Forwarder sends the traffic on the post-failure shortest path to the
   node immediately following the failed node in the segment list.
   [I-D.ietf-rtgwg-segment-routing-ti-lfa] describes how the immediate
   neighbors of a failed node can accomplish this by forwarding based on
   the first two segments in the segment list. The forwarding described
   in this draft was originally intended to be used for only a short
   period of time, to provide fast-reroute protection until the IGP
   converges. The current document proposes to extend this behavior on
   the Proxy Forwarder until well after the IGP has converged.

   If the faulty node is a label adhesion node, the Binding-SIDs cannot
   be exchanged to the label stack for its identity, and the traffic
   will be lost before it reaches the faulty node.

In this document, the proxy mechanism is provided in the neighbor
node of the faulty node of the forwarding path to implement traffic
forwarding after the node with the label adhesion fails on the SR-TE
loose path.

2. Extensions to IGP for Proxy Forwarding

When a node has segment routing proxy forwarding capability, it
advertises this capability. The capability indicates that the node
has the ability to do proxy forwarding for the global SID of each of
its neighbors. When an neighbor in a SR-TE path fails, the traffic
can be forwarded to the proxy node, which fast re-routes the traffic
to the node immediately following the failed node on the SR-TE path.

2.1. Extensions to OSPF

2.1.1. Advertising Proxy Forwarding

When a node P has the capability to do a SR proxy forwarding for all
its neighboring nodes for protecting the failures of these nodes,
node P advertises its SR proxy forwarding capability in its router
information opaque LSA, which contains a Router Functional
Capabilities TLV of the format as shown in Figure 1.

```
 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            |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                    Functional Capabilities                    |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```
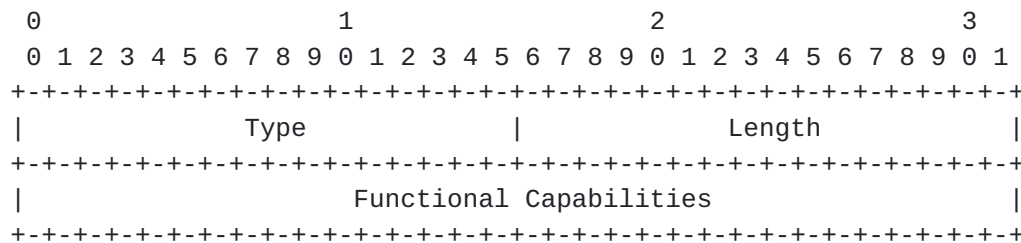
Figure 1: Router Functional Capabilities TLV

One bit (called PF bit) in the Functional Capabilities field of the
TLV is used to indicate node P's SR proxy forwarding capability.
When this bit is set to one by node P, it indicates that node P is
capable of doing a SR proxy forwarding for its neighboring nodes.

For a node X in the network, it learns the prefix/node SID of node N,
which is originated and advertised by node N. It creates a proxy
prefix/node SID of node N for node P if node P is capable of doing SR
proxy forwarding for node N. The proxy prefix/node SID of node N for
node P is a copy of the prefix/node SID of node N originated by node
N, but stored under (or say, associated with) node P.

In normal operations, node X prefers to use the prefix/node SID of
node N. When node N fails, node X prefers to use the proxy prefix/

node SID of node N. Thus node X will forward the traffic targeting
to the prefix/node SID of node N to node P when node N fails, and
node P will do a SR proxy forwarding for node N and forwarding the
traffic to its final destination without going through node N. After
node N fails, node X will keep the FIB entry to the proxy prefix/node
SID of node N for a given period of time.

Note that the behaviors of normal IP forwarding and routing
convergences in a network are not changed at all by the SR proxy
forwarding. For example, the next hop used by BGP is an IP address
(or prefix). The IGP and BGP converge in normal ways for changes in
the network. The packet with its IP destination to this next hop is
forwarded according to the IP forwarding table (FIB) derived from IGP
and BGP routes.

If node P can not do a SR proxy forwarding for all its neighboring
nodes, but for some of them, then it advertises the node SID of each
of the nodes as a proxy node SID, indicating that it is able to do
proxy forwarding for the node SID.

A new TLV, called Proxy Node SIDs TLV, is defined for node P to
advertise the node SIDs of some of its neighboring nodes. It has the
format as shown in Figure 2.

```
  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 (TBD1)          |             Length            |
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
 |                      Node SID Sub-TLVs                        |
 :                                                              :
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```
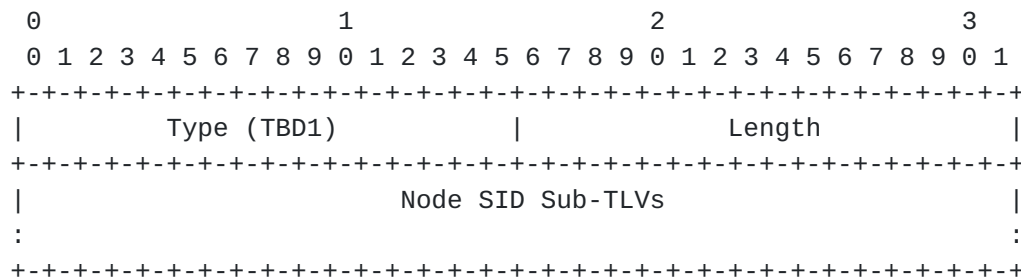
                   Figure 2: OSPF Proxy Node SIDs TLV

The Type (TBD1) is to be assigned by IANA. The TLV contains a number
of Node SID Sub-TLVs. The Length is the total size of the Node SID
Sub-TLVs included in the TLV. A Node SID Sub-TLV is the Prefix SID
Sub-TLV defined in [I-D.ietf-ospf-segment-routing-extensions].

A proxy forwarding node P originates an Extended Prefix Opaque LSA
containing this new TLV. The TLV includes the Node SID Sub-TLVs for
the node SIDs of some of P's neighboring nodes. For each of some of
P's neighboring nodes, the Node SID Sub-TLV for its prefix/node SID
is included the TLV. This prefix/node SID is called a proxy prefix/
node SID.

A proxy forwarding node will originate an Extended Prefix Opaque LSA, which includes a Proxy Node SIDs TLV. The format of the LSA is shown in Figure 3.

For a proxy forwarding node P, having a number of neighboring nodes, P originates and maintains an Extended Prefix Opaque LSA, which includes a Proxy Node SIDs TLV. The TLV contains the Prefix/Node SID Sub-TLV for each of some of the neighboring nodes after node P creates the corresponding proxy forwarding entries for protecting the failure of some of the neighboring nodes.

```
 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
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|            LS age              |    Options    |   LS Type     |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
| Opaque Type(7)|                Opaque ID                      |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                     Advertising Router                        |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                     LS sequence number                        |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|          LS checksum          |              Length           |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                                                               |
:                             TLVs                              :
:                  (including Proxy Node SIDs TLV)              :
|                                                               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```
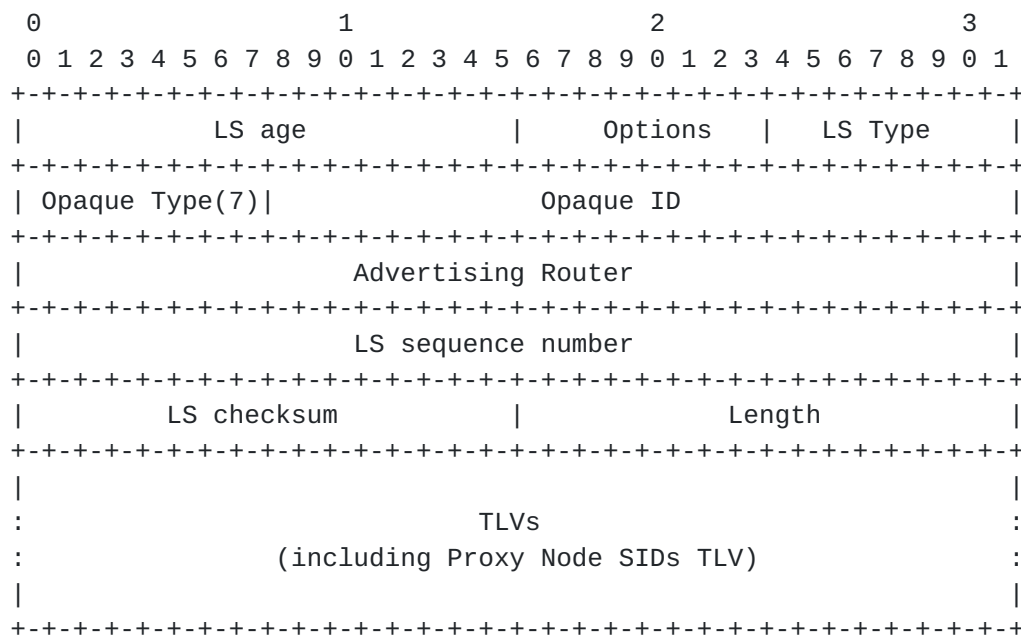
Figure 3: OSPFv2 Extended Prefix Opaque LSA

When an neighboring node fails, P maintains the LSA with the TLV containing the Prefix/Node SID Sub-TLV for the neighboring node for a given period of time. After the given period of time, the Prefix/ Node SID Sub-TLV for the neighboring node is removed from the TLV in the LSA and then after a given time the corresponding proxy forwarding entries for protecting the failure of the neighboring node is removed.

For a node X in the network, it learns the prefix/node SID of node N and the proxy prefix/node SID of node N. The former is originated and advertised by node N, and the latter is originated and advertised by the proxy forwarding node P of node N. Note that the proxy Prefix/Node SID Sub-TLV for node N does not contain a prefix of node N, and the prefix is the prefix associated with the prefix/node SID of node N originated by node N.

In normal operations, node X prefers to use the prefix/node SID of
node N. When node N fails, node X prefers to use the proxy prefix/
node SID of node N. Thus node X will forward the traffic targeting
to node N to node P when node N fails, and node P will do a proxy
forwarding for node N and forwarding the traffic to its destination
without going through node N.

2.1.2. Advertising Binding Segment

For a binding segment (or binding for short) on a node A, which
consists of a binding SID and a list of segments, node A advertises
an LSA containing the binding (i.e., the binding SID and the list of
the segments). The LSA is advertised only to each of the node A's
neighboring nodes. For OSPFv2, the LSA is a opaque LSA of LS type 9
(i.e., a link local scope LSA).

A binding segment is represented by binding segment TLV of the format
as shown in Figure 4.

```
 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 (TBD2)         |             Length            |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|            Reserved           |BindingSID Type|   SIDs Type   |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
~                   Binding SID Sub-TLV/value                   ~
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
~                     SID Sub-TLVs/values                       ~
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```
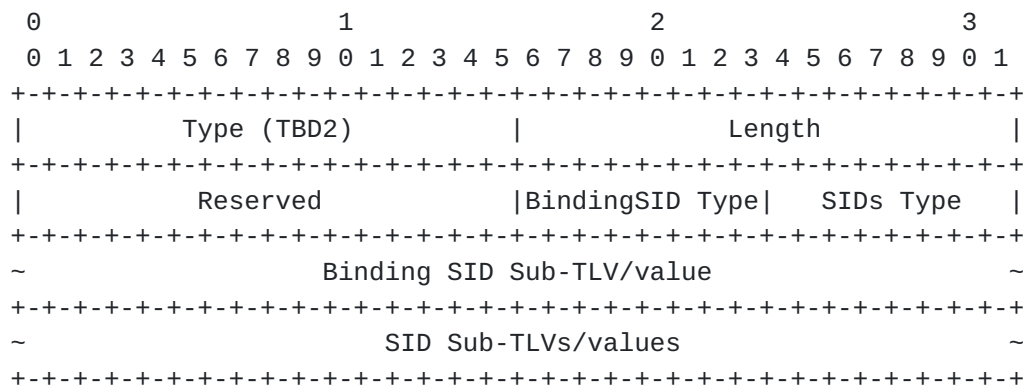
Figure 4: OSPF Binding Segment TLV

It comprises a binding SID and a list of segments (SIDs). The fields
of this TLV are defined as follows:

Type: 2 octets, its value (TBD2) is to be assigned by IANA.

Length: 2 octets, its value is (4 + length of Sub-TLVs/values).

Binding SID Type (BT): 1 octet indicates whether the binding SID is
represented by a Sub-TLV or a value included in the TLV. For the
binding SID represented by a value, it indicates the type of binding
SID. The following BT values are defined:

o BT = 0: The binding SID is represented by a Sub-TLV (i.e., Binding
SID Sub-TLV) in the TLV. A binding SID Sub-TLV is a SID/Label Sub-

TLV defined in [I-D.ietf-ospf-segment-routing-extensions]. BT != 0
indicates that the binding SID is represented by a value.

o BT = 1: The binding SID value is a label, which is represented by
the 20 rightmost bits. The length of the value is 3 octets.

o BT = 2: The binding SID value is a 32-bit SID. The length of the
value is 4 octets.

SIDs Type (ST): 1 octet indicates whether the list of segments (SIDs)
are represented by Sub-TLVs or values included in the TLV. For the
SIDs represented by values, it indicates the type of SIDs. The
following ST values are defined:

o ST = 0: The SIDs are represented by Sub-TLVs (i.e., SID Sub-TLVs)
in the TLV. A SID Sub-TLV is an Adj-SID Sub-TLV, a Prefix-SID Sub-
TLV or a SID/Label Sub-TLV defined in
[I-D.ietf-ospf-segment-routing-extensions]. ST != 0 indicates that
the SIDs are represented by values.

o ST = 1: Each of the SID values is a label, which is represented by
the 20 rightmost bits. The length of the value is 3 octets.

o ST = 2: Each of the SID values is a 32-bit SID. The length of the
value is 4 octets.

The opaque LSA of LS Type 9 containing the binding segment (i.e., the
binding SID and the list of the segments) has the format as shown in
Figure 5. It may have Opaque Type of x (the exact type is to be
assigned by IANA) for Binding Segment Opaque LSA.

```
     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
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |             LS age             |     Options   | LS Type (9)  |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    | Opaque Type(x)|                 Opaque ID                     |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |                     Advertising Router                        |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |                     LS sequence number                        |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |        LS checksum             |              Length          |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |                                                               |
    :                    Binding Segment TLVs                       :
    |                                                               |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```
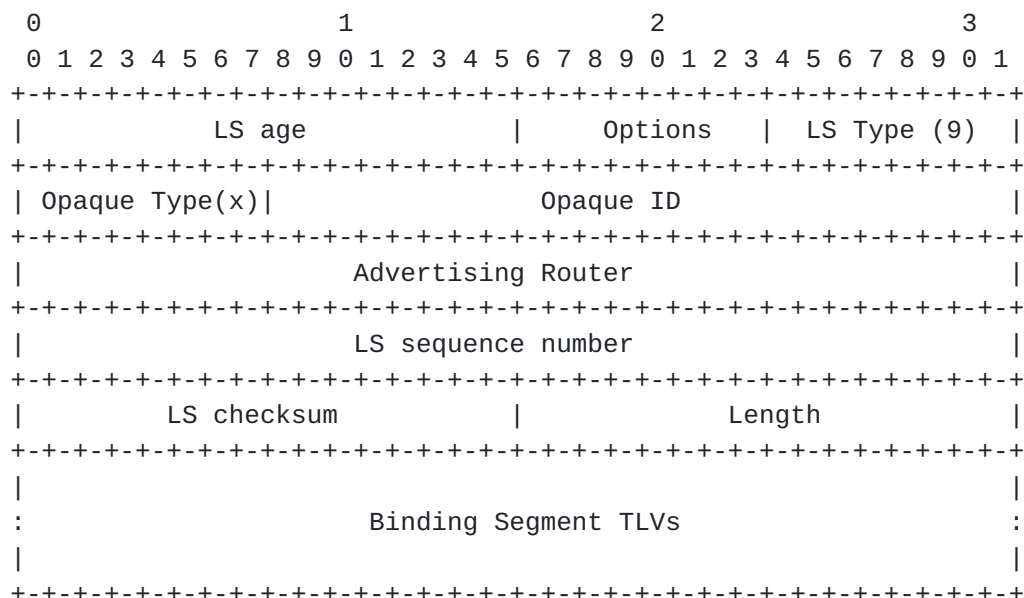
Figure 5: OSPFv2 Binding Segment Opaque LSA

For every binding on a node A, the LSA originated by A contains a binding segment TLV for it.

For node A running OSPFv3, it originates a link-local scoping LSA of a new LSA function code (TBD3) containing binding segment TLVs for the bindings on it. The format of the LSA is illustrated in Figure 6.

```
     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
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |             LS age             |0|0|0|       BS-LSA (TBD3)    |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |                       Link State ID                           |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |                     Advertising Router                        |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |                     LS Sequence Number                        |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |        LS checksum             |              Length          |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |                                                               |
    :                    Binding Segment TLVs                       :
    |                                                               |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```
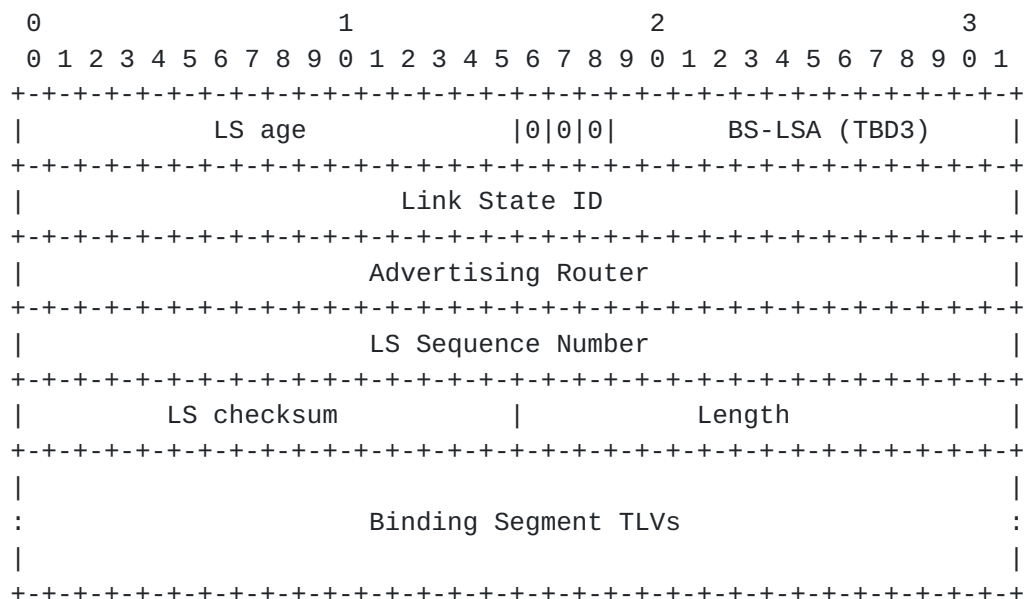
Figure 6: OSPFv3 Binding Segment Opaque LSA

The U-bit is set to 0, and the scope is set to 00 for link-local
scoping.

2.2. Extensions to IS-IS

2.2.1. Advertising Proxy Forwarding

When a node P has the capability to do a SR proxy forwarding for its
neighboring nodes for protecting the failures of them, node P
advertises its SR proxy forwarding capability in its LSP, which
contains a Router Capability TLV of Type 242 including a SR
capabilities sub-TLV of sub-Type 2.

One bit (called PF bit as shown in Figure 7) in the Flags field of
the SR capabilities sub-TLV is defined to indicate node P's SR proxy
forwarding capability. When this bit is set to one by node P, it
indicates that node P is capable of doing a SR proxy forwarding for
its neighboring nodes.

```
 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 (2)    |    Length     |    Flags      |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                    Range                       |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
//              SID/Label Sub-TLV (variable)                  //
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+


                 0  1  2  3  4  5  6  7
                +--+--+--+--+--+--+--+--+
                | I| V|PF|              |
                +--+--+--+--+--+--+--+--+
                          Flags
```
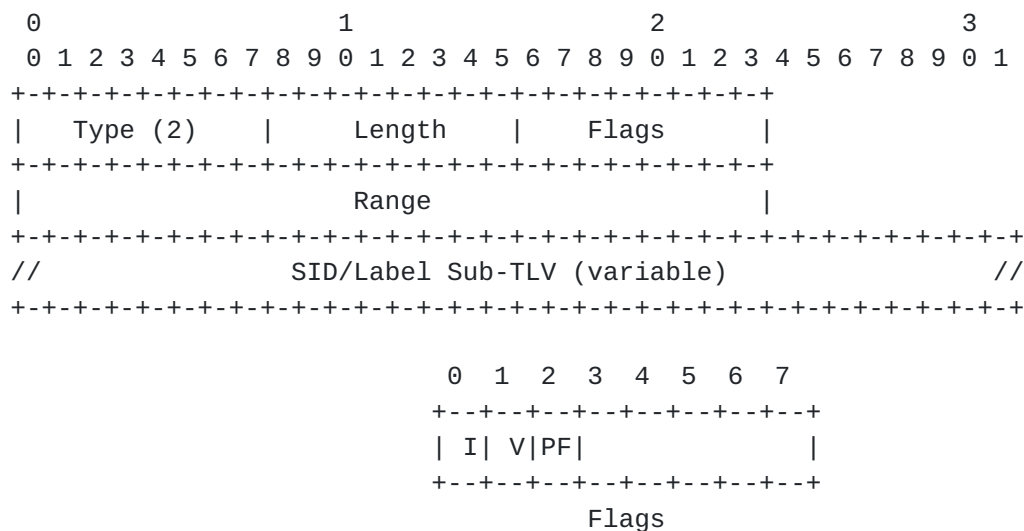
Figure 7: SR Capabilities sub-TLV

If node P can not do a SR proxy forwarding for all its neighboring
nodes, but for some of them, then it advertises the node SID of each
of the nodes as a proxy node SID, indicating that it is able to do
proxy forwarding for the node SID.

The IS-IS SID/Label Binding TLV (suggested value 149) is defined in
[I-D.ietf-isis-segment-routing-extensions]. A Proxy Forwarder uses
the SID/Label Binding TLV to advertise the node SID of its
neighboring node. The Flags field of the SID/Label Binding TLV is
extended to include a P flag as shown in Figure 8. The prefix/node

SID in prefix/node SID Sub-TLV included in SID/Label Binding TLV is
identified as a proxy forwarding prefix/node SID.

```
    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     |     Flags     |    RESERVED   |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |           Range               | Prefix Length |    Prefix     |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   //               Prefix (continued, variable)                 //
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |                     SubTLVs (variable)                       |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+


                           0 1 2 3 4 5 6 7
                          +-+-+-+-+-+-+-+-+
                          |F|M|S|D|A|P|   |
                          +-+-+-+-+-+-+-+-+
                               Flags
```
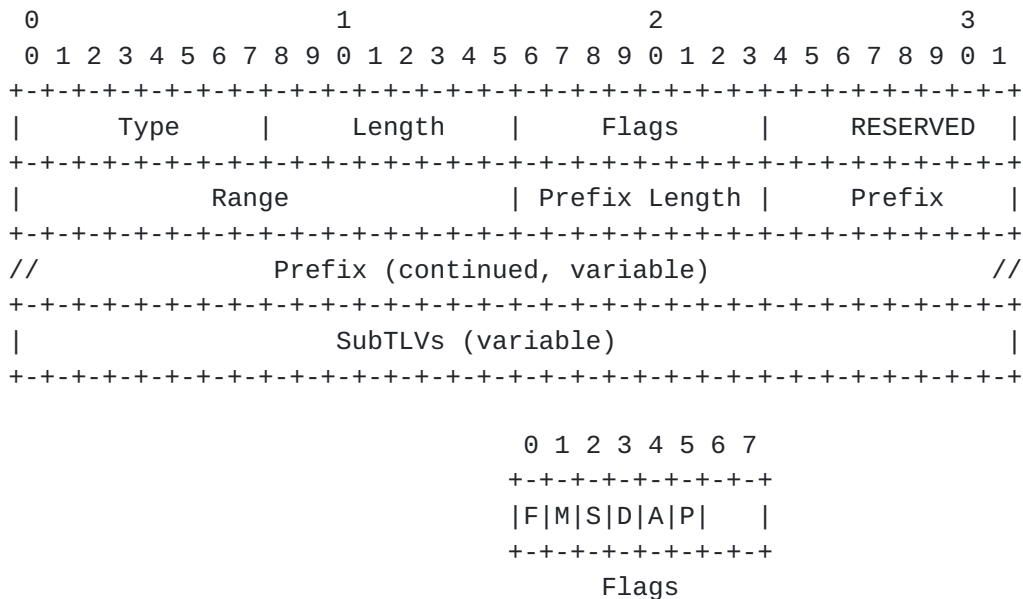
Figure 8: SID/Label Binding TLV

Where:

P-Flag: Proxy forwarding flag. If set, this prefix/node SID is
advertised by the proxy node. This TLV is used to announce that the
node has the ability to proxy forward the prefix/node SID.

When the P-flag is set in the SID/Label Binding TLV, the following
usage rules apply.

The Range, Prefix Length and Prefix field are not used. They should
be set to zero on transmission and ignored on receipt.

SID/Label Binding TLV contains a number of prefix/node SID Sub-TLVs.
The TLV advertised by a proxy forwarding node P contains prefix/node
SID Sub-TLVs for the node SIDs of P's neighbor nodes. Each of the
Sub-TLVs is a prefix/node SID Sub-TLV defined in
[I-D.ietf-isis-segment-routing-extensions]. From the SID in a
prefix/node SID Sub-TLV advertised by the Proxy Forwarding node, its
prefix can be obtained through matching corresponding prefix/node SID
advertised by the neighbor/protected node using TLV-135 (or 235, 236,
or 237) together with the prefix/node SID Sub-TLV.

2.2.2. **Advertising Binding Segment**

[I-D.ietf-spring-segment-routing-policy] has defined the usage of
binding-SID. For supporting binding SID proxy forwarding, a new IS-
IS TLV, called Binding Segment TLV, is defined. It contains a
binding SID and a list of segments (SIDs). This TLV may be
advertised in IS-IS Hello (IIH) PDUs, LSPs, or in Circuit Scoped Link
State PDUs (CS-LSP) [RFC7356]. Its format is shown in Figure 9.

```
 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     |BindingSID Type|   SIDs Type   |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
~                 Binding SID value/Sub-TLV                     ~
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
~                    SID values/Sub-TLVs                        ~
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```
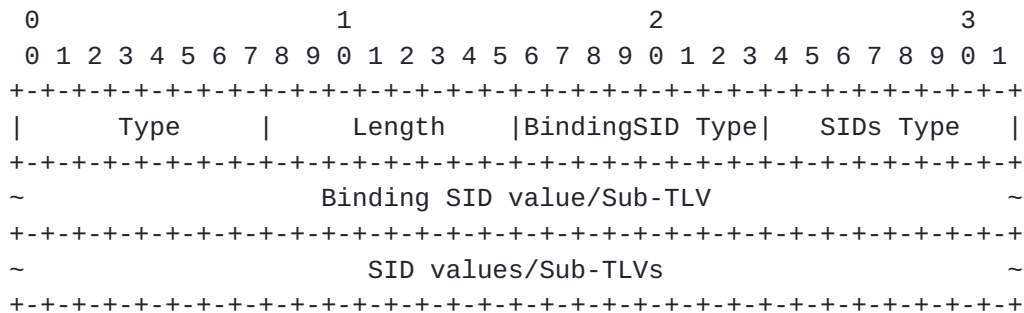
                  Figure 9: IS-IS Binding Segment TLV

The fields of this TLV are defined as follows:

Type: 1 octet Suggested value 152 (to be assigned by IANA)

Length: 1 octet (2 + length of Sub-TLVs/values).

Binding SID Type (BT): 1 octet indicates whether the binding SID is
represented by a Sub-TLV or a value included in the TLV. For the
binding SID represented by a value, it indicates the type of binding
SID. The following BT values are defined:

o BT = 0: The binding SID is represented by a Sub-TLV (i.e., binding
SID Sub-TLV) in the TLV. A binding SID Sub-TLV is a SID/Label Sub-
TLV defined in [I-D.ietf-isis-segment-routing-extensions]. BT != 0
indicates that the binding SID is represented by a value.

o BT = 1: The binding SID value is a label, which is represented by
the 20 rightmost bits. The length of the value is 3 octets.

o BT = 2: The binding SID value is a 32-bit SID. The length of the
value is 4 octets.

SIDs Type (ST): 1 octet indicates whether the SIDs are represented by
Sub-TLVs or values included in the TLV. For the SIDs represented by
values, it indicates the type of SIDs. The following ST values are
defined:

o ST = 0: The SIDs are represented by Sub-TLVs (i.e., SID Sub-TLVs)
in the TLV. A SID Sub-TLV is an Adj-SID Sub-TLV, a Prefix-SID Sub-
TLV or a SID/Label Sub-TLV defined in
[I-D.ietf-isis-segment-routing-extensions]. ST != 0 indicates that
the SIDs are represented by values.

o ST = 1: Each of the SID values is a label, which is represented by
the 20 rightmost bits. The length of the value is 3 octets.

o ST = 2: Each of the SID values is a 32-bit SID. The length of the
value is 4 octets.

3. Building Proxy Forwarding Table

Figure 10 is used to illustrate the SR proxy forwarding approach.
Each node N has SRGB = [N000-N999]. RT1 is an ingress node of SR
domain. RT3 is a failure node. RT2 is a Point of Local Repair (PLR)
node, i.e., a proxy forwarding node. Three label stacks are shown in
the figure. Label Stack 1 uses only adjacency-SIDs and represents
the path RT1->RT2->RT3->RT4->RT5. Label Stack 2 uses only node-SIDs
and represents the ECMP-aware path RT1->RT3->RT4->RT5. Label Stack 3
uses a node-SID and a binding SID. The Binding-SID with label=100 at
RT3 represents the ECMP-aware path RT3->RT4->RT5. So Label Stack 3,
which consists of the node-SID for RT3 following by Binding-SID=100,
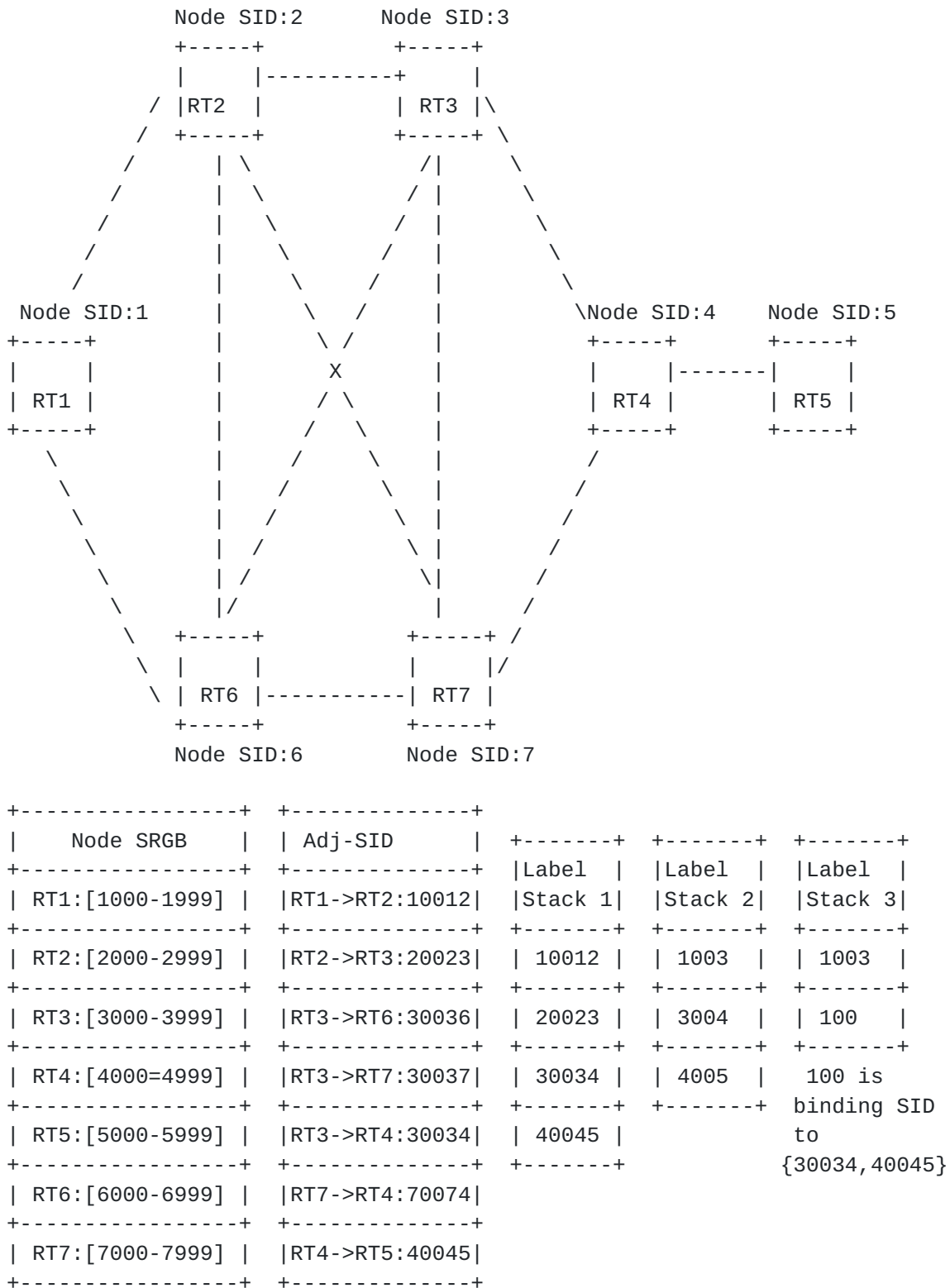represents the ECMP-aware path RT1->RT3->RT4->RT5.

```
          Node SID:2      Node SID:3
           +-----+          +-----+
           |     |----------+      |
          / |RT2 |          | RT3 |\
         /  +-----+          +-----+ \
        /      | \          /|     \
       /       |  \        / |      \
      /        |   \      /  |       \
     /         |    \    /   |        \
    /          |     \  /    |         \
  Node SID:1   |      \/     |       \Node SID:4   Node SID:5
  +-----+      |      /\     |         +-----+      +-----+
  |     |      |     X       |         |     |      |------| |     |
  | RT1 |      |    / \      |         | RT4 |      | RT5 |
  +-----+      |   /   \     |         +-----+      +-----+
    \          |  /     \    |          /
     \         | /       \   |         /
      \        |/         \  |        /
       \       |           \ |       /
        \      |            \|      /
         \  +-----+          +-----+ /
          \ |     |          |     |/
           \| RT6 |----------| RT7 |
            +-----+          +-----+
          Node SID:6      Node SID:7
```

```
+-----------------+  +---------------+
|    Node SRGB    |  | Adj-SID       |  +-------+  +-------+  +-------+
+-----------------+  +---------------+  |Label  |  |Label  |  |Label  |
| RT1:[1000-1999] |  |RT1->RT2:10012|  |Stack 1|  |Stack 2|  |Stack 3|
+-----------------+  +---------------+  +-------+  +-------+  +-------+
| RT2:[2000-2999] |  |RT2->RT3:20023|  | 10012 |  | 1003  |  | 1003  |
+-----------------+  +---------------+  +-------+  +-------+  +-------+
| RT3:[3000-3999] |  |RT3->RT6:30036|  | 20023 |  | 3004  |  | 100   |
+-----------------+  +---------------+  +-------+  +-------+  +-------+
| RT4:[4000=4999] |  |RT3->RT7:30037|  | 30034 |  | 4005  |    100 is
+-----------------+  +---------------+  +-------+  +-------+  binding SID
| RT5:[5000-5999] |  |RT3->RT4:30034|  | 40045 |              to
+-----------------+  +---------------+  +-------+           {30034,40045}
| RT6:[6000-6999] |  |RT7->RT4:70074|
+-----------------+  +---------------+
| RT7:[7000-7999] |  |RT4->RT5:40045|
+-----------------+  +---------------+
```

Figure 10: Topology of SR-TE Path

3.1. Advertising Proxy Forwarding

If the Point of Local Repair (PLR), for example, RT2, has the
capability to do a SR proxy forwarding for all its neighboring nodes,
it must advertise this capability. If the PLR can not do a SR proxy
forwarding for all its neighboring nodes, but for some of them, for
example, RT3, then it uses proxy Node SIDs TLV to advertise the
prefix-SID learned from RT3. The TLV contains the Sub-TLV/value for
the prefix/node SID of RT3 as a proxy SID. When RT3 fails, RT2 needs
to maintain the Sub-TLV/value for a period of time. When the proxy
forwarding table corresponding to the fault node is deleted (see
section 3.2), the Sub-TLV/value is withdrawn. The nodes in the
network (for example, RT1) learn the prefix/node SID TLV advertised
by RT3 and the proxy Node SIDs TLV advertised by RT2. When RT3 is
normal, the nodes prefer prefix/node SID TLV. When the RT3 fails,
the proxy prefix/node SIDs TLV advertised by RT2 is preferred.

3.2. Building Proxy Forwarding Table

A SR proxy node P needs to build an independent proxy forwarding
table for each neighbor N. The proxy forwarding table for node N
contains the following information:

1: Node N's SRGB range and the difference between the SRGB start
value of node P and that of node N;

2: All adjacency-SID of N and Node-SID of the node pointed to by node
N's adjacency-SID.

3: The binding-SID of N and the label stack associated with the
binding-SID.

Node P (PLR) uses a proxy forwarding table based on the next segment
to find a node N as a backup forwarding entry to the adj-SID and
Node-SID of node N. When node N fails, the proxy forwarding table
needs to be maintained for a period of time, which is recommended for
30 minutes.

Node RT3 in the topology of Figure 1 is node N, and node RT2 is node
P (PLR). RT2 builds the proxy forwarding table for RT3. The
structure of the table and how to build the table is a local
implementation issue.

4. Node Protection for Segment List

Segment Routing Traffic Engineering supports the creation of explicit
paths using adjacency-SIDs, node-SIDs, and binding-SIDs. The label
stack is a combination of one or more of adjacency-SIDs, node-SIDs,

and binding-SIDs. This Section shows how a proxy node uses the SR
proxy forwarding mechanism to protect traffic to the destination node
when the next segment of label stack is adjacency-SIDs, node-SIDs, or
binding-SIDs, respectively.

4.1. Next Segment is an Adjacency Segment

As shown in Figure 1, Label Stack 1 {10012, 20023, 30034, 40045}
represents SR-TE strict explicit path RT1->RT2->RT3->RT4->RT5. When
RT3 fails, node RT2 acts as a PLR, and uses next adj-SID (30034) of
the label stack to lookup the proxy forwarding table built by RT2
locally for RT3. The path returned is the label forwarding path to
RT3's next hop node RT4, which bypasses RT3. The specific steps are
as follows:

a. RT1 pops top adj-SID 10012, and forwards the packet to RT2;

b. RT2 uses the label 20023 to identify the next hop node RT3, which
has failed. RT2 pops label 20023 and queries the Proxy Forwarding
Table corresponding to RT3 with label 30034. The Proxy Forwarding
Table corresponding to RT3 returns an outgoing interface and label
stack representing a path to RT4 that does not pass through RT3. In
this case, outgoing interface to RT7 with label stack 7004, satisfies
this requirement.

c. So the packet leaves RT2 out the interface to RT7 with label
stack {7004, 40045}. RT4 forwards it to RT4, where the original path
is rejoined.

d. RT2 forwards packets to RT7. RT7 queries the local routing table
to forward the packet to RT4.

4.2. Next Segment is a Node Segment

As shown in Figure 1, Label Stack 2 {1003, 3004, 4005} represents SR-
TE loose path RT1->RT3->RT4->RT5, where 1003 is the node SID of RT3.

When the node RT3 fails, the proxy forwarding TLV advertised by the
RT2 is preferred to direct the traffic of the RT1 to the PLR node
RT2. Node RT2 acts as a PLR node and queries the proxy forwarding
table locally built for RT3. The path returned is the label
forwarding path to RT3's next hop node RT4, which bypasses RT3. The
specific steps are as follows:

a. RT1 swaps label 1003 to out-label 2003 to RT3.

b. RT2 receives the label forwarding packet whose top label of label
stack is 2003, and searches for the local Routing Table, the behavior
found is to lookup Proxy Forwarding table due to RT3 failure.

c. RT2 uses 2003 as the in-label to lookup Proxy Forwarding table,
and the query result is forwarding the packet to RT4.

d. Then RT2 queries the Routing Table to RT4, using the primary or
backup path to RT4. The next hop is RT7.

e. RT2 forwards the packet to RT7. RT7 queries the local routing
table to forward the packet to RT4.

f. After RT1 convergences, node SID 1003 is preferred to the proxy
SID implied/advertised by RT2.

4.3. Next Segment is a Binding Segment

As shown in Figure 1, Label Stack 3 {1003, 100} represents SR-TE
loose path RT1->RT3->RT4->RT5, where 100 is a Binding-SID, which
represents segment list {30034, 40045}.

When the node RT3 fails, the proxy forwarding SID implied or
advertised by the RT2 is preferred to forward the traffic of the RT1
to the PLR node RT2. Node RT2 acts as a PLR node and uses Binding-
SID to query the proxy forwarding table locally built for RT3. The
path returned is the label forwarding path to RT3's next hop node
(RT4), which bypasses RT3. The specific steps are as follows:

a. RT1 swaps label 1003 to out-label 2003 to RT3.

b. RT2 receives the label forwarding packet whose top label of label
stack is 2003, and searches for the local Routing Table, the behavior
found is to lookup Proxy Forwarding table due to RT3 failure.

c. RT2 uses Binding-SID:100 (label 2003 has pop) as the in-label to
lookup the Next Label record of the Proxy Forwarding Table, the
behavior found is to swap to Segment list {30034, 40045}.

d. RT2 swaps Binding-SID:100 to Segment list {30034, 40045}, and
uses the 3034 to lookup the Next Label record of the Proxy Forwarding
table again. The behavior found is to forward the packet to RT4.

e. RT2 queries the Routing Table to RT4, using primary or backup
path to RT4. The next hop is RT7.

f. RT2 forwards packets to RT7. RT7 queries the local routing table
to forward the packet to RT4.

5. Security Considerations

   The extensions to OSPF and IS-IS described in this document result in
   two types of behaviors in data plane when a node in a network fails.
   One is that for a node, which is a upstream (except for the direct
   upstream) node of the failed node along a SR-TE path, it continues to
   send the traffic to the failed node along the SR-TE path for an
   extended period of time. The other is that for a node, which is the
   direct upstream node of the failed node, it fast re-routes the
   traffic around the failed node to the direct downstream node of the
   failed node along the SR-TE path. These behaviors are internal to a
   network and should not cause extra security issues.

6. IANA Considerations

6.1. OSPFv2

   Under Subregistry Name "OSPF Router Functional Capability Bits"
   within the "Open Shortest Path First v2 (OSPFv2) Parameters"
   [RFC7770], IANA is requested to assign one bit for Proxy Forwarding
   Capability as follows:

   +============+==================+==================+
   | Bit number | Capability Name | Reference |
   +============+==================+==================+
   | 31 | Proxy Forwarding | This document |
   +------------+------------------+------------------+

   Under Registry Name "OSPFv2 Extended Prefix Opaque LSA TLVs"
   [RFC7684], IANA is requested to assign one new TLV value for OSPF
   Proxy Node SIDs as follows:

   +============+====================+================+
   | TLV Value | TLV Name | Reference |
   +============+====================+================+
   | 2 | Proxy Node SIDs TLV | This document |
   +------------+--------------------+----------------+

   Under Registry Name "Opaque Link-State Advertisements (LSA) Option
   Types" [RFC5250], IANA is requested to assign new Opaque Type
   registry values for Binding Segment LSA as follows:

   +================+==================+================+
   | Registry Value | Opaque Type | Reference |
   +================+==================+================+
   | 10 | Binding Segment | This document |
   +----------------+------------------+----------------+

   IANA is requested to create and maintain new registries:

    o OSPFv2 Binding Segment Opaque LSA TLVs

   Initial values for the registry are given below. The future
   assignments are to be made through IETF Review [RFC5226].

```
   Value           TLV Name                    Definition
   -----           -----------------------     ----------
   0               Reserved
   1               Binding Segment TLV         This Document
   2-32767         Unassigned
   32768-65535     Reserved
```

6.2. OSPFv3

   Under Registry Name "OSPFv3 LSA Function Codes", IANA is requested to
   assign new registry values for Binding Segment LSA as follows:

```
   +========+========================+================+
   | Value  | LSA Function Code Name | Reference      |
   +========+========================+================+
   | 16     | Binding Segment LSA    | This document  |
   +--------+------------------------+----------------+
```

   IANA is requested to create and maintain new registries:

    o OSPFv3 Binding Segment LSA TLVs

   Initial values for the registry are given below. The future
   assignments are to be made through IETF Review [RFC5226].

```
   Value           TLV Name                    Definition
   -----           -----------------------     ----------
   0               Reserved
   1               Binding Segment TLV         This Document
   2-32767         Unassigned
   32768-65535     Reserved
```

6.3. IS-IS

   Under Registration "Segment Routing Capability" in the "sub-TLVs for
   TLV 242" registry [I-D.ietf-isis-segment-routing-extensions], IANA is
   requested to assign one bit flag for Proxy Forwarding Capability as
   follows:

```
+============+=======================+===============+
| Bit number | Capability Name       | Reference     |
+============+=======================+===============+
|     2      | Proxy Forwarding (PF) | This document |
+-----------+-----------------------+--------------+
```

Under Registration "Segment Identifier/Label Binding TLV 149"
[I-D.ietf-isis-segment-routing-extensions], IANA is requested to
assign one bit P-Flag as follows:

```
+============+=================+===============+
| Bit number | Flag Name       | Reference     |
+============+=================+===============+
|     5      | P-Flag          | This document |
+-----------+-----------------+--------------+
```

Under Registry Name: IS-IS TLV Codepoints, IANA is requested to
assign one new TLV value for IS-IS Binding Segment as follows:

```
+========+=====================+===============+
| Value  | TLV Name            | Reference     |
+========+=====================+===============+
|  152   | Binding Segment TLV | This Document |
+--------+---------------------+--------------+
```

7. Acknowledgements

The authors would like to thank Peter Psenak, Acee Lindem, Les
Ginsberg, Bruno Decraene and Jeff Tantsura for their comments to this
work.

8. References

8.1. Normative References

[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-25 (work in progress), May 2019.

[I-D.ietf-ospf-segment-routing-extensions]
          Psenak, P., Previdi, S., Filsfils, C., Gredler, H.,
          Shakir, R., Henderickx, W., and J. Tantsura, "OSPF
          Extensions for Segment Routing", draft-ietf-ospf-segment-
          routing-extensions-27 (work in progress), December 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>.

   [RFC5226] Narten, T. and H. Alvestrand, "Guidelines for Writing an
              IANA Considerations Section in RFCs", RFC 5226,
              DOI 10.17487/RFC5226, May 2008,
              <https://www.rfc-editor.org/info/rfc5226>.

   [RFC5250] Berger, L., Bryskin, I., Zinin, A., and R. Coltun, "The
              OSPF Opaque LSA Option", RFC 5250, DOI 10.17487/RFC5250,
              July 2008, <https://www.rfc-editor.org/info/rfc5250>.

   [RFC7356] Ginsberg, L., Previdi, S., and Y. Yang, "IS-IS Flooding
              Scope Link State PDUs (LSPs)", RFC 7356,
              DOI 10.17487/RFC7356, September 2014,
              <https://www.rfc-editor.org/info/rfc7356>.

   [RFC7684] Psenak, P., Gredler, H., Shakir, R., Henderickx, W.,
              Tantsura, J., and A. Lindem, "OSPFv2 Prefix/Link Attribute
              Advertisement", RFC 7684, DOI 10.17487/RFC7684, November
              2015, <https://www.rfc-editor.org/info/rfc7684>.

   [RFC7770] Lindem, A., Ed., Shen, N., Vasseur, JP., Aggarwal, R., and
              S. Shaffer, "Extensions to OSPF for Advertising Optional
              Router Capabilities", RFC 7770, DOI 10.17487/RFC7770,
              February 2016, <https://www.rfc-editor.org/info/rfc7770>.

8.2. Informative References

   [I-D.hegde-spring-node-protection-for-sr-te-paths]
              Hegde, S., Bowers, C., Litkowski, S., Xu, X., and F. Xu,
              "Node Protection for SR-TE Paths", draft-hegde-spring-
              node-protection-for-sr-te-paths-07 (work in progress),
              July 2020.

   [I-D.ietf-rtgwg-segment-routing-ti-lfa]
              Litkowski, S., Bashandy, A., Filsfils, C., Decraene, B.,
              Francois, P., Voyer, D., Clad, F., and P. Camarillo,
              "Topology Independent Fast Reroute using Segment Routing",
              draft-ietf-rtgwg-segment-routing-ti-lfa-03 (work in
              progress), March 2020.

   [I-D.ietf-spring-segment-routing-policy]
              Filsfils, C., Talaulikar, K., Voyer, D., Bogdanov, A., and
              P. Mattes, "Segment Routing Policy Architecture", draft-
              ietf-spring-segment-routing-policy-08 (work in progress),
              July 2020.

   [I-D.sivabalan-pce-binding-label-sid]
              Sivabalan, S., Filsfils, C., Tantsura, J., Hardwick, J.,
              Previdi, S., and C. Li, "Carrying Binding Label/Segment-ID
              in PCE-based Networks.", draft-sivabalan-pce-binding-
              label-sid-07 (work in progress), July 2019.

   [RFC5462] Andersson, L. and R. Asati, "Multiprotocol Label Switching
              (MPLS) Label Stack Entry: "EXP" Field Renamed to "Traffic
              Class" Field", RFC 5462, DOI 10.17487/RFC5462, February
              2009, <https://www.rfc-editor.org/info/rfc5462>.

Authors' Addresses

   Zhibo Hu
   Huawei Technologies
   Huawei Bld., No.156 Beiqing Rd.
   Beijing 100095
   China

   Email: huzhibo@huawei.com

   Huaimo Chen
   Futurewei
   Boston, MA
   USA

   Email: Huaimo.chen@futurewei.com

   Junda Yao
   Huawei Technologies
   Huawei Bld., No.156 Beiqing Rd.
   Beijing 100095
   China

   Email: yaojunda@huawei.com

Chris Bowers
Juniper Networks
1194 N. Mathilda Ave.
Sunnyvale, CA 94089
USA

Email: cbowers@juniper.net

Yongqing
China Telecom
109, West Zhongshan Road, Tianhe District
Guangzhou 510000
China

Email: zhuyq8@chinatelecom.cn