IS-IS

In This Chapter

This chapter provides information to configure Intermediate System to Intermediate System (IS-IS).

Topics in this chapter include:

Configuring IS-IS

Intermediate-system-to-intermediate-system (IS-IS) is a link-state interior gateway protocol (IGP) which uses the Shortest Path First (SPF) algorithm to determine routes. Routing decisions are made using the link-state information. IS-IS evaluates topology changes and, if necessary, performs SPF recalculations.

Entities within IS-IS include networks, intermediate systems, and end systems. In IS-IS, a network is an autonomous system (AS), or routing domain, with end systems and intermediate systems. A router is an intermediate system. End systems are network devices which send and receive protocol data units (PDUs), the OSI term for packets. Intermediate systems send, receive, and forward PDUs.

End system and intermediate system protocols allow routers and nodes to identify each other. IS-IS sends out link-state updates periodically throughout the network, so each router can maintain current network topology information.

IS-IS supports large ASs by using a two-level hierarchy. A large AS can be administratively divided into smaller, more manageable areas. A system logically belongs to one area. Level 1 routing is performed within an area. Level 2 routing is performed between areas. The routers can be configured as Level 1, Level 2, or both Level 1/2.

Figure 14 displays an example of an IS-IS routing domain.

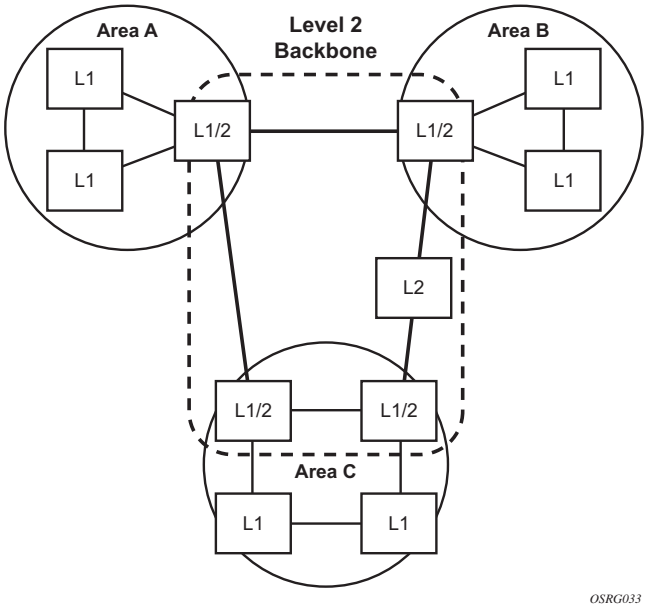

Figure 14: IS-IS Routing Domain

Routing

OSI IS-IS routing uses two-level hierarchical routing. A routing domain can be partitioned into areas. Level 1 routers know the topology in their area, including all routers and end systems in their area but do not know the identity of routers or destinations outside of their area. Level 1 routers forward traffic with destinations outside of their area to a Level 2 router in their area.

Level 2 routers know the Level 2 topology, and know which addresses are reachable by each Level 2 router. Level 2 routers do not need to know the topology within any Level 1 area, except to the extent that a Level 2 router can also be a Level 1 router within a single area. By default, only Level 2 routers can exchange PDUs or routing information directly with external routers located outside the routing domain.

In IS-IS, there are two types of routers:

- Level 1 intermediate systems — Routing is performed based on the area ID portion of the ISO address called the *network entity title* (NET). Level 1 systems route within an area. They recognize, based on the destination address, whether the destination is within the area. If so, they route toward the destination. If not, they route to the nearest Level 2 router.
- Level 2 intermediate systems — Routing is performed based on the area address. They route toward other areas, disregarding other area's internal structure. A Level 2 intermediate system can also be configured as a Level 1 intermediate system in the same area.

The Level 1 router's area address portion is manually configured (see ISO Network Addressing on page 454). A Level 1 router will not become a neighbor with a node that does not have a common area address. However, if a Level 1 router has area addresses A, B, and C, and a neighbor has area addresses B and D, then the Level 1 router will accept the other node as a neighbor, as address B is common to both routers. Level 2 adjacencies are formed with other Level 2 nodes whose area addresses do not overlap. If the area addresses do not overlap, the link is considered by both routers to be Level 2 only and only Level 2 LSPDUs flow on the link.

Within an area, Level 1 routers exchange LSPs which identify the IP addresses reachable by each router. Specifically, zero or more IP address, subnet mask, and metric combinations can be included in each LSP. Each Level 1 router is manually configured with the IP address, subnet mask, and metric combinations, which are reachable on each interface. A Level 1 router routes as follows:

- If a specified destination address matches an IP address, subnet mask, or metric reachable within the area, the PDU is routed via Level 1 routing.
- If a specified destination address does not match any IP address, subnet mask, or metric combinations listed as reachable within the area, the PDU is routed towards the nearest Level 2 router.

Level 2 routers include in their LSPs, a complete list of IP address, subnet mask, and metrics specifying all the IP addresses which reachable in their area. This information can be obtained from a combination of the Level 1 LSPs (by Level 1 routers in the same area). Level 2 routers can also report external reachability information, corresponding to addresses reachable by routers in other routing domains or autonomous systems.

IS-IS Frequently Used Terms

- Area — An area is a routing sub-domain which maintains detailed routing information about its own internal composition, and also maintains routing information which allows it to reach other routing sub-domains. Areas correspond to the Level 1 sub-domain.
- End system — End systems send NPDUs to other systems and receive NPDUs from other systems, but do not relay NPDUs. This International Standard does not specify any additional end system functions beyond those supplied by ISO 8473 and ISO 9542.
- Neighbor — A neighbor is an adjacent system reachable by traversing a single sub-network by a PDU.
- Adjacency — An adjacency is a portion of the local routing information which pertains to the reachability of a single neighboring end or intermediate system over a single circuit. Adjacencies are used as input to the decision process to form paths through the routing domain. A separate adjacency is created for each neighbor on a circuit and for each level of routing (Level 1 and Level 2) on a broadcast circuit.
- Circuit — The subset of the local routing information base pertinent to a single local Subnetwork Point of Attachments (SNPAs).
- Link — The communication path between two neighbors. A link is up when communication is possible between the two SNPAs.
- Designated IS — The intermediate system on a LAN which is designated to perform additional duties. In particular, the designated IS generates link-state PDUs on behalf of the LAN, treating the LAN as a pseudonode.
- Pseudonode — Where a broadcast sub-network has *n* connected intermediate systems, the broadcast sub-network itself is considered to be a pseudonode. The pseudonode has links to each of the *n* intermediate systems and each of the ISs has a single link to the pseudonode (rather than *n*-1 links to each of the other intermediate systems). Link-state PDUs are generated on behalf of the pseudonode by the designated IS.
- Broadcast sub-network — A multi-access subnetwork that supports the capability of addressing a group of attached systems with a single PDU.
- General topology sub-network — A topology that is modeled as a set of point-to-point links, each of which connects two systems. There are several generic types of general topology subnetworks, multipoint links, permanent point-to-point links, dynamic and static point-to-point links.
- Routing sub-domain — A routing sub-domain consists of a set of intermediate systems and end systems located within the same routing domain.
- Level 2 sub-domain — Level 2 sub-domain is the set of all Level 2 intermediate systems in a routing domain.

ISO Network Addressing

IS-IS uses ISO network addresses. Each address identifies a point of connection to the network, such as a router interface, and is called a Network Service Access Point (NSAP).

An end system can have multiple NSAP addresses, in which case the addresses differ only by the last byte (called the *n-selector*). Each NSAP represents a service that is available at that node. In addition to having multiple services, a single node can belong to multiple areas.

Each network entity has a special network address called a Network Entity Title (NET). Structurally, a NET is identical to an NSAP address but has an n-selector of 00. Most end systems have one NET. Intermediate systems can have up to three area IDs (area addresses).

NSAP addresses are divided into three parts. Only the area ID portion is configurable.

- Area ID — A variable length field between 1 and 13 bytes long. This includes the Authority and Format Identifier (AFI) as the most significant byte and the area ID.
- System ID — A six-byte system identification. This value is not configurable. The system ID is derived from the system or router ID.
- Selector ID — A one-byte selector identification that must contain zeros when configuring a NET. This value is not configurable. The selector ID is always 00.

Of the total 20 bytes comprising the NET, only the first 13 bytes, the area ID portion, can be manually configured. As few as one byte can be entered or, at most, 13 bytes. If less than 13 bytes are entered, the rest is padded with zeros.

Routers with common area addresses form Level 1 adjacencies. Routers with no common NET addresses form Level 2 adjacencies, if they are capable (Figure 15).

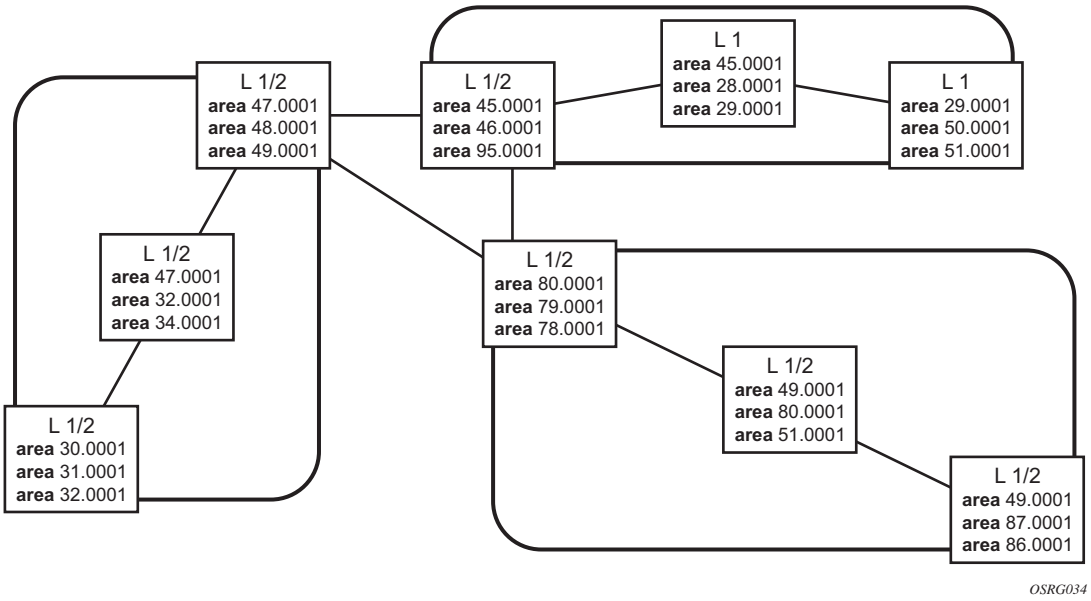

Figure 15: Using Area Addresses to Form Adjacencies

IS-IS PDU Configuration

The following PDUs are used by IS-IS to exchange protocol information:

- IS-IS hello PDU — Routers with IS-IS enabled send hello PDUs to IS-IS-enabled interfaces to discover neighbors and establish adjacencies.
- Link-state PDUs — Contain information about the state of adjacencies to neighboring IS-IS systems. LSPs are flooded periodically throughout an area.
- Complete sequence number PDUs — In order for all routers to maintain the same information, CSNPs inform other routers that some LSPs can be outdated or missing from their database. CSNPs contain a complete list of all LSPs in the current IS-IS database.
- Partial sequence number PDUs (PSNPs) — PSNPs are used to request missing LSPs and acknowledge that an LSP was received.

IS-IS Operations

The routers perform IS-IS routing as follows:

- Hello PDUs are sent to the IS-IS-enabled interfaces to discover neighbors and establish adjacencies.
- IS-IS neighbor relationships are formed if the hello PDUs contain information that meets the criteria for forming an adjacency.
- SRs can build a link-state PDU based upon their local interfaces that are configured for IS-IS and prefixes learned from other adjacent routers.
- SRs flood LSPs to the adjacent neighbors except the neighbor from which they received the same LSP. The link-state database is constructed from these LSPs.
- A Shortest Path Tree (SPT) is calculated by each IS, and from this SPT the routing table is built.

IS-IS Route Summarization

IS-IS IPv4 route summarization allows users to create aggregate IPv4 addresses that include multiple groups of IPv4 addresses for a given IS-IS level. IPv4 Routes redistributed from other routing protocols also can be summarized. It is similar to the OSPF area-range command. IS-IS IPv4 route summarization helps to reduce the size of the LSDB and the IPv4 routing table, and it also helps to reduce the chance of route flapping.

IPv4 route summarization supports:

- Level 1, Level 1-2, and Level 2
- Route summarization for the IPv4 routes redistributed from other protocols
- Metric used to advertise the summary address will be the smallest metric of all the more specific IPv4 routes.

Partial SPF Calculation

IS-IS supports partial SPF calculation, also referred to as partial route calculation.When an event does not change the topology of the network, IS-IS will not perform full SPF but will instead perform an IP reach calculation for the impacted routes. Partial SPF is performed at the receipt of IS-IS LSPs with changes to IP reach TLVs and in general, for any IS-IS LSP TLV and sub-TLV change that does not impact the network topology.

IS-IS MT-Topology Support

Multi-Topology IS-IS (MT-ISIS) support within SR OS allows for the creation of different topologies within IS-IS that contribute routes to specific route tables for IPv4 unicast, IPv6 unicast, IPv4 multicast, and IPv6 multicast. This capability allows for non-congruent topologies between these different routing tables. As a result, networks are able to control which links or nodes are to be used for forwarding different types of traffic.

For example, MT-ISIS could allow all links to carry IPv4 traffic, while only a subset of links can also carry IPv6 traffic.

SR OS supports the following Multi-Topologies:

- IPv4 Unicast – MT-ID 0
- IPv6 Unicast – MT-ID 2
- IPv4 Multicast – MT-ID 3
- IPv6 Multicast – MT-ID 4

Native IPv6 Support

IS-IS IPv6 TLVs for IPV6 routing is supported in SR OS. This support is considered native IPv6 routing within IS-IS. However, it has limitations in that IPv4 and IPv6 topologies must be congruent, otherwise traffic may be blackholed. Service providers should ensure that the IPv4 topology and IPv6 topologies are the same if native IPv6 routing is used within IS-IS.

IS-IS Administrative Tags

IS-IS admin tags enable a network administrator to configure route tags to tag IS-IS route prefixes. These tags can subsequently be used to control Intermediate System-to-Intermediate System (IS-IS) route redistribution or route leaking.

The IS-IS support for route tags allows the tagging of IP addresses of an interface and use the tag to apply administrative policy with a route map. A network administrator can also tag a summary route and then use a route policy to match the tag and set one or more attributes for the route.

Using these administrative policies allow the operator to control how a router handles the routes it receives from and sends to its IS-IS neighboring routers. Administrative policies are also used to govern the installation of routes in the routing table.

Route tags allow:

- Policies to redistribute routes received from other protocols in the routing table to IS-IS.
- Policies to redistribute routes between levels in an IS-IS routing hierarchy.
- Policies to summarize routes redistributed into IS-IS or within IS-IS by creating aggregate (summary) addresses.

Setting Route Tags

IS-IS route tags are configurable in the following ways:

- Setting a route tag for an IS-IS interface.
- Setting a route tag on an IS-IS passive interface.
- Setting a route tag for a route redistributed from another protocol to IS-IS.
- Setting a route tag for a route redistributed from one IS-IS level to another IS-IS level.
- Setting a route tag for an IS-IS default route.
- Setting a route tag for an IS-IS summary address.

Using Route Tags

Although an operator on this or on a neighboring IS-IS router has configured setting of the IS-IS administrative tags, it will not have any effect unless policies are configure to instruct how to process the given tag value.

Policies can process tags where IS-IS is either the origin, destination or both origin and destination protocol.

```
config>router>policy-options>policy-statement>entry>from
     config>router>policy-options>policy-statement>entry>action tag tag-value
     config>router>policy-options>policy-statement# default-action tag tag-value
```

Unnumbered Interface Support

IS-IS supports unnumbered point-to-point interface with both Ethernet and PPP encapsulations.

Unnumbered interfaces borrow the address from other interfaces such as system or loopback interfaces and uses it as the source IP address for packets originated from the interface. This feature supports both dynamic and static ARP for unnumbered interfaces to allow interworking with unnumbered interfaces that may not support dynamic ARP.

An unnumbered interface is an IPv4 capability only used in cases where IPv4 is active (IPv4-only and mixed IPv4/IPv6 environments). When configuring an unnumbered interface, the interface specified for the unnumbered interface (system or other) must have an IPv4 address. Also, the interface type for the unnumbered interface will automatically be point-to-point. The unnumbered option can be used in IES and VPRN access interfaces, as well as in a network interface with MPLS support.

Segment Routing in Shortest Path Forwarding

Segment routing adds to IS-IS and OSPF routing protocols the ability to perform shortest path routing and source routing using the concept of abstract segment. A segment can represent a local prefix of a node, a specific adjacency of the node (interface/next-hop), a service context, or a specific explicit path over the network. For each segment, the IGP advertises an identifier referred to as Segment ID (SID).

When segment routing is used together with MPLS data plane, the SID is a standard MPLS label. A router forwarding a packet using segment routing will thus push one or more MPLS labels. This is the scope of the features described in this section.

Segment routing using MPLS labels can be used in both shortest path routing applications and in traffic engineering applications. This section focuses on the shortest path forwarding applications.

When a received IPv4 prefix SID is resolved, the Segment Routing module programs the ILM with a swap operation and also an LTN with a push operation both pointing to the primary/LFA NHLFE. An IPv4 SR tunnel to the prefix destination is also added to the TTM.

The SR tunnel in TTM is available to be used in the following contexts:

- IPv4 BGP shortcut and IPv4 BGP label route
- VLL, LDP VPLS, IES/VPRN spoke-interface, R-VPLS, BGP EVPN.
- BGP-AD VPLS, BGP-VPLS, BGP VPWS when the **use-provisioned-sdp** option is enabled in the binding to the PW template.
- Intra-AS BGP VPRN for VPN-IPv4 and VPN-IPv6 prefixes with both auto-bind and explicit SDP.
- Multicast over IES/VPRN spoke interface with spoke-sdp riding an SR tunnel.

Segment routing introduces the remote LFA feature which expands the coverage of the LFA by computing and automatically programming SR tunnels which are used as backup next-hops. The SR shortcut tunnels terminate on a remote alternate node which provides loop-free forwarding for packets of the resolved prefixes. When the **loopfree-alternate** option is enabled in an IS-IS or OSPF instance, SR tunnels are protected with an LFA backup next-hop. If the prefix of a given SR tunnel is not protected by the base LFA, the remote LFA will automatically compute a backup next-hop using an SR tunnel if the **remote-lfa** option is also enabled in the IGP instance.

Configuring Segment Routing in Shortest Path

The user enables segment routing in an IGP routing instance using the following sequence of commands.

First, the user configures the global label block, referred to as Segment Routing Global Block (SRGB), which will be reserved for assigning labels to segment routing prefix SIDs originated by this router. This range is carved from the system dynamic label range and is not instantiated by default:

configure>router>mpls-labels>sr-labels start *start-value* **end** *end-value*

Next, the user enables the context to configure segment routing parameters within a given IGP instance:

configure> router>isis>segment-routing

configure> router>ospf>segment-routing

The key parameter is the configuration of the prefix SID index range and the offset label value which this IGP instance will use. Because each prefix SID represents a network global IP address, the SID index for a prefix must be unique network-wide. Thus, all routers in the network are expected to configure and advertise the same prefix SID index range for a given IGP instance. However, the label value used by each router to represent this prefix, i.e., the label programmed in the ILM, can be local to that router by the use of an offset label, referred to as a start label:

Local Label (Prefix SID) = start-label + {SID index}

The label operation in the network becomes thus very similar to LDP when operating in the independent label distribution mode (RFC 5036) with the difference that the label value used to forward a packet to each downstream router is computed by the upstream router based on advertised prefix SID index using the above formula.

The following is an example of a router advertising its loopback address and the resulting packet label encapsulation throughout the network.

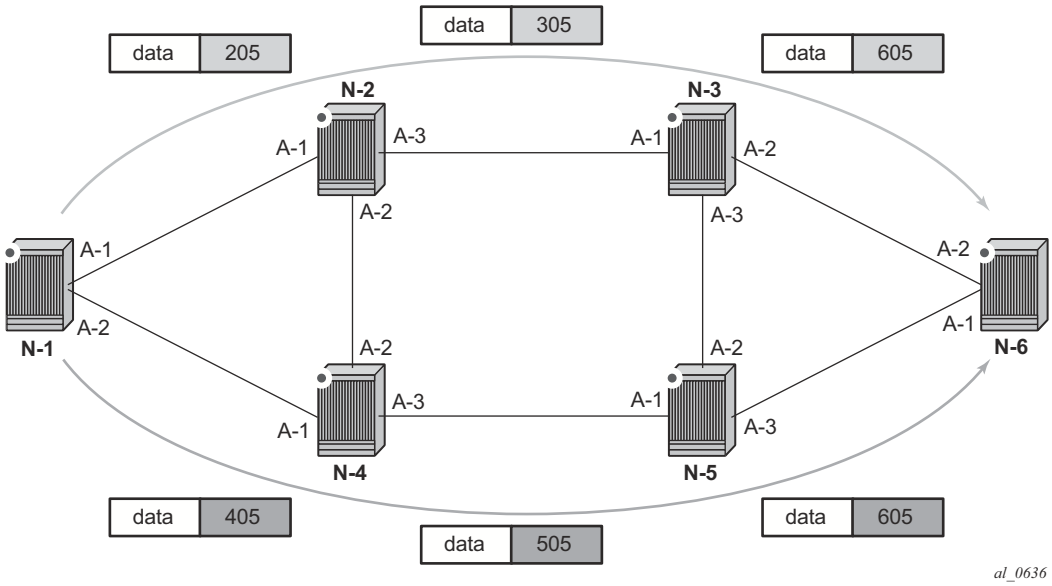

Figure 16: Packet Label Encapsulation using Segment Routing Tunnel

Router N-6 advertises loopback 10.10.10.1/32 with a prefix index of 5.Routers N-1 to N-6 are configured with the same SID index range of [1,100] and an offset label of 100 to 600 respectively. The following are the actual label values programmed by each router for the prefix of PE2:

- N-6 has a start label value of 600 and programs an ILM with label 605.
- N-3 has a start label of 300 and swaps incoming label 305 to label 605.

- N-2 has a start label of 200 and swaps incoming label 205 to label 305.

Similar operations are performed by N-4 and N-5 for the bottom path.

N-1 has an SR tunnel to N-6 with two ECMP paths. It pushes label 205 when forwarding an IP or service packet to N-6 via downstream next-hop N-2 and pushes label 405 when forwarding via downstream next-hop N-4.

The CLI for configuring the prefix SID index range and offset label value for a given IGP instance is as follows:

configure> router>isis>segment-routing>prefix-sid-range { global | start-label label-value max-index index-value}

configure> router>ospf>segment-routing>prefix-sid-range { global | start-label label-value max-index index-value}

There are two mutually-exclusive modes of operation for the prefix SID range on the router. In the global mode of operation, the user configures the global value and this IGP instance will assume the start label value is the lowest label value in the SRGB and the prefix SID index range size equal to the range size of the SRGB. Once one IGP instance selected the **global** option for the prefix SID range, all IGP instances on the system will be restricted to do the same.

The user must shutdown the segment routing context and delete the **prefix-sid-range** command in all IGP instances in order to change the SRGB. Once the SRGB is changed, the user must re-enter the **prefix-sid-range** command again. The SRGB range change will fail if an already allocated SID index/label goes out of range.

In the per-instance mode of operation, the user partitions the SRGB into non-overlapping sub-ranges among the IGP instances. The user thus configures a subset of the SRGB by specifying the start label value and the prefix SID index range size. Note that all resulting net label values (start-label + index) must be within the SRGB or the configuration will fail. Furthermore, the code checks for overlaps of the resulting net label value range across IGP instances and will strictly enforce that these ranges do not overlap.

The user must shutdown the segment routing context of an IGP instance in order to change the SID index/label range of that IGP instance using the **prefix-sid-range** command. In addition, any range change will fail if an already allocated SID index/label goes out of range.

The user can, however, change the SRGB on the fly as long as it does not reduce the current per-IGP instance SID index/label range defined with the **prefix-sid-range**. Otherwise, the user must shutdown the segment routing context of the IGP instance and delete and re-configure the **prefix-sid-range** command.

Finally, the user brings up segment routing on that IGP instances by un-shutting the context:

c**onfigure> router>isis>segment-routing>no shutdown**

configure> router>ospf>segment-routing>no shutdown

This command will fail if the user has not previously enabled the **router-capability** option in the IGP instance. Segment routing is a new capability and needs to be advertised to all routers in a given domain so that routers which support the capability will only program the node SID in the data path towards neighbors which support it.

configure> router>isis>advertise-router-capability {area|as}

configure> router>ospf>advertise-router-capability {link|area|as}

Note that the IGP segment routing extensions are area-scoped. As a consequence, the user must configure the flooding scope to **area** in OSPF and to **area** or **as** in IS-IS, otherwise performing **no shutdown** of the segment-routing node fail.

The **segment-routing** command is mutually exclusive with the **rsvp-shortcut** and **advertise-tunnel-link** options under IGP, because an SR tunnel cannot resolve to an RSVP tunnel next-hop.

Next, the user assigns a node SID index or label to the prefix representing the primary address of an IPv4 network interface of type **loopback** using one of the following commands:

- **configure> router>isis>interface>ipv4-node-sid index** *value*
- **configure> router>ospf>interface>node-sid index** *value*
- **configure> router>isis>interface>ipv4-node-sid label** *value*
- **configure> router>ospf>interface>node-sid label** *value*

Only a single node SID can be assigned to an interface. The secondary address of an IPv4 interface cannot be assigned a node SID index and does not inherit the SID of the primary IPv4 address.

Above commands should fail if the network interface is not of type loopback or if the interface is defined in an IES or a VPRN context. Also, assigning the same SID index/label value to the same interface in two different IGP instances is not allowed within the same node.

Also, for OSPF the protocol version number and the instance number dictates if the node-SID index/label is for an IPv4 or IPv6 address of the interface. Specifically, the support of address families in OSPF is as follows:

- ospfv2: always ipv4 only

The value of the label or index SID is taken from the range configured for this IGP instance. When using the global mode of operation, a new segment routing module checks that the same index or label value cannot be assigned to more than one loopback interface address. When using the per-instance mode of operation, this check is not required because the index, and thus the label ranges, of the various IGP instances are not allowed to overlap.

Segment Routing Operational Procedures

Prefix Advertisement and Resolution

Once segment routing is successfully enabled in the IS-IS or OSPF instance, the router will perform the following operations. See Control Protocol Changes for details of all TLVs and sub-TLVs for both IS-IS and OSPF protocols.

1. Advertise the Segment Routing Capability Sub-TLV to routers in all areas/levels of this IGP instance. However, only neighbors with which it established an adjacency will interpret the SID/label range information and use it for calculating the label to swap to or push for a given resolved prefix SID.
2. Advertise the assigned index for each configured node SID in the new prefix SID sub-TLV with the N-flag (node-SID flag) set. Then the segment routing module programs the incoming label map (ILM) with a pop operation for each local node SID in the data path.
3. Assign and advertise automatically an adjacency SID label for each formed adjacency over a network IP interface in the new Adjacency SID sub-TLV. Note the following points:
   - → Adjacency SID is advertised for both numbered and unnumbered network IP interface.
   - → Adjacency SID for parallel adjacencies between two IGP neighbors is not supported.
   - → Adjacency SID will not be advertised for an IES interface because access interfaces do not support MPLS.
   - → The adjacency SID must be unique per instance and per adjacency. Furthermore, ISIS MT=0 can establish an adjacency for both IPv4 and IPv6 address families over the same link and in such a case a different adjacency SID is assigned to each next-hop. However, the existing IS-IS implementation will assign a single Protect-Group ID (PG-ID) to the adjacency and as such when the state machine of a BFD session tracking the IPv4 or IPv6 next-hop times out, an action is triggered for the prefixes of both address families over that adjacency.

   The segment routing module programs the incoming label map (ILM) with a swap to an implicit null label operation, for each advertised adjacency SID.
4. Resolve received prefixes and, if a prefix SID sub-TLV exists, the Segment Routing module programs the ILM with a swap operation and an LTN with a push operation, both pointing to the primary/LFA NHLFE. An SR tunnel is also added to the TTM. Note that if a node SID resolves over an IES interface, the data path will not be programmed and a trap will be raised. Thus, only next-hops of an ECMP set corresponding to network IP interfaces are programmed in data path; next-hops corresponding to IES interfaces are not programmed. If however the user configures the interface as network on one side and IES on the other side, MPLS packets for the SR tunnel received on the access side will be dropped.
5. LSA filtering will cause SIDs not to be sent in one direction which means some node SIDs will not be resolved in parts of the network upstream of the advertisement suppression

Note that the SID/Label Binding TLV is supported in receive side and processed. It will, however, not be generated by the router.

When the user enables segment routing in a given IGP instance, the main SPF and LFA SPF are computed normally and the primary next-hop and LFA backup next-hop for a received prefix are added to RTM without the label information advertised in the prefix SID sub-TLV. In all cases, the segment routing (SR) tunnel is not added into RTM.

Error and Resource Exhaustion Handling

When the prefix corresponding to a node SID is being resolved, the following procedures are followed:

1. SR OS supports assigning different prefix-SID indexes and labels to the same prefix in different IGP instances. While other routers that receive these prefix SIDs will program a single route into RTM, based on the winning instance ID as per RTM route type preference. SR OS will add two tunnels to this destination prefix in TTM. This provides for the support of multiple topologies for the same destination prefix.

   For example: In two different instances (L2, IS-IS instance 1 and L1, IS-IS instance 2—see Figure 17), Router D has the same prefix destination, with different SIDs (SIDx and SIDy).

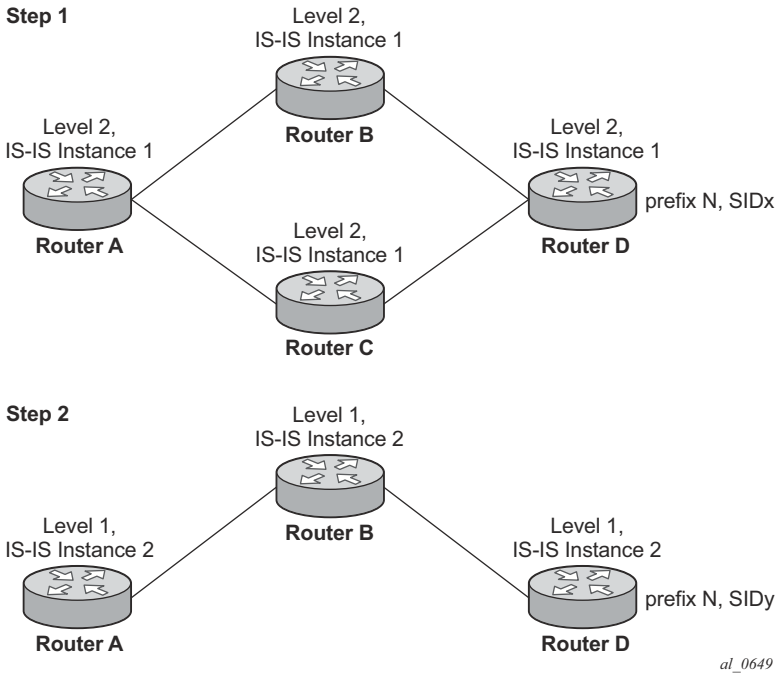

Figure 17: Programming multiple tunnels to the same destination

Assume the following route type preference in RTM and tunnel type preference in TTM are configured:

- ROUTE_PREF_ISIS_L1_INTER (RTM) 15
- ROUTE_PREF_ISIS_L2_INTER (RTM) 18
- ROUTE_PREF_ISIS_TTM 10
- Note that the TTM tunnel type preference is not used by the SR module. It is put in the TTM and will be used by other applications such a VPRN auto-bind and BGP shortcut to select a TTM tunnel.

→ Step 1: Router A performs the following resolution within the single IS-IS instance 1 , level 2

– All metrics are the same, and ECMP = 2.

– For prefix N, the RTM entry is:

– prefix N

– nhop1 = B

– nhop2 = C

– preference 18

– For prefix N, the SR tunnel TTM entry is:

– tunnel-id 1: prefix N-SIDx

– nhop1 = B

– nhop2 = C

– tunl-pref 10

→ Step 2: Add IS-IS instance 2 (Level 1) in the same setup, but in routers A, B, and C only.

– For prefix N, the RTM entry is:

– prefix N

– nhop1 = B

– preference 15

RTM will prefer L1 route over L2 route.

– For prefix N, there are two SR tunnel entries in TTM:

– SR entry for L2:

– tunnel-id 1: prefix N-SIDx

– nhop1 = B

– nhop2 = C

– tunl-pref 10

– SR entry for L1:

– tunnel-id 2: prefix N-SIDy

– nhop1 = B

– tunl-pref 10

2. While SR OS does not allow assigning the same SID index or label to different routes of the same prefix within the same IGP instance, it will resolve only one of them if received from another SR implementation and based on the RTM active route selection.
3. While SR OS does not allow assigning different SID indexes or labels to different routes of the same prefix within the same IGP instance, it will resolve only one of them if received from another SR implementation and based on the RTM active route selection.

   The selected SID will be used for ECMP resolution to all neighbors. If the route is inter-area and the conflicting SIDs are advertised by different ABRs, ECMP towards all ABRs will used the selected SID.
4. If any of the following conditions are true, the router logs a trap and an syslog error message and will not program the ILM and NHLFE for the prefix SID:
   → Received prefix SID index falls outside of the locally configured SID range.
   → one or more resolved ECMP next-hops for a received prefix SID did not advertise SR Capability sub-TLV.
   → Received prefix SID index falls outside the advertised SID range of one or more resolved ECMP next-hops.
5. Received duplicate prefix-SID index or label for different prefixes within the same IGP instance
   → Program ILM/NHLFE for the first one, log a trap and a syslog error message, and do not program the subsequent one in data path.
6. Received duplicate prefix-SID index for different prefixes across IGP instances
   → In global SID index range mode of operation, the resulting ILM label values will be the same across the IGP instances. The router programs ILM/NHLFE for the prefix of the winning IGP instance based on the RTM route type preference. The router logs a trap and a syslog error message, and does not program the subsequent prefix SIDs in data path.
   → In per-instance SID index range mode of operation, the resulting ILM label will have different values across the IGP instances. The router programs ILM/NHLFE for each prefix as expected.

7. Received duplicate prefix-SID index for the same prefix across IGP instances

   → In global SID index range mode of operation, the resulting ILM label values will be the same across the IGP instances. The router programs ILM/NHLFE for the prefix of the winning IGP instance based on the RTM route type preference. The router logs a trap and a syslog error message, and does not program the other prefix SIDs in data path.

   → In per-instance SID index range mode of operation, the resulting ILM label will have different values across the IGP instances. The router programs ILM/NHLFE for each prefix as expected.

   The behavior in the case of a global SID index range is illustrated by the IS-IS example in Figure 18.

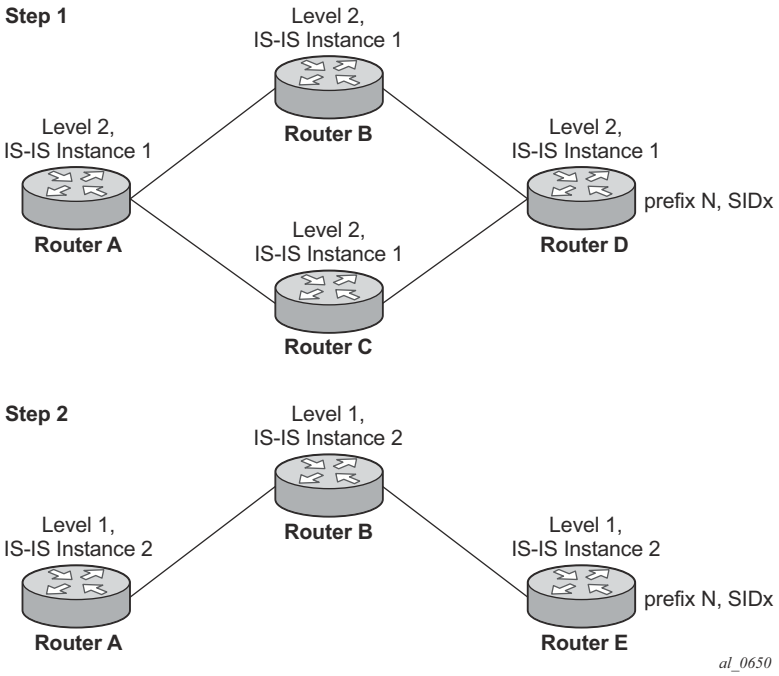

Figure 18: Handling of Same Prefix and SID in different IS-IS Instances

Assume the following route type preference in RTM and tunnel type preference in TTM are configured:

- ROUTE_PREF_ISIS_L1_INTER (RTM) 15
- ROUTE_PREF_ISIS_L2_INTER (RTM) 18
- ROUTE_PREF_ISIS_TTM 10

Note that the TTM tunnel type preference is not used by the SR module. It is put in the TTM and will be used by other applications such a VPRN auto-bind and BGP shortcut to select a TTM tunnel.

→ Step 1: Router A performs the following resolution within the single IS-IS instance 1 , level 2

- All metrics are the same, and ECMP = 2.
- For prefix N, the RTM entry is:
  - prefix N
  - nhop1 = B
  - nhop2 = C
  - preference 18
- For prefix N, the SR tunnel TTM entry is:
  - tunnel-id 1: prefix N-SIDx
  - nhop1 = B
  - nhop2 = C
  - tunl-pref 10

→ Step 2: Add IS-IS instance 2 (Level 1) in the same setup, but in routers A, B, and E only.

- For prefix N, the RTM entry is:
  - prefix N
  - nhop1 = B
  - preference 15

  RTM will prefer L1 route over L2 route.
- For prefix N, there is one SR tunnel entry in TTM:
  - SR entry for L2:
    - tunnel-id 2: prefix N-SIDx
    - nhop1 = B
    - rtm-pref 15
    - tunl-pref 10

  SR makes similar decision as RTM is based on rtm-pref 15, which is better than rtm-pref 18, so tunnel-id 2 is chosen.

8. System exhausted ILM resource while assigning an SID index/label to a local loopback interface.

   → Index allocation is failed and an error is returned in CLI. In addition, log a trap and a syslog error message.

9. System exhausted ILM, NHLFE, or any other IOM or CPM resource while resolving and programming a received prefix SID or programming a local adjacency SID.
   → The IGP instance goes into overload and a trap and syslog error message are generated. The segment routing module deletes the tunnel. The user must manually clear the IGP overload condition after freeing resources. Once IGP is brought back up, it will attempt to program at the next SPF all tunnels which previously failed the programming operation.

Segment Routing Tunnel Management

The segment routing module adds to TTM an SR tunnel entry for each resolved remote node SID prefix and programs the data path with the corresponding LTN with the push operation pointing to the primary and LFA backup NHLFEs . The LFA backup next-hop for a given prefix which was advertised with a node SID will only be computed if the **loopfree-alternate** option is enabled in the IS-IS or OSPF instance. The resulting SR tunnel which is populated in TTM will be automatically protected with FRR when an LFA backup next-hop exists for the prefix of the node SID.

With ECMP, a maximum of 32 primary next-hops (NHLFEs) are programmed for the same tunnel destination per IGP instance. ECMP and LFA next-hops are mutually exclusive as per existing implementation.

The default preference for SR tunnels in the TTM is set lower than LDP tunnels but higher than BGP tunnels to allow controlled migration of customers without disrupting their current deployment when they enable segment routing. The following is the setting of the default preference of the various tunnel types. This includes the preference of both SR tunnels based on shortest path (referred to as SR-ISIS and SR-OSPF).

The global default TTM preference for the tunnel types is as follows:

- ROUTE_PREF_RSVP 7
- ROUTE_PREF_LDP 9
- ROUTE_PREF_OSPF_TTM 10
- ROUTE_PREF_ISIS_TTM 11
- ROUTE_PREF_BGP_TTM 12
- ROUTE_PREF_GRE 255

The default value for SR-ISIS or SR-OSPF is the same regardless if one or more IS-IS or OSPF instances programmed a tunnel for the same prefix. The selection of an SR tunnel in this case will be based on lowest IGP instance-id.

The TTM is used in the case of BGP shortcuts, VPRN auto-bind, or BGP transport tunnel when the tunnel binding commands are configured to the **any** value which parses the TTM for tunnels in the protocol preference order. The user can choose to either go with the global TTM preference or list

explicitly the tunnel types they want to use. When they list the tunnel types explicitly, the TTM preference will still be used to select one type over the other. In both cases, a fallback to the next preferred tunnel type is performed if the selected one fails. Also, a reversion to a more preferred tunnel type is performed as soon as one is available. See BGP Shortcut Using Segment Routing Tunnel, BGP Label Route Resolution Using Segment Routing Tunnel, and Service Packet Forwarding with Segment Routing for the detailed service and shortcut binding CLI.

For SR-ISIS and SR-OSPF, the user can configure the preference of each specific IGP instance away from the above default values.

- **configure>router>isis>segment-routing>tunnel-table-pref preference <1..255>**
- **configure>router>ospf>segment-routing>tunnel-table-pref preference <1..255>**

The SR tunnel in TTM is available to all users of the TTM: BGP routes, VPRN auto-bind and explicit SDP binding, EVPN MPLS auto-bind, and L2 service with PW template auto-bind and with explicit SDP binding.

Local adjacency SIDs are not programmed into TTM but the remote ones can be used together with a node SID in a tunnel configuration in directed LFA feature.

Tunnel MTU Determination

The MTU of an SR tunnel populated into TTM is determined as in the case of an IGP tunnel; for example, LDP LSP, based on the outgoing interface MTU minus the label stack size. Segment routing, however, supports remote LFA which programs an LFA backup next-hop that adds another label to the tunnel for a total of two. Finally, directed LFA, if implemented by other routers in the network, can push additional labels but most of the common topologies will not exceed a total of three labels.

Based on the above, the user is provided with a CLI to configure the MTU of all SR tunnels within each IGP instance:

configure> router>isis (ospf)>segment-routing>tunnel-mtu *bytes*

There is no default value for this new command. If the user does not configure an SR tunnel MTU, the MTU will be fully determined by IGP as explained below.

The MTU of the SR tunnel is then determined as follows:

SR_Tunnel_MTU = MIN {Cfg_SR_MTU, IGP_Tunnel_MTU- 3 labels}

Where,

- *Cfg_SR_MTU* is the MTU configured by the user for all SR tunnels within a given IGP instance using the above CLI. If no value was configured by the user, the SR tunnel MTU will be fully determined by the IGP interface calculation explained next.

- *IGP_Tunnel_MTU* is the minimum of the IS-IS or OSPF interface MTU among all the ECMP paths or among the primary and LFA backup paths of this SR tunnel.

The SR tunnel MTU is dynamically updated anytime any of the above parameters used in its calculation changes. This includes when the set of the tunnel next-hops changes or the user changes the configured SR MTU or interface MTU value.

Remote LFA with Segment Routing

The user enables the remote LFA next-hop calculation by the IGP LFA SPF by appending the following new option in the existing command which enables LFA calculation:

- **configure> router>isis>loopfree-alternate remote-lfa**
- **configure> router>ospf>loopfree-alternate remote-lfa**

SPF performs the remote LFA additional computation following the regular LFA next-hop calculation when both of the following conditions are met:

- The remote-lfa option is enabled in an IGP instance.
- The LFA next hop calculation did not result in protection for one or more prefixes resolved to a given interface.

Remote LFA extends the protection coverage of LFA-FRR to any topology by automatically computing and establishing/tearing-down shortcut tunnels, also referred to as repair tunnels, to a remote LFA node which puts the packets back into the shortest without looping them back to the node which forwarded them over the repair tunnel. A repair tunnel can in theory be an RSVP LSP, an LDP-in-LDP tunnel, or an SR tunnel. In SR OS, this feature is restricted to use an SR repair tunnel to the remote LFA node.

The remote LFA algorithm for link protection is described in RFC 7490. Unlike the regular LFA calculation, which is calculated per prefix, the LFA algorithm for link protection is a per-link LFA SPF calculation. As such, it provides protection for all destination prefixes which share the protected link by using the neighbor on the other side of the protected link as a proxy for all these destinations. Assume the topology in Figure 19.

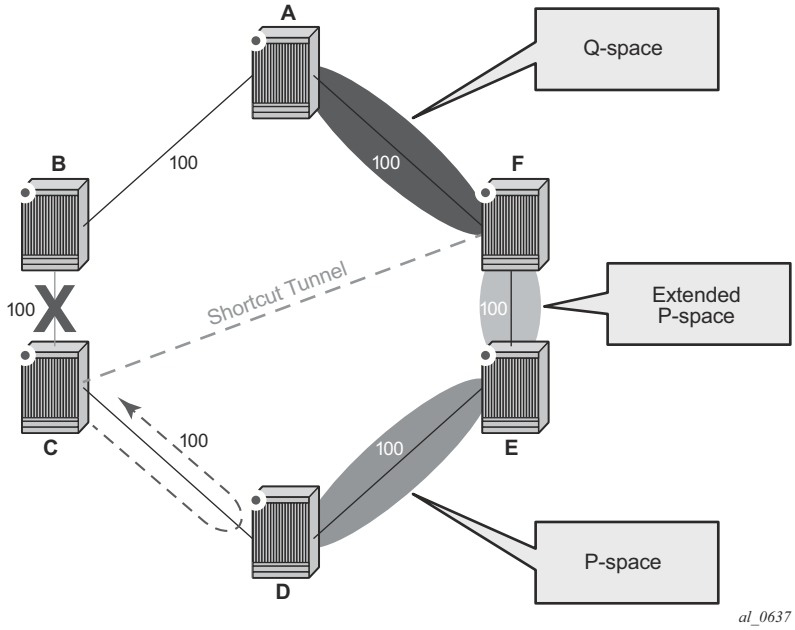

Figure 19: Remote LFA Algorithm

When the LFA SPF in node C computes the per-prefix LFA next-hop, prefixes which use link C-B as the primary next-hop will have no LFA next-hop due to the ring topology. If node C used node link C-D as a back-up next-hop, node D would loop a packet back to node C. The remote LFA then runs the following algorithm, referred to as the "PQ Algorithm" in RFC 7490:

1. Compute the extended P space of Node C with respect to link C-B: set of nodes reachable from node C without any path transiting the protected link (link C-B). This yields nodes D, E, and F.

   The determination of the extended P space by node C uses the same computation as the regular LFA by running SPF on behalf of each of the neighbors of C.

   Note that the RFC 7490 initially introduced the concept of P space, which would have excluded node F because, from the node C perspective, node C has a couple of ECMP paths, one of which goes via link C-B. However, because the remote LFA next-hop is activated when link C-B fails, this rule can be relaxed and node F can be included, which then yields the extended P space.

The user can limit the search for candidate P nodes to reduce the amount of SPF calculations in topologies where many eligible P nodes can exist. A CLI command is provided to configure the maximum IGP cost from node C for a P node to be eligible:

→ **configure> router>isis>loopfree-alternate remote-lfa max-pq-cost** *value*

→ **configure> router>ospf>loopfree-alternate remote-lfa max-pq-cost** *value*

2. Compute the Q space of node B with respect to link C-B: set of nodes from which the destination proxy (node B) can be reached without any path transiting the protected link (link C-B).

   The Q space calculation is effectively a reverse SPF on node B. In general, one reverse SPF is run on behalf of each of C neighbors to protect all destinations resolving over the link to the neighbor. This yields nodes F and A in the example of Figure 19.

   The user can limit the search for candidate Q nodes to reduce the amount of SPF calculations in topologies where many eligible Q nodes can exist. The CLI above is also used to configure the maximum IGP cost from node C for a Q node to be eligible.

3. Select the best alternate node: this is the intersection of extended P and Q spaces. The best alternate node or PQ node is node F in the example of Figure 19. From node F onwards, traffic follows the IGP shortest path.

   If many PQ nodes exist, the lowest IGP cost from node C is used to narrow down the selection and if more than one PQ node remains, the node with lowest router-id is selected.

The details of the label stack encoding when the packet is forwarded over the remote LFA next-hop is shown in Figure 20.

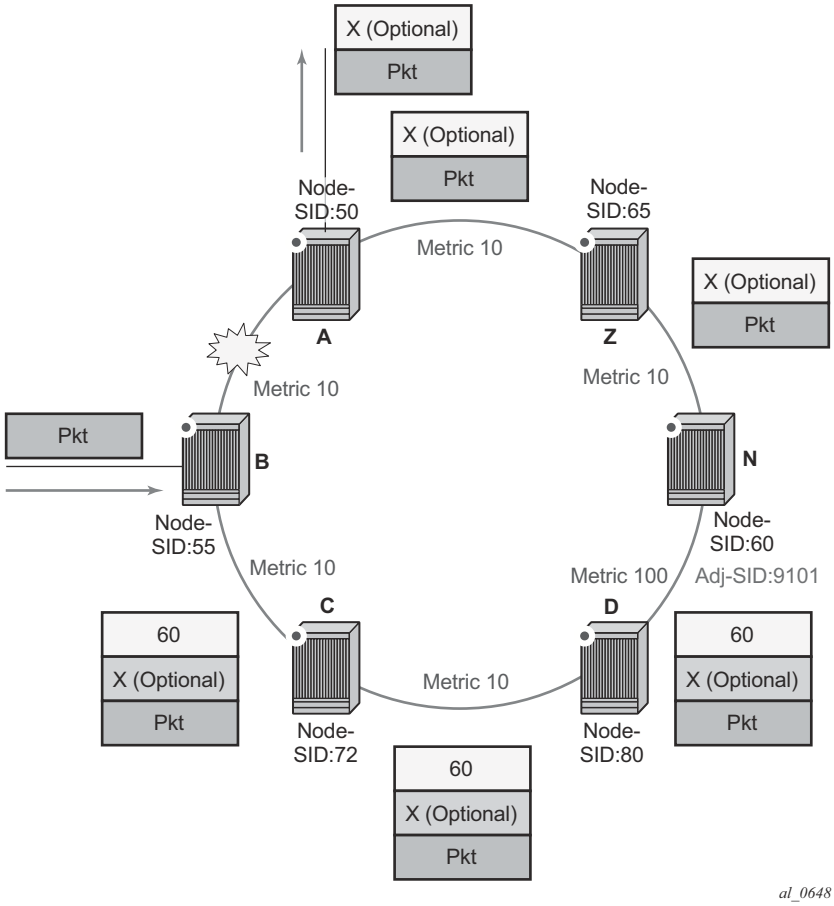

Figure 20: Remote LFA Next-Hop in Segment Routing

The label corresponding to the node SID of the PQ node is pushed on top of the original label of the SID of the resolved destination prefix. If node C has resolved multiple node SIDs corresponding to different prefixes of the selected PQ node, it will push the lowest node SID label on the packet when forwarded over the remote LFA backup next-hop.

If the PQ node is also the advertising router for the resolved prefix, the label stack is compressed in some cases depending on the IGP:

- In IS-IS, the label stack is always reduced to a single label, which is the label of the resolved prefix owned by the PQ node.
- In OSPF, the label stack is reduced to the single label of the resolved prefix when the PQ node advertised a single node SID in this OSPF instance. If the PQ node advertised a node SID for multiple of its loopback interfaces within this same OSPF instance, the label stack

is reduced to a single label only in the case where the SID of the resolved prefix is the lowest SID value.

The following rules and limitations apply to the remote LFA implementation:

- LFA policy is currently supported for IP next-hops only. It is not supported with tunnel next-hops when IGP shortcuts are used for LFA backup. Remote LFA is also a tunnel next-hop and, as such, a user configured LFA policy will not be applied in the selection of a remote LFA backup next-hop when multiple candidates are available.
- As a result, if an LFA policy is applied and does not find an LFA IP next-hop for a set of prefixes, the remote LFA SPF will be run to search for a remote LFA next-hop for the same prefixes. The selected remote LFA next-hops, if found, may not satisfy the LFA policy constraints.
- If the user excludes a network IP interface from being used as an LFA next-hop using the CLI command **loopfree-alternate-exclude** under the interface's IS-IS or OSPF context, the interface will also be excluded from being used as the outgoing interface for a remote LFA tunnel next-hop.
- As with the regular LFA algorithm, the remote LFA algorithm will compute a backup next-hop to the ABR advertising an inter-area prefix and not to the destination prefix itself.

Data Path Support

A packet received with a label matching either a node SID or an adjacency SID will be forwarded according to the ILM type and operation, as described in Table 8.

Table 8: Data Path Support

Label type	Operation
Top label is a local node SID	Label is popped and the packet is further processed. If the popped node SID label is the bottom of stack label, the IP packet is looked up and forwarded in the appropriate FIB.
Top or next label is a remote node SID	Label is swapped to the calculated label value for the next-hop and forwarded according to the primary or backup NHLFE. With ECMP, a maximum of 32 primary next-hops (NHLFEs) are programmed for the same destination prefix and for each IGP instance. ECMP and LFA next-hops are mutually exclusive as per existing implementation.

Table 8: Data Path Support

Label type	Operation
Top or next label is an adjacency SID	Label is popped and the packet is forwarded out on the interface to the next-hop associated with this adjacency SID label. In effect, the data path operation is modeled like a swap to an implicit-null label instead of a pop.
Next label is BGP 3107 label	The packet is further processed according to the ILM operation as in current implementation. • The BGP label may be popped and the packet looked up in the appropriate FIB. • The BGP label may be swapped to another BGP label. • The BGP label may be stitched to an LDP label.
Next label is a service label	The packet is looked up and forwarded in the Layer 2 or VPRN FIB as in current implementation.

A router forwarding an IP or a service packet over an SR tunnel pushes a maximum of three transport labels with a remote LFA next-hop, and four transport labels when the P and Q nodes are at most one hop away from each other (directed LFA if implemented by Node B). This is illustrated in Figure 21.

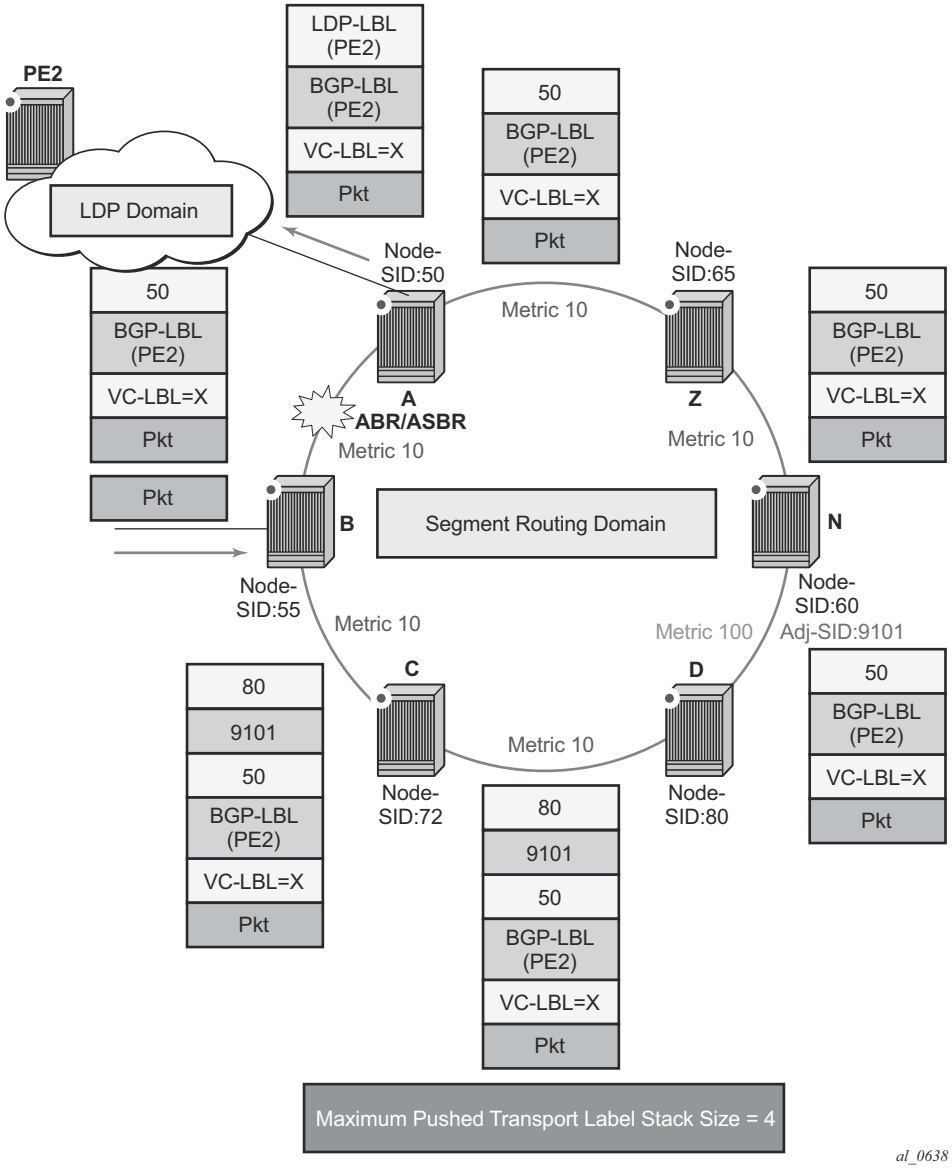

Figure 21: Maximum Pushed Transport Label Stack in Shortest Path Forwarding with Segment Routing

Assume that a VPRN service in node B forwards a packet received on a SAP to a destination VPN-IPv4 prefix X advertised by a remote PE2 via ASBR/ABR node A. Router B is in a segment routing domain while PE2 is in an LDP domain. BGP label routes are used to distribute the PE /32 loopbacks between the two domains.

When node B forwards over the primary next-hop for prefix X, it pushes the node SID of the ASBR followed by the BGP 3107 label of PE2, followed by the service label for prefix X. When the directed LFA next-hop is activated, node B pushes a couple or more segment routing labels: the node SID for the remote LFA backup node (node D) and the adjacency SID for the next-hop to node N via link D-N.

When node D receives the packet while the directed LFA next-hop is activated, it pops the top segment routing label which corresponds to a local node SID. It then examines the ILM for the next label which corresponds to the adjacency SID. This results in popping this label and forcing the forwarding of the packet over link D-N.

When the ABR/ASBR node receives the packet from either node B or node Z, it pops the segment routing label which corresponds to a local node SID, then swaps the BGP label and pushes the LDP label of PE2 which is the next-hop of the BGP label route.

Hash label support

When the **hash-label** option is enabled in a service context, the insertion of the hash label into the bottom of the label stack of a packet forwarded using segment routing transport tunnel is performed.

Control Protocol Changes

This section describes both IS-IS Control Protocol Changes and OSPF Control Protocol Changes.

IS-IS Control Protocol Changes

New TLV/sub-TLVs are defined in *draft-ietf-isis-segment-routing-extensions* and are supported in the implementation of segment routing in IS-IS. Specifically:

- the prefix SID sub-TLV
- the adjacency SID sub-TLV
- the SID/Label Binding TLV
- SR-Capabilities Sub-TLV
- SR-Algorithm Sub-TLV

This section describes the behaviors and limitations of the IS-IS support of segment routing TLV and sub-TLVs.

SR OS supports advertising the IS router capability TLV (RFC 4971) only for topology MT=0. As a result, the segment routing capability Sub-TLV can only be advertised in MT=0 which restricts the segment routing feature to MT=0.

Similarly, if prefix SID sub-TLVs for the same prefix are received in different MT numbers of the same IS-IS instance, then only the one in MT=0 will be resolved as long as there is a IP Reachability TLV received for same prefix. When the prefix SID index is also duplicated, an error is logged and a trap is generated, as explained in Error and Resource Exhaustion Handling.

I and V flags are set to 1 and 0 respectively when originating the SR capability sub-TLV but are not checked when the sub-TLV is received. Only the SRGB range is processed.

The algorithm field is set to 0, meaning Shortest Path First (SPF) algorithm based on link metric, when originating the SR-Algorithm capability sub-TLV but is not checked when the sub-TLV is received.

Only IPv4 prefix and adjacency SID sub-TLVs will be originated within MT=0. IPv6 prefix and adjacency SID sub-TLVs can however be received and ignored. The user can get a dump of the octets of the received but not-supported sub-TLVs using the existing **show** command.

SR OS originates a single prefix SID sub-TLV per IS-IS IP reachability TLV and processes the first prefix SID sub-TLV only if multiple are received within the same IS-IS IP reachability TLV.

SR OS encodes the 32 bit index in the prefix SID sub-TLV. The 24 bit label is not supported.

SR OS originates a prefix SID sub-TLV with the following encoding of the flags.

- The R-flag is set if the prefix SID sub-TLV, along with its corresponding IP reachability TLV, is propagated between levels. See below for more details about prefix propagation.
- The N-flag is always set because SR OS supports prefix SID of type node SID only.

- The P-Flag (no-PHP flag) is always set, meaning that the label for the prefix SID will be pushed by the PHP router when forwarding to this router. SR OS PHP router will process properly a received prefix SID with the P-flag set to zero and will use implicit-null for the outgoing label towards the router which advertised it as long as the P-Flag is also set to 1.
- The E-flag (Explicit-Null flag) is always set to zero. An SR OS PHP router will, however, process properly a received prefix SID with the E-flag set to 1 and, when the P-flag is also set to 1, it will push explicit-null for the outgoing label towards the router which advertised it.
- The V-flag is always set to 0 to indicate an index value for the SID.
- The L-flag is always set to 0 to indicate that the SID index value is not locally significant.
- The algorithm field is always set to zero to indicate Shortest Path First (SPF) algorithm based on link metric and is not checked on a received prefix SID sub-TLV.

SR OS will still resolve a prefix SID sub-TLV received without the N-flag set but with the prefix length equal to 32. A trap, however, is raised by IS-IS.

SR OS will not resolve a prefix SID sub-TLV received with the N flag set and a prefix length different than 32. A trap is raised by IS-IS.

SR OS resolves a prefix SID received within a IP reachability TLV based on the following route preference:

- SID received via L1 in a prefix SID sub-TLV part of IP reachability TLV
- SID received via L2 in a prefix SID sub-TLV part of IP reachability TLV

A prefix received in an IP reachability TLV is propagated, along with the prefix SID sub-TLV, by default from L1 to L2 by an L1L2 router. A router in L2 will set up an SR tunnel to the L1 router via the L1L2 router, which acts as an LSR.

A prefix received in an IP reachability TLV is not propagated, along with the prefix SID sub-TLV, by default from L2 to L1 by an L1L2 router. If the user adds a policy to propagate the received prefix, then a router in L1 will set up an SR tunnel to the L2 router via the L1L2 router, which acts as an LSR.

If a prefix is summarized by an ABR, the prefix SID sub-TLV will not be propagated with the summarized route between levels. To propagate the node SID for a /32 prefix, route summarization must be disabled.

SR OS does not propagate into IS-IS the prefix SID sub-TLV of external routes. Thus, when the corresponding prefix is redistributed from another protocol such as OSPF, the prefix SID is removed. SR OS will, however, accept the prefix SID sub-TLV of an external route if received from another router and will process it the same way as for an internal route.

SR OS originates an adjacency SID sub-TLV with the following encoding of the flags:

- the F-flag is set to zero to indicate the IPv4 family for the adjacency encapsulation.
- the B-Flag is set to zero and is not processed on receipt
- the V-flag is always set to 1
- the L-flag is always set to 1
- the S-flag is set to zero as assigning adjacency SID to parallel links between neighbors is not supported. An adjacency received SID with S-Flag set will not be processed.
- the weight octet is not supported and is set to all zeros.

SR OS does not originate the SID/Label Binding TLV but can process it properly if received. Note the following rules and limitations:

- Only the Mapping Server Prefix-SID Sub-TLV within the TLV is processed and the ILMs installed if the prefixes in the provided range are resolved.
- The range and FEC prefix fields are processed. Each FEC prefix is resolved normally, as for the prefix SID sub-TLV, meaning there must be an IP Reachability TLV received for the exact matching prefix.
- If the same prefix is advertised with both a prefix SID sub-TLV and a mapping server Prefix-SID sub-TLV. The resolution follows the following route preference:
  - → SID received via L1 in a mapping server Prefix-SID sub-TLV
  - → SID received via L2 in a mapping server Prefix-SID sub-TLV
  - → SID received via L1 in a prefix SID sub-TLV part of IP reachability TLV
  - → SID received via L2 in a prefix SID sub-TLV part of IP reachability TLV
- No leaking of the entire TLV is performed between levels. However, an L1L2 router will propagate the prefix-SID sub-TLV from the SID/Label binding TLV (received from a mapping server) into the IP Reachability TLV if the latter is propagated between levels.
- The mapping server which advertised the SID/Label Binding TLV does not need to be in the shortest path for the FEC prefix.
- If the same FEC prefix is advertised in multiple binding TLVs by different routers, the SID in the binding TLV of the first router which is reachable will be used. If that router becomes unreachable, the next reachable one will be used.
- No check is performed if the content of the binding TLVs from different mapping servers are consistent or not.
- Any other sub-TLV, for example, the SID/Label Sub-TLV, ERO metric and unnumbered interface ID ERO, will be ignored but the user can get a dump of the octets of the received but not-supported sub-TLVs using the existing IGP **show** command.

OSPF Control Protocol Changes

New TLV/sub-TLVs are defined in *draft-ietf-ospf-segment-routing-extensions-04* and are required for the implementation of segment routing in OSPF. Specifically:

- the prefix SID sub-TLV part of the OSPFv2 Extended Prefix TLV
- the prefix SID sub-TLVpart of the OSPFv2 Extended Range Prefix TLV
- the adjacency SID sub-TLV part of the OSPFv2 Extended Link TLV
- SID/Label Range Capability TLV
- SR-Algorithm Capability TLV

This section describes the behaviors and limitations of the OSPF support of segment routing TLV and sub-TLVs.

SR OS currently supports IPv4 prefix and adjacency SIDs within OSPFv2 instances.

SR OS originates a single prefix SID sub-TLV per OSPFv2 Extended Prefix TLV and processes the first one only if multiple prefix SID sub-TLVs are received within the same OSPFv2 Extended Prefix TLV.

SR OS encodes the 32-bit index in the prefix SID sub-TLV. The 24-bit label or variable IPv6 SID is not supported.

SR OS originates a prefix SID sub-TLV with the following encoding of the flags.

- The NP-Flag is always set, meaning that the label for the prefix SID will be pushed by the PHP router when forwarding to this router. The SR OS PHP router will properly process a received prefix SID with the NP-flag set to zero and will use implicit-null for the outgoing label towards the router which advertised it.
- The M-Flag is always unset because SR OS does not support originating a mapping server prefix-SID sub-TLV.
- The E-flag is always set to zero. The SR OS PHP router will properly process a received prefix SID with the E-flag set to 1, and when the NP-flag is also set to 1 it will push explicit-null for the outgoing label towards the router which advertised it.
- The V-flag is always set to 0 to indicate an index value for the SID.
- The L-flag is always set to 0 to indicate that the SID index value is not locally significant.
- The algorithm field is always set to zero to indicate Shortest Path First (SPF) algorithm based on link metric and is not checked on a received prefix SID sub-TLV.

SR OS resolves a prefix SID received within an Extended Prefix TLV based on the following route preference:

- SID received via an intra-area route in a prefix SID sub-TLV part of Extended Prefix TLV

- SID received via an inter-area route in a prefix SID sub-TLV part of Extended Prefix TLV

SR OS originates an adjacency SID sub-TLV with the following encoding of the flags.

- The F-flag is unset to indicate the Adjacency SID refers to an adjacency with outgoing IPv4 encapsulation.
- The B-flag is set to zero and is not processed on receipt.
- The V-flag is always set.
- The L-flag is always set.
- The S-flag is not supported.
- The weight octet is not supported and is set to all zeros.

SR OS does not originate the OSPFv2 Extended Range Prefix TLV but can process it properly if received. Note the following rules and limitations:

- Only the prefix SID sub-TLV within the TLV is processed and the ILMs installed if the prefixes are resolved.
- The range and address prefix fields are processed. Each prefix is resolved normally as for a prefix SID received within the Extended Prefix TLV, meaning there must be an Extended Prefix TLV received for the exact matching prefix.
- If the same prefix is advertised with both a prefix SID sub-TLV in a IP reachability TLV and a mapping server Prefix-SID sub-TLV, the resolution follows the following route preference:
  - → the SID received via an intra-area route in a prefix SID sub-TLV part of Extended Prefix TLV
  - → the SID received via an inter-area route in a prefix SID sub-TLV part of Extended Prefix TLV
  - → the SID received via intra-area route in a mapping server Prefix-SID sub-TLV
  - → the SID received via a inter-area route in a mapping server Prefix-SID sub-TLV
- No leaking of the entire TLV is performed between levels. However, an ABR will propagate the prefix-SID sub-TLV from the Extended Prefix Range TLV (received from a mapping server) into an Extended Prefix TLV if the latter is propagated between areas.
- The mapping server which advertised the OSPFv2 Extended Range Prefix TLV does not need to be in the shortest path for the FEC prefix.
- If the same FEC prefix is advertised in multiple OSPFv2 Extended Range Prefix TLVs by different routers, the SID in the TLV of the first router which is reachable will be used. If that router becomes unreachable, the next reachable one will be used.
- No check is performed to determine whether or not the contents of the OSPFv2 Extended Range Prefix TLVs received from different mapping servers are consistent.

- Any other sub-TLV, for example, the ERO metric and unnumbered interface ID ERO, will be ignored but the user can get a dump of the octets of the received but not-supported sub-TLVs using the existing IGP **show** command.

SR OS supports propagation on ABR of external prefix LSA into other areas with routeType set to 3 as per Section 6.2 of *draft-ietf-ospf-segment-routing-extensions-04*.

SR OS supports propagation on ABR of external prefix LSA with route type 7 from NSSA area into other areas with route type set to 5 as per Section 6.3 of *draft-ietf-ospf-segment-routing-extensions-04*.

When the user configures an OSPF import policy, the outcome of the policy applies to prefixes resolved in RTM and the corresponding tunnels in TTM. So, a prefix removed by the policy will not appear as both a route in RTM and as an SR tunnel in TTM.

The algorithm field is set to 0 (Shortest Path First (SPF) algorithm based on link metric) when originating the SR-Algorithm capability TLV but is not checked when the sub-TLV is received.

BGP Shortcut Using Segment Routing Tunnel

The user enables the resolution of IPv4 prefixes using SR tunnels to BGP next-hops in TTM with the following command:

```
configure>router> bgp>next-hop-resolution
            shortcut-tunnel
                  [no] family {ipv4}
                        resolution {any|disabled|filter}
                        resolution-filter
                              [no] sr-isis
                              [no] sr-ospf
                        [no] disallow-igp
                        exit
                  exit
            exit
```

When **resolution** is set to **any**, any supported tunnel type in BGP shortcut context will be selected following TTM preference. The following tunnel types are supported in a BGP shortcut context and in order of preference: RSVP, LDP, Segment Routing and BGP.

When the **sr-isis** or **sr-ospf** value is enabled, an SR tunnel to the BGP next-hop is selected in the TTM from the lowest preference ISIS or OSPF instance. If many instances have the same lowest preference from the lowest numbered IS-IS or OSPF instance

See the BGP chapter for more details.

BGP Label Route Resolution Using Segment Routing Tunnel

The user enables the resolution of RFC 3107 BGP label route prefixes using SR tunnels to BGP next-hops in TTM with the following command:

```
configure>router>bgp>next-hop-resolution>
            label-route-transport-tunnel
                  [no] family {ipv4, vpn}
                        resolution {any|disabled|filter}
                        resolution-filter
                              [no] sr-isis
                              [no] sr-ospf
                        exit
                  exit
            exit
```

When the **resolution** option is explicitly set to **disabled**, the default binding to LDP tunnel resumes. If **resolution** is set to **any**, any supported tunnel type in BGP label route context will be selected following TTM preference.

The following tunnel types are supported in a BGP label route context and in order of preference: RSVP, LDP, and Segment Routing.

When the **sr-isis** or **sr-ospf** is specified using the **resolution-filter** option, a tunnel to the BGP next-hop is selected in the TTM from the lowest numbered ISIS or OSPF instance.

Refer to the BGP section of the SR OS Routing Protocols Guide for more details.

Service Packet Forwarding with Segment Routing

A couple of new SDP subtypes of the MPLS type are added to allow service binding to an SR tunnel programmed in TTM by OSPF or IS-IS:

***A:7950 XRS-20# configure service sdp 100 mpls create**

***A:7950 XRS-20>config>service>sdp$ sr-ospf**

***A:7950 XRS-20>config>service>sdp$ sr-isis**

The SDP of type **sr-isis** or **sr-ospf** can be used with the **far-end** option. When the **sr-isis** or **sr-ospf** value is enabled, a tunnel to the far-end address is selected in the TTM from the lowest preference ISIS or OSPF instance. If many instances have the same lowest preference from the lowest numbered IS-IS or OSPF instance

The **tunnel-far-end** option is not supported. In addition, the **mixed-lsp-mode** option does not support the **sr-isis** and **sr-ospf** tunnel types.

The signaling protocol for the service labels for an SDP using an SR tunnel can be configured to static (**off**), T-LDP (**tldp**), or BGP (**bgp**).

SR tunnels can be used in VPRN and BGP EVPN with the **auto-bind-tunnel** command. See Next-hop Resolution Using Tunnels for more information.

Both VPN-IPv4 and VPN-IPv6 (6VPE) are supported in a VPRN or BGP EVPN service using segment routing transport tunnels with the **auto-bind-tunnel** command.

Refer to the *SR OS Layer 3 Services Guide* and the BGP chapter of the *SR OS Routing Protocols Guide* for more details of the VPRN **auto-bind-tunnel** CLI command.

The following are the service contexts which are supported with SR tunnels:

- VLL, LDP VPLS, IES/VPRN spoke-interface, R-VPLS, BGP EVPN.
- BGP-AD VPLS, BGP-VPLS, BGP VPWS when the use-provisioned-sdp option is enabled in the binding to the PW template.
- Intra-AS BGP VPRN for VPN-IPv4 and VPN-IPv6 prefixes with both auto-bind and explicit SDP.
- Multicast over IES/VPRN spoke interface with spoke-sdp riding an SR tunnel.

The following service contexts are not supported:

- Inter-AS VPRN.
- Dynamic MS-PW, PW-switching.
- BGP-AD VPLS, BGP-VPLS, BGP VPWS with auto-generation of SDP using an SR tunnel when binding to a PW template.

Mirror Services and Lawful Intercept

The user can configure a spoke-SDP bound to an SR tunnel to forward mirrored packets from a mirror source to a remote mirror destination. In the configuration of the mirror destination service at the destination node, the **remote-source** command must use a spoke-sdp with VC-ID which matches the one the user configured in the mirror destination service at the mirror source node. The far-end option is not supported with an SR tunnel.

This also applies to the configuration of the mirror destination for an LI source.

Configuration at mirror source node:

```
config mirror mirror-dest 10

            no spoke-sdp <sdp-id:vc-id>
            spoke-sdp <sdp-id:vc-id> [create]
               egress
                  vc-label <egress-vc-label>
```

Notes:

- *sdp-id* matches an SDP which uses an SR tunnel
- for vc-label, both static and t-ldp egress vc labels are supported

Configuration at mirror destination node:

```
*A:7950 XRS-20# configure mirror mirror-dest 10 remote-source

            - spoke-sdp <SDP-ID>:<VC-ID> create <-- VC-ID matching that
               of spoke-sdp configured in mirror destination context at
               mirror source node.
                  ingress
                     vc-label <ingress-vc-label> <--- optional: both
                     static and t-ldp ingress vc label are supported.
                  exit
                  no shutdown
               exit
            exit
```

Notes:

- the **far-end** command is not supported with SR tunnel at mirror destination node; user must reference a spoke-SDP using a segment routing SDP coming from mirror source node:
  - → **far-end** *ip-address* [**vc-id** *vc-id*] [**ing-svc-label** *ingress-vc-label* | **tldp**] [**icb**]
  - → **no far-end** *ip-address*

- for vc-label, both static and t-ldp ingress vc labels are supported

Mirroring and LI are also supported with the PW redundancy feature when the endpoint spoke-sdp, including the ICB, is using an SR tunnel. Routable Lawful Intercept Encapsulation (**config>mirror>mirror-dest>encap**# **layer-3-encap**) when the remote L3 destination is reachable over an SR tunnel is also supported.

FIB Prioritization

The RIB processing of specific routes can be prioritized through the use of the **rib-priority command**. This command allows specific routes to be prioritized through the protocol processing so that updates are propagated to the FIB as quickly as possible.

The **rib-priority** command is configured within the global IS-IS routing context, and the administrator has the option to either specify a prefix-list or an IS-IS tag value. If a prefix list is specified than route prefixes matching any of the prefix list criteria will be considered high priority. If instead an IS-IS tag value is specified then any IS-IS route with that tag value will be considered high priority.

The routes that have been designated as high priority will be the first routes processed and then passed to the FIB update process so that the forwarding engine can be updated. All known high priority routes should be processed before the IS-IS routing protocol moves on to other standard priority routes. This feature will have the most impact when there are a large number of routes being learned through the IS-IS routing protocols.

IS-IS Configuration Process Overview

Figure 22 displays the process to provision basic IS-IS parameters.

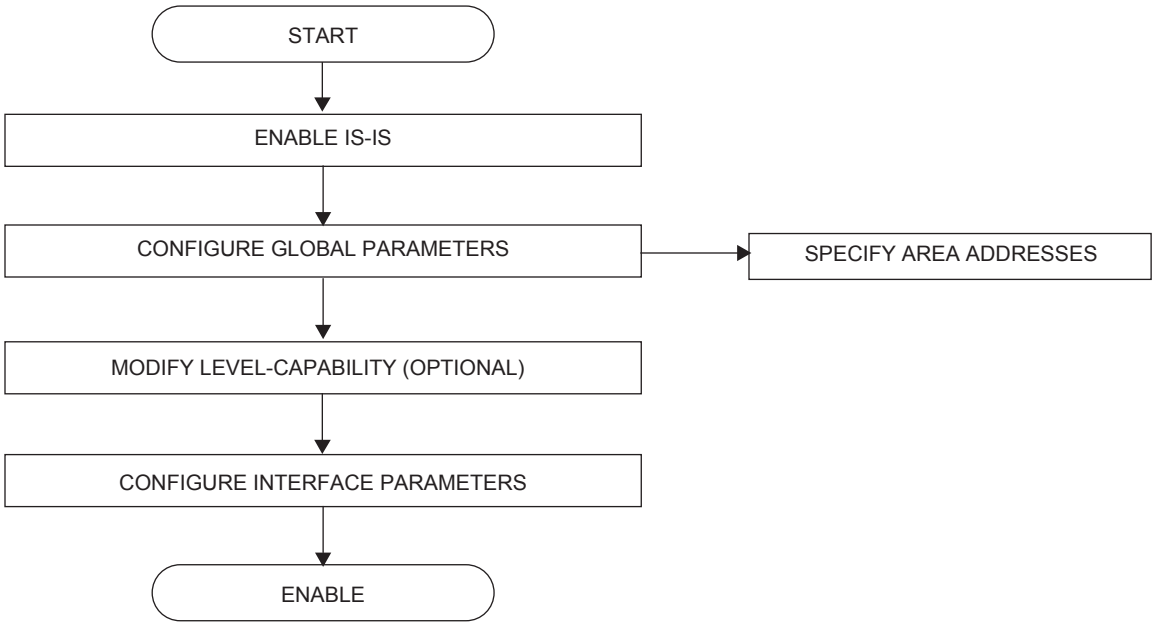

Figure 22: IS-IS Configuration and Implementation Flow

Configuration Notes

This section describes IS-IS configuration caveats.

General

- IS-IS must be enabled on each participating router.
- There are no default network entity titles.
- There are no default interfaces.
- By default, the routers are assigned a Level 1/Level 2 level capability.