

Chapter 5

Configure MPLS-Signaled LSPs

To configure Multiprotocol Label Switching (MPLS)-signaled label-switched paths (LSPs), you create an LSP that runs from the ingress router to the egress router. (For information about Label Distribution Protocol [LDP]-signaled LSPs, see “Configure LDP” on page 197.) To create the LSP, you configure only the ingress router; you do not have to configure any other routers. You can configure the LSP so that the JUNOS software makes all forwarding decisions, or you can configure some or all routers in the path. The LSP is set up by the Resource Reservation Protocol (RSVP), through RSVP signaling messages. The JUNOS software automatically negotiates, assigns, releases, and reuses labels. Automatically assigned labels have a value from 99,999 through 1,048,575.

To configure signaled LSPs across a network, perform the following tasks:

- Configure the Ingress Router for Signaled LSPs on page 45

- Configure All Other MPLS Routers for Signaled LSPs on page 78

- Enable RSVP on page 79

These sections provide information about special features related to signaled LSPs:

- Improve TED Accuracy with RSVP PathErr Messages on page 79

- Configure MPLS over GRE Tunnels on page 84

- Configure IPv6 Tunnels over MPLS on page 86

- LSP Attributes for GMPLS on page 89

For configuration examples, see “Examples: Configure Signaled LSPs” on page 81.

Configure the Ingress Router for Signaled LSPs

To configure signaled LSPs, perform the following tasks on the ingress router:

- Create a Named Path on page 46

- Create an LSP on page 48

- Configure Alternate Backup Paths Using Fate-Sharing on page 76

Create a Named Path

To configure signaled LSPs, you must first create one or more named paths on the ingress router. For each path, you can specify some or all transit routers in the path, or you can leave it empty.

Each path name can contain up to 32 characters and can include letters, digits, periods, and hyphens. The name must be unique within the ingress router. Once a named path is created, you can use the named path with the `primary` or `secondary` statement to configure LSPs at the `[edit protocols mpls label-switched-path label-path-name]` hierarchy level. You can specify the same named path on any number of LSPs.

To determine whether an LSP is associated with the primary or secondary path in an RSVP session, issue the `show rsvp session detail` command. For more information, see the *JUNOS Internet Software Operational Mode Command Reference: Protocols, Class of Service, Chassis, and Management*.

To create an empty path, create a named path by including the following form of the path statement at the `[edit protocols mpls]` hierarchy level. This form of the path statement is empty, which means that any path between the ingress and egress routers is accepted. In actuality, the path used tends to be the same path as is followed by destination-based, best-effort traffic.

```
[edit protocols mpls]
path path-name;
```

To create a path in which you specify some or all transit routers in the path, include the following form of the path statement at the `[edit protocols mpls]` hierarchy level, specifying one address for each transit router:

```
[edit protocols mpls]
path path-name {
  address | host name <strict | loose>;
}
```

In this form of the path statement, you specify one or more transit router addresses. Specifying the ingress and/or egress routers is optional. You can specify the address or host name of each transit router, although you do not need to list each transit router if its type is loose. Specify the addresses in order, starting with the ingress router (optional) or the first transit router, and continuing sequentially along the path up to the egress router (optional) or the router immediately before the egress router. You need to specify only one address per router hop. If you specify more than one address for the same router, only the first address is used; the additional addresses are ignored and truncated.

For each router address, you specify the type, which can be one of the following:

strict—(Default) The route taken from the previous router to this router is a direct path and cannot include any other routers. If *address* is an interface address, this router also ensures that the incoming interface is the one specified. Doing this is useful when there are parallel links between the previous router and this router. It also ensures that routing can be enforced on a per-link basis.

For strict addresses, you must ensure that the router immediately preceding the router you are configuring has a direct connection to that router. The address can be a loopback interface address, in which case the incoming interface is not checked.

loose—The route taken from the previous router to this router need not be a direct path and can include other routers and can be received on any interface. The address can be any interface address or the address of the loopback interface.

Examples: Create a Named Path

The following path, *to-hastings*, specifies the complete strict path from the ingress to the egress routers through 14.1.1.1, 13.1.1.1, 12.1.1.1 and 11.1.1.1, in that order. There cannot be any intermediate routers except the ones specified. However, there can be intermediate routers between 11.1.1.1 and the egress router because the egress router is not specifically listed in the path statement. To prevent intermediate routers before egress, configure the egress router as the last router, with a strict type.

```
[edit protocols mpls]
path to-hastings {
  14.1.1.1 strict;
  13.1.1.1 strict;
  12.1.1.1 strict;
  11.1.1.1 strict;
}
```

The following path, *alt-hastings*, allows any number of intermediate routers between routers 14.1.1.1 and 11.1.1.1. In addition, intermediate routers are permitted between 11.1.1.1 and the egress router.

```
[edit protocols mpls]
path alt-hastings {
  14.1.1.1 strict;
  11.1.1.1 loose;
}
```

Create an LSP

The second step in configuring signaled LSPs is to create one or more LSPs and define the properties associated with the label-switched path on the ingress router. To configure an LSP, include the label-switched-path statement at the [edit protocols mpls] hierarchy level:

```
[edit protocols mpls]
label-switched-path lsp-path-name {
  to address;
  from address;
  adaptive;
  admin-group {
    exclude group-names;
    include group-names;
  }
  auto-bandwidth {
    adjust-interval seconds;
    adjust-threshold percent;
    maximum-bandwidth bps;
    minimum-bandwidth bps;
    monitor-bandwidth;
  }
  bandwidth bps;
  class-of-service cos-value;
  description;
  disable;
  fast-reroute {
    fast-reroute bps;
    exclude group-names;
    hop-limit number;
    include group-names;
  }
  hop-limit number;
  install {
    destination-prefix/prefix-length <active>;
  }
  ldp-tunneling;
  link-protection;
  lsp-attributes {
    gpid gpid;
    signal-bandwidth type;
    switching-type type;
  }
  metric number;
  no-cspf;
  no-decrement-ttl;
  node-link-protection;
  optimize-timer seconds;
  preference preference;
  priority setup-priority hold-priority;
  (random | least-fill | most-fill);
  (record | no-record);
  retry-limit number;
  retry-timer seconds;
  standby;
}
```

```

traceoptions {
  file filename <replace> <size size> <files number> <no-stamp>
    <(world-readable | no-world-readable)>;
  flag flag <flag-modifier> <disable>;
}
primary path-name {
  adaptive;
  admin-group {
    exclude group-names;
    include group-names;
  }
  bandwidth bps;
  class-of-service cos-value;
  hop-limit number;
  no-cspf;
  no-decrement-ttl;
  optimize-timer seconds;
  preference preference;
  priority setup-priority hold-priority;
  (record | no-record);
  retry-limit number;
  retry-timer seconds;
  standby;
}
secondary path-name {
  adaptive;
  admin-group {
    exclude group-names;
    include group-names;
  }
  bandwidth bps;
  class-of-service cos-value;
  hop-limit number;
  no-cspf;
  no-decrement-ttl;
  optimize-timer seconds;
  preference preference;
  priority setup-priority hold-priority;
  (record | no-record);
  retry-limit number;
  retry-timer seconds;
  standby;
}
}

```

Each LSP must have a name, *lsp-path-name*, which can be up to 32 characters long and can contain letters, digits, periods (.), and hyphens (-). The name must be unique within the ingress router. For ease of management and identification, configure unique names across the entire domain.

When you configure LSPs, you can specify the following statements either for each LSP or for each path. (You configure LSPs at the [edit protocols mpls label-switched-path *lsp-path-name*] hierarchy level, and you configure paths at the [edit protocols mpls label-switched-path *lsp-path-name* primary] or [edit protocols mpls label-switched-path *lsp-path-name* secondary] hierarchy level.) For statements that you configure on a per-LSP basis, the value applies to all paths in the LSP. For statements that you configure on a per-path basis, the path value overrides the per-LSP value.

- adaptive
- admin-group
- auto-bandwidth
- bandwidth
- class-of-service
- hop-limit
- no-cspf
- optimize-timer
- preference
- priority
- record or no-record
- standby

For maintenance purposes, you can also configure the following attributes across all LSPs and any paths within those LSPs. (You can configure all the LSPs and paths at once at the [edit protocols mpls] hierarchy level.)

- admin-group
- bandwidth
- class-of-service
- no-decrement-ttl
- no-record
- optimize-timer
- preference
- priority
- standby

For each LSP, you can configure the following properties:

Configure the Address of the Egress Router on page 52

Configure the Address of the Ingress Router on page 52

Configure the Primary and Secondary LSPs on page 53

Configure the Description on page 53

Configure Fast Reroute on page 54

Configure Addresses to Associate with the LSP on page 58

Configure Path Connection Retry Information on page 59

Configure the LSP Metric on page 59

Configure CSPF Tie-Breaking on page 61

Configure Load Balancing for LSPs without CSPF on page 61

Configure Load Balancing for MPLS LSPs on page 61

Disable Normal TTL Decrementing on page 63

Configure Automatic Bandwidth Allocation on page 64

For each LSP and for each primary and secondary path, you can configure the following properties:

Disable Constrained-Path LSP Computation on page 66

Configure Administrative Groups on page 67

Configure the LSP Preference on page 69

Configure Whether to Record Path Routes on page 69

Configure the MPLS CoS Value on page 70

Rewrite IEEE 802.1p Packet Headers with the MPLS CoS Value on page 71

Configure an LSP to be Adaptive on page 71

Configure Priority and Preemption on page 72

Optimize Signaled LSPs on page 73

Configure the Maximum Path Length on page 74

Configure the Path Bandwidth on page 75

Configure the Standby State on page 75

Configure LSP Hold Time on page 76

Configure LDP Tunneling on page 76

Configure the Address of the Egress Router

When configuring an LSP, you must specify the address of the egress router by including the `to` statement at the `[edit protocols mpls label-switched-path lsp-path-name]` hierarchy level:

```
[edit protocols mpls label-switched-path lsp-path-name]  
  to address;
```

When you are setting up an LSP, the `to` statement is the only required statement. All other statements are optional.

After the LSP is established, the address of the egress router is installed as a host route in the routing table. Then, this route can be used by BGP to forward traffic.

To have the software send BGP traffic over an LSP, the address of the egress router is the same as the address of the BGP next-hop. You can specify the egress router's address as any one of the router's interface addresses or as the BGP router ID. If you specify a different address, even if the address is on the same router, BGP traffic is not sent over the LSP.

To determine the address of the BGP next-hop, use the `show route detail` command. To determine the destination address of an LSP, use the `show mpls lsp` command. To determine whether a route has gone through an LSP, use the `show route` or `show route forwarding-table` command. In the output of these last two commands, the `label-switched-path` or `push` keyword included with the route indicates it has passed through an LSP. Also, use the `traceroute` command to trace the actual path that the route leads to. This is another indication as to whether a route has passed through an LSP.

You also can manipulate the address of the BGP next-hop by defining a BGP import policy filter that sets the route's next-hop address.

Configure the Address of the Ingress Router

The local router always is considered to be the ingress router, which is the beginning of the LSP. The software automatically determines the proper outgoing interface and IP address to use to reach the next router in an LSP.

By default, the router ID is chosen as the address of the ingress router. To override the automatic selection of the source address, specify a source address in the `from` statement at the `[edit protocols mpls label-switched-path lsp-path-name]` hierarchy level:

```
[edit protocols mpls label-switched-path lsp-path-name]  
  from address;
```

The outgoing interface used by the LSP is not affected by the source address that you configure.

Configure the Primary and Secondary LSPs

By default, an LSP routes itself hop-by-hop toward the egress router. The LSP tends to follow the shortest path as dictated by the local routing table, usually taking the same path as destination-based, best-effort traffic. These paths are “soft” in nature because they automatically reroute themselves whenever a change occurs in a routing table or in the status of a node or link.

To configure the path so that it follows a particular route, create a named path using the path statement, as described in “Create a Named Path” on page 46. Then apply the named path by including the primary or secondary statement at the [edit protocols mpls label-switched-path *lsp-path-name*] hierarchy level:

```
[edit protocols mpls label-switched-path lsp-path-name]
primary path-name {
  ...
}
secondary path-name {
  ...
}
```

A named path can be referenced by any number of LSPs.

The primary statement creates the primary path, which is the LSP’s preferred path. The secondary statement creates an alternative path. If the primary path can no longer reach the egress router, the alternative path is used.

When the software switches from the primary to a secondary path, it continuously attempts to revert to the primary path, switching back to it when it is again reachable, but no sooner than the retry time specified in the retry-timer statement. (For more information, see “Configure Path Connection Retry Information” on page 59.)

You can configure zero or one primary path. If you do not configure a primary path, the first secondary path that is established is selected as the path.

You can configure zero or more secondary paths. All secondary paths are equal, and the software tries them in the order that they are listed in the configuration. The software does not attempt to switch among secondary paths. If the current secondary path is not available, the next one is tried. To create a set of equal paths, specify secondary paths without specifying a primary path.

If you do not specify any named paths, or if the path that you specify is empty, the software makes all routing decisions necessary to reach the egress router.

Configure the Description

To provide a textual description for the LSP, include the description statement at the [edit protocols mpls label-switched-path *lsp-path-name*] hierarchy level. Enclose any descriptive text that includes spaces in quotation marks (" "). Any descriptive text you include is displayed in the output of the show mpls lsp detail command and has no effect on the operation of the LSP.

```
[edit protocols mpls label-switched-path lsp-path-name]
description text;
```

The description text can be no more than 80 characters in length.

Configure Fast Reroute

Fast reroute provides a mechanism for automatically rerouting traffic on an LSP if a node or link in an LSP fails, thus reducing the loss of packets traveling over the LSP.

The following sections provide an overview of how fast reroute works and how to configure fast reroute:

Fast Reroute Overview on page 54

Detour Merging Procedure on page 56

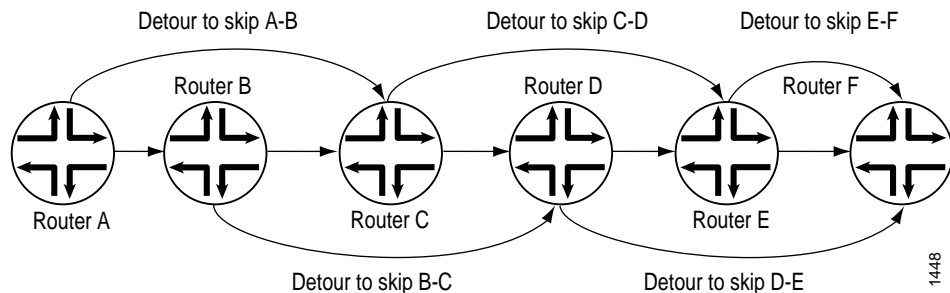
Detour Computations on page 57

Enable Fast Reroute on page 57

Fast Reroute Overview

Fast rerouting is accomplished by precomputing and pre-establishing a number of detours along the LSP. Figure 15 illustrates an LSP from Router A to Router F, showing some of the detours that are established for the LSP. Each detour is established by an upstream node with the intent of avoiding the link toward the immediate downstream node and the immediate downstream node itself. Each detour might traverse through one or more label-switched routers that are not shown in the figure.

Figure 15: Detours Established for an LSP Using Fast Reroute



If a node detects that a downstream link has failed (using a link-layer-specific liveness detection mechanism) or that a downstream node has failed (for example, using the RSVP neighbor hello protocol), the node quickly switches the traffic to the detour and, at the same time, signals the ingress router about the link or node failure. Figure 16 illustrates the detour taken when the link between Router B and Router C fails.

If the network topology is not rich enough (there are insufficient routers with insufficient links to other routers), some of the detours might not succeed. For example, the detour from Router A to Router C in Figure 15 cannot traverse link A-B and Router B. If such a path is not possible, the detour does not occur.

Note that after the node switches traffic to the detour, it might switch the traffic again to a newly calculated detour soon after. The initial detour route might not be the best route. Rather than verifying whether it is currently valid, the node simply switches traffic onto that route. After the node switches traffic to the initial detour, it recomputes the detour. If the node determines that the initial detour is still valid, traffic continues to flow over this detour. If the node determines that the initial detour is no longer valid, it again switches the traffic to the newly computed detour.

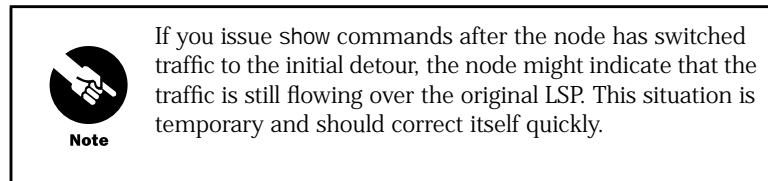
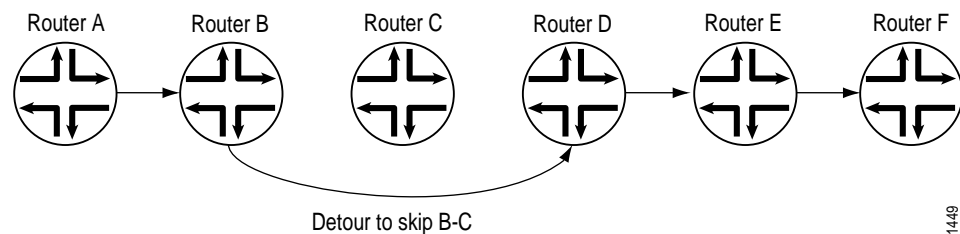


Figure 16: Detour after the Link from Router B to Router C Fails



The time required for a fast-rerouting detour to take effect depends on two independent time intervals:

Amount of time to detect that there is a link or node failure—This interval depends greatly on the link layer in use and the nature of the failure. For example, failure detection on an SDH/SONET link typically is much faster than on a Gigabit Ethernet link, and both are much faster than detection of a router failure.

Amount of time required to splice the traffic onto the detour—This operation is performed by the Packet Forwarding Engine (PFE), which requires little time to splice traffic onto the detour. The time needed can vary depending on the number of LSPs being switched to detours.

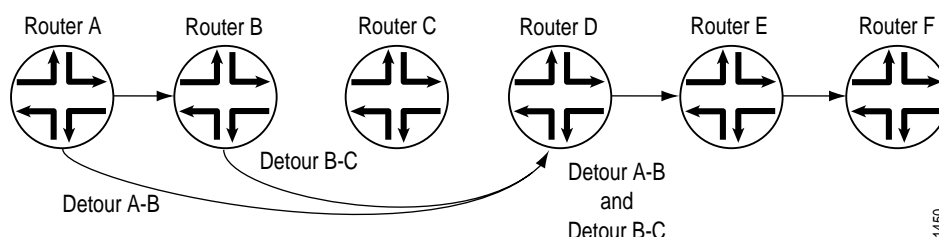
Fast reroute is a short-term patch to reduce packet loss. Because detour computation might not reserve adequate bandwidth, the detours might introduce congestion on the alternate links. The ingress router is the only router that is fully aware of LSP policy constraints and, therefore, is the only router able to come up with adequate long-term alternate paths.

Fast reroute protects traffic against any single point of failure between the ingress and egress routers. If there are multiple failures along an LSP, it is possible that fast reroute itself might fail. Also, fast reroute does not protect against failure of the ingress or egress routers.

Detours are created using RSVP and, like all RSVP sessions, they require extra state and overhead in the network. For this reason, each node establishes at most one detour for each router or transit node that has fast reroute enabled. Creating more than one detour for each LSP increases the overhead, but serves no practical purpose.

To reduce network overhead further, each detour attempts to merge back into the LSP as soon as possible after the failed node or link. If you can consider an LSP that travels through N router nodes, it is possible to create $N - 1$ detours. For instance, in Figure 17, the detour tries to merge back into the LSP at Router D instead of at Router E or Router F. Merging back into the LSP makes the detour scalability problem more manageable. If topology limitations prevent the detour from quickly merging back into the LSP, detours merge with other detours automatically.

Figure 17: Detours Merging into Other Detours



Detour Merging Procedure

This section describes the procedures used by a router to determine which LSP to select when the router receives Path messages from different interfaces with identical SESSION and SENDER_TEMPLATE objects. When this occurs, the router needs to merge the path states.

The router employs the following procedure to determine when and how to merge path states:

When all the Path messages do not include a FAST_REROUTE or a DETOUR object, or when the router is the egress of the LSP, no merging is required. The messages are processed according to RSVP-TE.

Otherwise, the router *must* record the path state in addition to the incoming interface. If the path messages do not share the same outgoing interface and next-hop router, the router considers them to be independent LSPs and does not merge them.

For all the Path messages that share the same outgoing interface and next-hop router, the router uses the following procedure to select the final LSP:

If only one LSP originates from this node, it is selected as the final LSP.

If only one LSP contains a FAST_REROUTE object, it is selected as the final LSP.

If there are several LSPs and some of them have a DETOUR object, eliminate those containing a DETOUR object from the final LSP selection process.

If several final LSP candidates remain (that is, there are still both DETOUR and protected LSPs), select the LSPs with FAST_REROUTE objects.

If none of the LSPs have FAST_REROUTE objects, select the ones without DETOUR objects. If all the LSPs have DETOUR objects, select them all.

Of the remaining LSP candidates, eliminate from consideration those that traverse nodes that other LSPs avoid.

If several candidate LSPs still remain, select the one with the shortest ERO path length. If more than one LSP has the same path length, select one randomly.

Once the final LSP has been identified, the router must only transmit the Path messages that correspond to this LSP. All other LSPs are considered merged at this node.

Detour Computations

Computing and setting up detours is done independently at each node. On a node, if an LSP has fast reroute enabled and if a downstream link or node can be identified, the router performs a CSPF computation using the information in the local Traffic Engineering Database (TED). For this reason, detours rely on your IGP supporting traffic engineering extensions. Without the TED, detours cannot be established.

CSPF initially attempts to find a path that skips the next downstream node. This provides protection against downstream failures in either nodes or links. If a node skipping path is not available, CSPF attempts to find a path on an alternate link to the next downstream node. This provides protection against downstream failures in links only. Detour computations might not succeed the first time. If a computation fails, the router recomputes detours approximately once every refresh interval until the computation succeeds. The RSVP metric for each detour is set to a value in the range of 10,000 through 19,999.

Enable Fast Reroute

To enable fast reroute on an LSP, include the fast-reroute statement at the [edit protocols mpls label-switched-path *lsp-path-name*] hierarchy level on the ingress router:

```
[edit protocols mpls label-switched-path lsp-path-name]
fast-reroute {
  bandwidth bps;
  (exclude group-names | no-exclude);
  hop-limit number;
  (include group-names | no-include);
}
```

You do not need to configure fast reroute on the LSP's transit and egress routers. Once fast reroute is enabled, the ingress router signals all the downstream routers that fast reroute is enabled on the LSP, and each downstream router does its best to set up detours for the LSP. If a downstream router does not support fast reroute, it ignores the request to set up detours and continues to support the LSP. A router that does not support fast reroute will cause some of the detours to fail, but otherwise has no impact on the LSP.

By default, no bandwidth is reserved for the rerouted path. To allocate bandwidth for the rerouted path, include the bandwidth statement. The bandwidth does not need to be identical to that allocated for the LSP.

Hop-limit constraints define how many more routers a detour is allowed to traverse compared to the LSP itself. By default, the hop limit is set to 6. For example, if an LSP traverses four routers, any detour for the LSP can be up to 10 (that is, 4 + 6) router hops, including the ingress and egress routers.

By default, a detour inherits the same administrative (coloring) group constraints as its parent LSP when CSPF is determining the alternate path. Administrative groups, also known as link coloring or resource class, are manually assigned attributes that describe the “color” of links, such that links with the same color conceptually belong to the same class. If you specify the include statement when configuring the parent LSP, all links traversed by the alternate session must have at least one color found in the list of groups. If you specify the exclude statement when configuring the parent LSP, none of the links must have a color found in the list of groups. For more information about administrative group constraints, see “Configure Administrative Groups” on page 67.

Configure Addresses to Associate with the LSP

By default, a host route toward the egress router is installed in the inet.3 routing table. (The host route address is the one you configure in the to statement.) Installing the host route allows BGP to perform next-hop resolution. It also prevents the host route from interfering with prefixes learned from dynamic routing protocols and stored in the inet.0 routing table.

Unlike the routes in the inet.0 table, routes in the inet.3 table are not copied to the Packet Forwarding Engine, and hence they cause no changes in the system forwarding table directly. You cannot ping or traceroute through these routes. The only use for inet.3 is to permit BGP to perform next-hop resolution. To examine the inet.3 table, use the show route table inet.3 command.

To inject additional routes into the inet.3 routing table, include the install statement at the [edit protocols mpls label-switched-path *lsp-path-name*] hierarchy level:

```
[edit protocols mpls label-switched-path lsp-path-name]
install {
  destination/mask <active>;
}
```

The specified routes are installed as aliases into the routing table when the LSP is established. Installing additional routes allows BGP to resolve next-hops within the specified prefix and to direct additional traffic for these next-hops to a particular LSP.

Including the active option with the install statement installs the specified prefix into the inet.0 routing table, which is the primary forwarding table. The result is a route that is installed in the forwarding table any time the LSP is established, which means you can ping or traceroute the route. Use this option with care, because this type of prefix is very similar to a static route.

You use alias routes for routers that have multiple addresses being used as BGP next-hops, or for routers that are not MPLS-capable. In either of these cases, the LSP can be configured to another MPLS-capable system within the local domain, which then acts as a “border” router. The LSP then terminates on the border router and, from that router, Layer 3 forwarding takes the packet to the true next-hop router.

In the case of an interconnect, the domain’s border router can act as the proxy router and can advertise the prefix for the interconnect if the border router is not setting the BGP next-hop to itself.

In the case of a POP that has routers that do not support MPLS, one router (for example, a core router) that supports MPLS can act as a proxy for the entire POP and can inject a set of prefixes that cover the POP. Thus, all routers within the POP can advertise themselves as IBGP next-hops, and traffic can follow the LSP to reach the core router. This means that normal IGP routing would prevail within the POP.

You cannot use the ping or traceroute commands on routes in the inet.3 routing table.

For BGP next-hop resolution, it makes no difference whether a route is in inet.0 or inet.3; the route with the best match (longest mask) is chosen. Among multiple best-match routes, the one with the highest preference value is chosen.

Configure Path Connection Retry Information

The ingress router might make many attempts to connect and reconnect to the egress router using the primary path. You can control how often the ingress router tries to establish a connection using the primary path and how long it waits between retry attempts.

The retry timer configures how long the ingress router waits before trying to connect again to the egress router using the primary path. The default retry time is 30 seconds. The time can be from 1 through 600 seconds. To modify this value, include the retry-timer statement at the [edit protocols mpls label-switched-path *lsp-path-name*] hierarchy level:

```
[edit protocols mpls label-switched-path lsp-path-name]
  retry-timer seconds;
```

By default, no limit is set to the number of times an ingress router attempts to establish or re-establish a connection to the egress router using the primary path. To limit the number of attempts, include the retry-limit statement at the [edit protocols mpls label-switched-path *lsp-path-name*] hierarchy level:

```
[edit protocols mpls label-switched-path lsp-path-name]
  retry-limit number;
```

The limit can be a value up to 10000. When the retry limit is exceeded, no more attempts are made to establish a path connection. At this point, intervention is required to restart the primary path.

If you set a retry limit, it is reset to 1 each time a successful primary path is created.

Configure the LSP Metric

The LSP metric is used to indicate the ease or difficulty of sending traffic over a particular LSP. Lower LSP metric values (lower cost) increase the likelihood of an LSP being used. Conversely, high LSP metric values (higher cost) decrease the likelihood of an LSP being used.

The LSP metric can be specified dynamically by the router or explicitly by the user as described in the following sections:

Configure a Dynamic LSP Metric on page 60

Configure a Static LSP Metric on page 60

Configure a Dynamic LSP Metric

If no specific metric is configured, an LSP attempts to track the IGP metric toward the same destination (the to address of the LSP). IGP includes OSPF, IS-IS, RIP, and static routes. BGP and other RSVP/LDP routes are excluded.

For example, if the OSPF metric toward a router is 20, all LSPs toward that router automatically inherit metric 20. If the OSPF toward a router later changes to a different value, all LSP metrics change accordingly. If there are no IGP routes toward the router, the LSP raises its metric to 65,535.

Note that in this case, the LSP metric is completely determined by IGP; it bears no relationship to the actual path the LSP is currently traversing. If LSP reroutes (such as through reoptimization), its metric does not change, and thus it remains transparent to users. Dynamic metric is the default behavior; no configuration is required.

Configure a Static LSP Metric

You can manually assign a fixed metric value to an LSP. Once configured using the metric statement at the [edit protocols mpls label-switched-path lsp-name] hierarchy level, the LSP metric is fixed and will not change:

```
[edit protocols mpls label-switched-path lsp-name]
metric number;
```

The LSP metric has several uses:

When there are parallel LSPs with the same egress router, the metrics are compared to see which LSP has the lowest metric value (the lowest cost) and therefore the preferred path to the destination. If the metrics are the same, the traffic is shared.

Adjusting the metric values can force traffic to prefer some LSPs over others, regardless of the underlying IGP metric.

When an IGP shortcut is enabled (see “IGP Shortcuts” on page 29), an IGP route might be installed in the routing table with an LSP as the next hop, if the LSP is on the shortest path to the destination. In this case, the LSP metric is added to the other IGP metrics to determine the total path metric. For example, if an LSP whose ingress router is X and egress router is Y is on the shortest path to destination Z, the LSP metric is added to the metric for the IGP route from Y to Z to determine the total cost of the path. If several LSPs are potential next hops, the total metrics of the paths are compared to determine which path is preferred (that is, has the lowest total metric). Or, IGP paths and LSPs leading to the same destination could be compared using the metric value to determine which path is preferred.

By adjusting the LSP metric, you can force traffic to prefer LSPs, to prefer the IGP path, or to share the load among them.

If router X and Y are BGP peers, and if there is an LSP between them, the LSP metric represents the total cost to reach Y from X. If for any reason the LSP reroutes, the underlying path cost might change significantly, but X’s cost to reach Y remains the same (the LSP metric), which allows X to report through BGP MED a stable metric to downstream neighbors. As long as Y remains reachable through the LSP, no changes are visible to downstream BGP neighbors.

Configure CSPF Tie-Breaking

When selecting a path for an LSP, CSPF uses a tie-breaking process if there are several equal-cost paths. For information about how CSPF selects a path, see “How CSPF Selects a Path” on page 26. To configure a random tie-breaking rule for CSPF to use to choose among equal-cost paths, include the random statement at the [edit protocols mpls path label-switched-path *lsp-path-name*] hierarchy level:

```
[edit protocols mpls path label-switched-path lsp-path-name]  
random;
```

To prefer the path with the least-utilized links, include the least-fill statement at the [edit protocols mpls path label-switched-path *lsp-path-name*] hierarchy level:

```
[edit protocols mpls path label-switched-path lsp-path-name]  
least-fill;
```

To prefer the path with the most-utilized links, include the most-fill statement at the [edit protocols mpls path label-switched-path *lsp-path-name*] hierarchy level:

```
[edit protocols mpls path label-switched-path lsp-path-name]  
most-fill;
```

Configure Load Balancing for LSPs without CSPF

An LSP tends to load-balance its placement by randomly selecting one of the equal-cost next hops and using it exclusively. The random selection is made independently at each transit router and is made by comparing IGP metrics alone. No consideration is given to bandwidth or congestion levels.

Configure Load Balancing for MPLS LSPs

Load balancing is used to evenly distribute traffic when:

There are multiple equal-cost nexthops over different interfaces to the same destination.

There is a single nexthop over an aggregated interface.

By default, when load balancing is used to help distribute traffic, the JUNOS software employs a hash algorithm to select a next-hop address to install into the forwarding table. Whenever the set of nexthops for a destination changes in any way, the next-hop address is reselected using the hash algorithm.

You can configure how the hash algorithm is used to load-balance traffic across a set of equal-cost LSPs. The hash algorithm can be configured to use the first MPLS label, the first two MPLS labels, or first MPLS label and the IP payload.

To configure what the hash algorithm uses to select a next-hop address, configure the hash-key statement at the [edit forwarding-options hash-key] hierarchy level:

```
[edit forwarding-options]
hash-key {
  family mpls {
    label-1;
    label-2;
    payload {
      ip;
    }
  }
}
```

To use the first MPLS label for the hash key, configure the label-1 statement at the [edit forwarding-options hash-key family mpls] hierarchy level:

```
[edit forwarding-options hash-key family mpls]
label-1;
```

To use the second MPLS label in the hash key, configure both the label-1 statement and the label-2 statement at the [edit forwarding-options hash-key family mpls] hierarchy level:

```
[edit forwarding-options hash-key family mpls]
label-1;
label-2;
```

To use the MPLS packet's IP payload (IPv4 or IPv6) in the hash key, configure the ip option at the [edit forwarding-options hash-key family mpls payload] hierarchy level:

```
[edit forwarding-options hash-key family mpls]
label-1;
payload {
  ip;
}
```

When you configure the ip option at the [edit forwarding-options hash-key family mpls payload] hierarchy level, you must also configure the label-1 statement at the [edit forwarding-options hash-key family mpls] hierarchy level. The IP payload can only be included if there is single MPLS label. If there are two MPLS labels, the packet is load-balanced based on the labels (the IP payload is excluded).



Note

The router determines if the MPLS payload is an IP packet by checking the byte containing the IP version number. If the IP version number is 4 (IPv4) or 6 (IPv6) then the packet is assumed to be an IP packet.

For more information on statements configured under the [edit forwarding-options] hierarchy level, see the *JUNOS Internet Software Configuration Guide: Policy Framework*.

Disable Normal TTL Decrementing

By default, the TTL field value in the packet header is decremented by 1 for every hop the packet traverses in the LSP, thereby preventing loops. If the TTL field value reaches 0, packets are dropped, and an ICMP error packet might be sent to the originating router.

If normal TTL decrement is disabled, the TTL field of IP packets entering LSPs are decremented by only 1 upon transiting the LSP, making the LSP appear as a one-hop router to diagnostic tools, such as traceroute. This is done by the ingress router, which pushes a label on IP packets with the TTL field in the label initialized to 255. The label's TTL field value is decremented by 1 for every hop the MPLS packet traverses in the LSP. On the penultimate hop of the LSP, the router pops the label but does not write the label's TTL field value to the IP packet's TTL field. Instead, when the IP packet reaches the egress router, the IP packet's TTL field value is decremented by 1.

When you use traceroute to diagnose problems with an LSP, traceroute sees the ingress router, although the egress router performs the TTL decrement. Note that this assumes that traceroute is initiated outside of the LSP. The behavior of traceroute is different if it is initiated from the ingress router of the LSP. In this case, the egress router would be the first router to respond to traceroute.

You can disable normal TTL decrementing in an LSP so that the TTL field value does not reach 0 before the packet reaches its destination, thus preventing the packet from being dropped. You can also disable normal TTL decrementing to make the MPLS cloud appear as a single hop, thereby hiding the network topology.

There are two ways to disable TTL decrementing:

On the ingress of the LSP, if you include the `no-decrement-ttl` statement at the [edit protocols mpls label-switched-path *lsp-path-name*] hierarchy level, the ingress router negotiates with all downstream routers using a proprietary RSVP object, to ensure all routers are in agreement. If negotiation succeeds, the whole LSP behaves as one hop to transit IP traffic.

```
[edit protocols mpls label-switched-path lsp-path-name]  
no-decrement-ttl;
```

Note that the RSVP object is proprietary to the JUNOS software and might not work with other software. This potential incompatibility only applies to RSVP-signaled LSPs, not to LDP-signaled LSPs. When you include the `no-decrement-ttl` statement, TTL hiding can be enforced on a per-LSP basis.

On the router, you can include the `no-propagate-ttl` statement at the [edit protocols mpls] hierarchy level. This statement applies to all LSPs, regardless of whether they are RSVP-signaled or LDP-signaled. Once set, all future LSPs traversing through this router behave as a single hop to IP packets. LSPs established before you configure this statement are not affected.

```
[edit protocols mpls]  
no-propagate-ttl;
```

If you include the `no-propagate-ttl` statement, make sure all routers are configured consistently within an MPLS domain; failing to do so might cause the IP packet TTL to increase while in transit within LSPs. This can happen, for example, when the ingress router has `no-propagate-ttl` configured but the penultimate router does not, so the penultimate router writes the MPLS TTL value (which starts from the ingress router as 255) into the IP packet.

The operation of the `no-propagate-ttl` statement is more interoperable with other vendors' equipment. However, you must ensure all routers are configured identically.

Configure Automatic Bandwidth Allocation

Automatic bandwidth allocation allows an MPLS tunnel to automatically adjust its bandwidth allocation based on the volume of traffic flowing through the tunnel. You can configure an LSP with minimal bandwidth, and this feature can dynamically adjust the LSP's bandwidth allocation based on current traffic patterns. The bandwidth adjustments do not interrupt traffic flow through the tunnel.

At the end of the time interval specified under the protocols `mpls label-switched-path auto-bandwidth` hierarchy level, the current maximum average bandwidth usage is compared to the allocated bandwidth for the LSP. If the LSP needs more bandwidth, an attempt is made to set up a new path where bandwidth is equal to the current maximum average usage. If the attempt is successful, the LSP's traffic is routed through the new path and the old path is removed. If the attempt fails, the LSP continues to use its current path.



Note

Note that you might not be able to use this feature to adjust the bandwidth of fast-reroute LSPs. Because the LSPs use fixed filter (FF) reservation style, when a new path is signaled, the bandwidth might be double-counted. This can prevent a fast-reroute LSP from ever adjusting its bandwidth when automatic bandwidth allocation is enabled.

Configure MPLS Statistics

To enable automatic bandwidth allocation, you first need to configure MPLS statistics. As part of this configuration, include the `auto-bandwidth` statement at the protocols `mpls statistics` hierarchy level. You can also use the `interval` statement to configure the interval for calculating the average bandwidth usage. This setting applies to all LSPs configured on the router. You can set the adjustment interval (configured at the protocols `mpls label-switched-path label-switched-path-name auto-bandwidth` hierarchy level) on specific LSPs.

Configure the MPLS and automatic bandwidth allocation statistics as follows:

```
[edit]
protocols {
  mpls {
    statistics {
      auto-bandwidth;
      file filename size size files number <no-stamp>;
      interval seconds;
    }
  }
}
```

Configure the Maximum and Minimum Bounds of the LSP's Bandwidth

You can maintain the LSP's bandwidth between minimum and maximum bounds by specifying values for the minimum-bandwidth and maximum-bandwidth statements at the [edit protocols mpls label-switched-path *label-switched-path-name* auto-bandwidth] hierarchy level. Specify the bandwidth reallocation interval in seconds using the adjust-interval statement.

Configure automatic bandwidth allocation as follows:

```
[edit protocols mpls label-switched-path label-switched-path-name]
auto-bandwidth {
  adjust-interval seconds;
  adjust-threshold percent;
  minimum-bandwidth bps;
  maximum-bandwidth bps;
}
```

Configure the Threshold for Automatic Bandwidth Adjustment

Use the adjust-threshold statement to specify the sensitivity of the automatic bandwidth adjustment of an LSP to changes in bandwidth utilization. You can set the threshold for when to trigger automatic bandwidth adjustments. When configured, bandwidth demand for the current interval is determined and compared to the LSP's current bandwidth allocation. If the percentage difference in bandwidth is greater than or equal to the specified adjust-threshold percentage, the LSP's bandwidth is adjusted to the current bandwidth demand.

For example, assume that the current bandwidth allocation is 100 Mbps and that the percentage configured for the adjust-threshold statement is 15 percent. If the bandwidth demand increases to 110 Mbps, the bandwidth allocation is not adjusted. However, if the bandwidth demand increases to 120 Mbps (20 percent over the current allocation) or decreases to 80 Mbps (20 percent under the current allocation), the bandwidth allocation is increased to 120 Mbps or decreased to 80 Mbps, respectively.

Set the adjust-threshold statement as follows:

```
[edit protocols mpls label-switched-path label-switched-path-name]
auto-bandwidth {
  adjust-threshold percent;
}
```

Configure Passive Bandwidth Utilization Monitoring

The monitor-bandwidth statement is used to switch to a passive bandwidth utilization monitoring mode. In this mode, no automatic bandwidth adjustments are made, but the maximum average bandwidth utilization is continuously monitored and recorded.

Set the monitor-bandwidth statement as follows:

```
[edit protocols mpls label-switched-path label-switched-path-name]
auto-bandwidth {
  monitor-bandwidth;
}
```

If you have configured an LSP with primary and secondary paths, the automatic bandwidth allocation statistics are carried over to the secondary path if the primary path fails. For example, consider a primary path whose adjustment interval is half complete and whose maximum average bandwidth usage is currently calculated as 50 Mbps. If the primary path suddenly fails, the time remaining for the next adjustment and the maximum average bandwidth usage are carried over to the secondary path.

Disable Constrained-Path LSP Computation

If the IGP is a link-state protocol (such as IS-IS or OSPF) and it supports extensions that allow the current bandwidth reservation on each router's link to be reported, constrained-path LSPs are computed by default.

The JUNOS implementations of IS-IS and OSPF include the extensions that support constrained-path LSP computation.

IS-IS—These extensions are enabled by default. To disable this support, include the `disable` statement at the `[edit protocols isis traffic-engineering]` hierarchy level, as discussed in the *JUNOS Internet Software Configuration Guide: Routing and Routing Protocols*.

OSPF—These extensions are disabled by default. To enable this support, include the `traffic-engineering` statement in the configurations of all routers running OSPF, as described in the *JUNOS Internet Software Configuration Guide: Routing and Routing Protocols*.

If IS-IS is enabled on a router or you enable OSPF traffic engineering extensions, MPLS performs the constrained-path LSP computation by default.

Constrained-path LSP computation works as follows:

1. LSPs advertise their link information in the IGP's link-state packets.
2. These packets are flooded throughout the network, providing link information to all nodes.
3. This link information is placed into the traffic engineering database (TED) and provides each ingress router with LSP topology information and recent LSP bandwidth reservation information.
4. When computing complete paths for LSPs, the ingress router uses the information in the TED in conjunction with the requirements you configure for the LSP, including bandwidth (configured with the `bandwidth` statement), hop limit (configured with the `hop-limit` statement), and the address of the egress router (configured with the `to` statement).

Constrained-path LSPs have a greater chance of being established quickly and successfully for the following reasons:

The LSP computation takes into account the current bandwidth reservation.

Constrained-path LSPs reroute themselves away from node failures and congestion.

When constrained-path LSP computation is enabled, you can configure the LSP so that it is periodically reoptimized, as described in "Optimize Signaled LSPs" on page 73.

When an LSP is being established or when an existing LSP fails, the constrained-path LSP computation is repeated periodically at the interval specified by the retry timer until the LSP is set up successfully. Once the LSP is set up, no recomputation is done. For more information about the retry timer, see “Configure Path Connection Retry Information” on page 59.

By default, constrained-path LSP computation is enabled. You might want to disable constrained-path LSP computation when all nodes do not support the necessary traffic engineering extensions. To disable constrained-path LSP computation, include the `no-cspf` statement at the `[edit protocols mpls label-switched-path lsp-path-name]` or `[edit protocols mpls label-switched-path lsp-path-name (primary | secondary)]` hierarchy level:

```
no-cspf;
```

Configure Administrative Groups

Administrative groups, also known as link coloring or resource class, are manually assigned attributes that describe the “color” of links, such that links with the same color conceptually belong to the same class. You can use administrative groups to implement a variety of policy-based LSP setups.

Administrative groups are meaningful only when constrained-path LSP computation is enabled.

Administrative groups require three levels of configuration. First, configure a table of group names at the `[edit protocols mpls]` hierarchy level:

```
[edit protocols mpls]
admin-groups {
  group-name group-value;
}
```

You can assign up to 32 names and values (in the range 0 through 31), which define a series of names and their corresponding values. The administrative names and values must be identical across all routers within a single domain.

To configure administrative groups, follow these steps:

1. Define multiple levels of service quality:

```
[edit]
protocols {
  mpls {
    admin-groups {
      best-effort 1;
      copper 2;
      silver 3;
      gold 4;
      violet 5;
    }
  }
}
```

2. Define administrative groups for an interface. These groups identify the administrative groups to which an interface belongs. You can assign multiple groups to an interface.

```
[edit]
protocols {
  mpls {
    interface interface name {
      admin-group [ group-name group-name... ];
    }
  }
}
```

If you do not include the `admin-group` statement, an interface does not belong to any group.

IGPs use the group information to build link-state packets, which are then flooded throughout the network, providing information to all nodes in the network. At any router, the IGP topology, as well as administrative groups of all the links, are available.

Changing the interface's administrative group affects only new LSPs. Existing LSPs on the interface are not preempted or recomputed to keep the network stable. If LSPs need to be removed because of a group change, issue the `clear rsvp session` command.

3. Configure an administrative group constraint for each LSP or for each primary or secondary LSP path, at the `[edit protocols mpls label-switched-path lsp-path-name]` or `[edit protocols mpls label-switched-path lsp-path-name (primary | secondary)]` hierarchy level:

```
[edit]
protocols {
  mpls {
    label-switched-path lsp-path-name {
      to address;
      ...
      primary path-name {
        admin-group {
          exclude [ group-name group-name ... ];
          include [ group-name group-name ... ];
        }
      }
      secondary path-name {
        admin-group {
          exclude [ group-name group-name ... ];
          include [ group-name group-name ... ];
        }
      }
      admin-group {
        exclude [ group-name group-name ... ];
        include [ group-name group-name ... ];
      }
    }
  }
}
```

If you omit the include or exclude statements, the path computation proceeds unchanged using constrained-path LSP computation. If you configure an exclude list, all chosen links must not have a color in the exclude list. If you configure an include list, all chosen links must have at least one color in the include list. Links that have no color are automatically disqualified by any include or exclude list.



Note

Changing the LSP's administrative group causes an immediate recomputation of the route; therefore, the LSP might be rerouted.

Configure the LSP Preference

As an option, you can configure multiple LSPs between the same pair of ingress and egress routers. This is useful for balancing the load among the LSPs because all LSPs, by default, have the same preference level. To prefer one LSP over another, set different preference levels for individual LSPs. The LSP with the lowest preference value is used. The default preference of all LSPs is 7, which is lower (more preferred) than all learned routes except for direct interface routes.

To change the default preference value, include the preference statement at the [edit protocols mpls], [edit protocols mpls label-switched-path *lsp-path-name*], or [edit protocols mpls label-switched-path *lsp-path-name* (primary | secondary)] hierarchy level:

```
preference preference;
```

Configure Whether to Record Path Routes

The JUNOS implementation of RSVP supports the Record Route Object, which allows an LSP to actively record the routers through which it transits. You can use this information for troubleshooting and to prevent routing loops. By default, path route information is recorded. To disable recording, include the no-record statement at the [edit protocols mpls], [edit protocols mpls label-switched-path *lsp-path-name*], or [edit protocols mpls label-switched-path *lsp-path-name* (primary | secondary)] hierarchy level:

```
no-record;
```

Configure the MPLS CoS Value

When IP traffic enters an LSP tunnel, the ingress router marks all packets with a class-of-service (CoS) value, which is used to place the traffic into a transmission priority queue. On the router, for SDH/SONET and T3 interfaces, each interface has four transmit queues. The CoS value is encoded as part of the MPLS header and remains in the packets until the MPLS header is removed when the packets exit from the egress router. The routers within the LSP utilize the CoS value set at the ingress router. The CoS value is encoded using the class-of-service bits (also known as the EXP or experimental bits). For more information, see “Label Allocation” on page 21.

MPLS class of service works in conjunction with the router’s general CoS functionality. If you do not configure any CoS features, the default general CoS settings are used. For MPLS class of service, you might want to prioritize how the transmit queues are serviced by configuring weighted round-robin, and to configure congestion avoidance using Random Early Detection (RED). The general CoS features are described in the *JUNOS Internet Software Configuration Guide: Network Interfaces and Class of Service*.

When traffic enters an LSP tunnel, the CoS bits in the MPLS header are set in one of two ways. In the first way, the number of the output queue into which the packet was buffered and the Packet Loss Priority (PLP) bit are written into the MPLS header and are used as the packet’s CoS value. This behavior is the default, and no configuration is required. The *JUNOS Internet Software Configuration Guide: Network Interfaces and Class of Service* explains the IP CoS values, and summarizes how the CoS bits are treated.

In the second way, you set a fixed CoS value on all packets entering the LSP tunnel. This means that all packets entering the LSP receive the same class of service. To do this, include the class-of-service statement at the [edit protocols mpls], [edit protocols mpls label-switched-path *lsp-path-name*], or [edit protocols mpls label-switched-path *lsp-path-name* (primary | secondary)] hierarchy level:

```
class-of-service cos-value;
```

The CoS value can be a decimal number from 0 through 7. This number corresponds to a 3-bit binary number. The high-order 2 bits of the CoS value select which transmit queue to use on the outbound interface card.

The low-order bit of the CoS value is treated as the PLP bit and is used to select the RED drop profile to use on the output queue. If the low-order bit is 0, the non-PLP drop profile is used, and if the low-order bit is 1, the PLP drop profile is used. It is generally expected that RED will more aggressively drop packets that have the PLP bit set. For more information about RED and drop profiles, see the *JUNOS Internet Software Configuration Guide: Network Interfaces and Class of Service*.



Note

Configuring the PLP drop profile to drop packets more aggressively (for example, setting the CoS value from 6 to 7) decreases the likelihood of traffic getting through.

Table 2 summarizes how MPLS CoS values correspond to the transmit queue and PLP bit. Note that in MPLS, the mapping between the CoS bit value and the output queue is hard-coded. You cannot configure the mapping for MPLS; you can configure it only for IPv4 traffic flows, as described in the *JUNOS Internet Software Configuration Guide: Network Interfaces and Class of Service*.

Table 2: MPLS CoS Values

MPLS CoS Value	Bits	Transmit Queue	PLP Bit
0	000	0	Not set
1	001	0	Set
2	010	1	Not set
3	011	1	Set
4	100	2	Not set
5	101	2	Set
6	110	3	Not set
7	111	3	Set

Because the CoS value is part of the MPLS header, the value is associated with the packets only as they travel through the LSP tunnel. The value is not copied back to the IP header when the packets exit from the LSP tunnel.

Rewrite IEEE 802.1p Packet Headers with the MPLS CoS Value

For Ethernet interfaces installed on a T-series platform with a peer connection to an M-series router or a T-series platform, you can rewrite both MPLS CoS and IEEE 802.1p bits to a configured value (the MPLS CoS bits are also known as the EXP or experimental bits). This allows you to pass the configured value to the Layer 2 VLAN path. To rewrite both the MPLS CoS and IEEE 802.1p bits, you must include the EXP and IEEE 802.1p rewrite rules in the class-of-service interface configuration. The EXP rewrite table is applied when you configure the IEEE 802.1p and EXP rewrite rules.

For information on how to configure the EXP and IEEE 802.1p rewrite rules, see the *JUNOS Internet Software Configuration Guide: Network Interfaces and Class of Service*.

For information on the CoS bits, see “Label Allocation” on page 21 and “Configure the MPLS CoS Value” on page 70.

Configure an LSP to be Adaptive

An LSP occasionally might need to reroute itself. Reasons include the following:

- Continuous reoptimization process is configured with the optimize-timer statement.

- The current path has connectivity problems.

- The LSP is preempted by another LSP configured with the priority statement and is forced to reroute.

- The explicit-path information for an active LSP is modified, or the LSP’s bandwidth is increased.

You can configure an LSP to be *adaptive* when it is attempting to reroute itself. When it is adaptive, the LSP holds onto existing resources until the new path is successfully established and traffic has been cut over to the new LSP. To retain its resources, an adaptive LSP does the following:

- Maintains existing paths and allocated bandwidths—This ensures that the existing path is not torn down prematurely and allows the current traffic to continue flowing while the new path is being set up.

- Avoids double-counting for links that share the new and old paths—Double-counting occurs when an intermediate router does not recognize that the new and old paths belong to the same LSP and counts them as two separate LSPs, requiring separate bandwidth allocations. If some links are close to saturation, double-counting might cause the setup of the new path to fail.

By default, adaptive behavior is disabled. You can include the adaptive statement in two different hierarchy levels. If you specify the adaptive statement at the LSP hierarchy level [edit protocols mpls label-switched-path *lsp-path-name*], adaptive behavior is enabled on all primary/secondary paths of the LSP. This means both the primary and secondary paths share the same bandwidth on common links.

```
[edit protocols mpls label-switched-path lsp-path-name]  
adaptive;
```

If you specify the adaptive statement at the primary or secondary hierarchy level [edit protocols mpls label-switched-path *lsp-path-name* (primary | secondary)], adaptive behavior is enabled only on the path on which it is specified. Bandwidth double-counting happens between different paths.

```
[edit protocols mpls label-switched-path lsp-path-name (primary | secondary)]  
adaptive;
```

Configure Priority and Preemption

When there is insufficient bandwidth to establish a more important LSP, you might want to tear down a less important existing LSP to free up the bandwidth. You do this by preempting the existing LSP.

Whether an LSP can be preempted is determined by two properties associated with the LSP:

- Setup priority—Determines whether a new LSP that preempts an existing LSP can be established. For preemption to occur, the setup priority of the new LSP must be higher than that of the existing LSP. Also, the act of preempting the existing LSP must produce sufficient bandwidth to support the new LSP. That is, preemption occurs only if the new LSP can be set up successfully.

- Hold priority—Determines the degree to which an LSP holds onto its session reservation after the LSP has been set up successfully. When the hold priority is high, the existing LSP is less likely to give up its reservation and hence it is unlikely that the LSP can be preempted.

You cannot configure an LSP with a high setup priority and a low hold priority because permanent preemption loops might result if two LSPs are allowed to preempt each other. You must configure the hold priority to be higher than or equal to the setup priority.

The setup priority also defines the relative importance of LSPs on the same ingress router. When the software starts, when a new LSP is established, or during fault recovery, the setup priority determines the order in which LSPs are serviced. Higher priority LSPs tend to be established first and hence enjoy more optimal path selection.

To configure the LSP's preemption properties, include the priority statement at the [edit protocols mpls], [edit protocols mpls label-switched-path *lsp-path-name*], or [edit protocols mpls label-switched-path *lsp-path-name* (primary | secondary)] hierarchy level:

```
priority setup-priority hold-priority;
```

Both *setup-priority* and *hold-priority* can be a value from 0 through 7. The value 0 corresponds to the highest priority, and the value 7 to the lowest. By default, an LSP has a setup priority of 7 (that is, it cannot preempt any other LSPs) and a hold priority of 0 (that is, other LSPs cannot preempt it). These defaults are such that preemption does not happen. When you are configuring these values, the setup priority should always be less than or equal to the hold priority.

Optimize Signaled LSPs

Once an LSP has been established, topology or resources changes might, over time, make the path suboptimal. A subsequent recomputation might be able to determine a more optimal path.

If reoptimization is enabled, an LSP can be rerouted through different paths by constrained-path recomputations. However, if reoptimization is disabled, the LSP has a fixed path and cannot take advantage of newly available network resources. The LSP is fixed until the next topology change breaks the LSP and forces a recomputation.

Reoptimization is not related to failover. A new path is always computed when topology failures occur that disrupt an established path.

Because of the potential system overhead involved, you need to control carefully the frequency of reoptimization. Network stability might suffer when reoptimization is enabled. By default, *optimize-timer* is set to 0 (that is, it is disabled).

Configuring LSP optimization is meaningful only when constrained-path LSP computation is enabled, which is the default behavior. For more information about constrained-path LSP computation, see "Disable Constrained-Path LSP Computation" on page 66.

To enable path reoptimization, include the *optimize-timer* statement at the [edit protocols mpls], [edit protocols mpls label-switched-path *lsp-path-name*], or [edit protocols mpls label-switched-path *lsp-path-name* (primary | secondary)] hierarchy level:

```
optimize-timer seconds;
```

Once you have configured the *optimize-timer* statement, the reoptimization timer continues its countdown to the configured value even if you delete the *optimize-timer* statement from the configuration.

If you change the value configured for the *optimize-timer* statement, the old value is still used for the next optimization. The optimization after that uses the new value. You can force JUNOS to use a new value immediately by deleting the old value, committing the configuration, and then configuring the new value for the *optimize-timer* statement and committing the configuration again.

After reoptimization is run, the result is accepted only if it meets the following criteria:

1. The new path is not higher in IGP metric. (The metric for the old path is updated during computation, so if a recent link metric changed somewhere along the old path, it is accounted for.)
2. If the new path has the same IGP metric, it is not more hops away.
3. The new path does not cause preemption. (This is to reduce the ripple effect of preemption causing more preemption.)
4. The new path does not worsen congestion overall. This is done by comparing the percentage of available bandwidth on each link traversed by the new and old paths, starting from the most congested links.

When all the above conditions are met, then:

5. If the new path has a lower IGP metric, it is accepted.
6. If the new path has an equal IGP metric and lower hop count, it is accepted.
7. If you choose least-fill as a load-balancing algorithm and if the new path reduces congestion by at least 10 percent aggregated over all links it traversed, it is accepted. For random or most-fill algorithms, this rule does not apply.
8. Otherwise, the new path is rejected.

To disable items 2, 3, 4 and 6 above, enter the clear `mpls optimize-aggressive` command or at the `[edit protocols mpls]` hierarchy level, include the `optimize-aggressive` statement:

```
optimize-aggressive;
```

Including the `optimize-aggressive` statement makes the reoptimization process more aggressive. Not only does it tend to reroute more often, it also limits the reoptimization algorithm to be based on the IGP metric only.

Configure the Maximum Path Length

By default, each LSP can traverse a maximum of 255 hops, including the ingress and egress routers. To modify this value, include the `hop-limit` statement at the `[edit protocols mpls]`, `[edit protocols mpls label-switched-path lsp-path-name]`, or `[edit protocols mpls label-switched-path lsp-path-name (primary | secondary)]` hierarchy level:

```
hop-limit number;
```

The number of hops can be from 2 through 255. (A path with two hops consists of the ingress and egress routers only.)

Configure the Path Bandwidth

Each LSP has a bandwidth value. This value is included in the sender's Tspec field in RSVP path setup messages. To specify a bandwidth value, include the bandwidth statement at the [edit protocols mpls], [edit protocols mpls label-switched-path *lsp-path-name*], or [edit protocols mpls label-switched-path *lsp-path-name* (primary | secondary)] hierarchy level:

```
bandwidth bps;
```

You specify the bandwidth value in bits per second, with a higher value implying a greater user traffic volume. The default bandwidth is 0 bits per second.

A nonzero bandwidth requires transit routers to reserve capacity along the outbound links for the path. This is done using RSVP's reservation scheme. Any failure in bandwidth reservation (such as failures at RSVP policy control or admission control) might cause the LSP setup to fail.

Configure the Standby State

By default, secondary paths are set up only as needed. To have the system maintain a secondary path in a hot-standby state indefinitely, include the standby statement at the [edit protocols mpls label-switched-path *lsp-path-name* secondary] hierarchy level:

```
standby;
```

The hot-standby state is meaningful only on secondary paths. Maintaining a path in a hot-standby state enables swift cutover to the secondary path when downstream routers on the current active path indicate connectivity problems. Though it is possible to configure the standby statement at the [edit protocols mpls label-switched-path *lsp-path-name* primary *path-name*] hierarchy level, it has no effect on router behavior.

If you configure the standby statement at the [edit protocols mpls] or the [edit protocols mpls label-switched-path *lsp-path-name*] the hot-standby state is activated on all secondary paths configured beneath that hierarchy level.

The hot-standby state has two advantages:

- It eliminates the call-setup delay during network topology changes. Call setup can suffer from significant delays when network failures trigger large numbers of LSP reroutes at the same time.

- A cutover to the secondary path can be made before RSVP learns that an LSP is down. There can be significant delays between the time the first failure is detected by protocol machinery (which can be an interface down, a neighbor becoming unreachable, a route becoming unreachable, or a transient routing loop being detected) and the time an LSP actually fails (which requires a timeout of soft state information between adjacent RSVP routers). When topology failures occur, hot-standby secondary paths can usually achieve the smallest cutover delays with minimal disruptions to user traffic.

When the primary path is considered to be stable again, traffic is automatically switched from the standby secondary path back to the primary path. The switch is performed no faster than twice the retry-timer interval and only if the primary path exhibits stability throughout the entire switch interval.

The drawback of the hot-standby state is that more state information must be maintained by all the routers along the path, which requires overhead from each of the routers.

Configure LSP Hold Time

When an LSP changes from being up to being down, or from down to up, this transition takes effect immediately in the router software and hardware. However, when advertising LSPs into IS-IS, you may want to damp LSP transitions, thereby not advertising the transition until a certain period of time has transpired (known as the hold time). In this case, if the LSP goes from up to down, the LSP is not advertised as being down until it has remained down for the hold-time period. Transitions from down to up are advertised into IS-IS immediately. Note that LSP damping only affects IS-IS advertisements of the LSP; other routing software and hardware react immediately to LSP transitions.

To damp LSP transitions, you can include the advertisement-hold-time statement at the [edit protocols mpls] hierarchy level:

```
[edit protocols mpls]
  advertise-hold-time seconds;
```

seconds can be a value from 0 through 65,535 seconds. The default is 5 seconds.

Configure LDP Tunneling

To correctly identify an LDP session associated with an RSVP LSP, ensure that the RSVP LSP endpoint address is the same as the transport address of the LDP peer.

Configure Alternate Backup Paths Using Fate-Sharing

You can create a database of information that CSPF uses to compute one or more backup paths to use in case the primary path becomes unstable. The database describes the relationships between elements of the network, such as routers and links. Because these network elements share the same fate, this relationship is called *fate sharing*.

You can configure backup paths that minimize the number of shared links and fiber paths with the primary paths as much as possible to ensure that, if a fiber is cut, the minimum amount of data is lost and that a path still exists to the destination.

For a backup path to work optimally, it must not share links or physical fiber paths with the primary path. This ensures that a single point of failure will not affect the primary and backup paths at the same time.

To configure fate sharing, include the fate-sharing statement at the [edit routing-options] hierarchy level:

```
[edit routing-options]
  fate-sharing {
    group group-name {
      cost value;
      from address <to address>;
    }
  }
```

Each fate-sharing group must have a name, which can be up to 32 characters long and can contain letters, digits, periods (.) and hyphens (-). You can define up to 512 groups.

Fate-sharing groups contain three types of objects:

Point-to-point links—Identified by the IP addresses at each end of the link. Unnumbered point-to-point links are typically identified by borrowing IP addresses from other interfaces. Order is not important; from 1.2.3.4 to 1.2.3.5 and from 1.2.3.5 to 1.2.3.4 have the same meaning.

Nonpoint-to-point links—Include links on a LAN interface (such as Gigabit Ethernet interfaces), or NBMA interfaces, (such as ATM or Frame Relay). You identify these links by their individual interface address. For example, if the LAN interface 192.168.200.0/24 has four routers attached to it, each router link is individually identified:

```
from 192.168.200.1; # LAN interface of router 1
from 192.168.200.2; # LAN interface of router 2
from 192.168.200.3; # LAN interface of router 3
from 192.168.200.4; # LAN interface of router 4
```

You can list the addresses in any order.

A router node—Identified by its configured router ID.

All objects in a group share certain similarities. For example, you can define a group for all fibers sharing the same fiber conduit, all optical channels that share the same fiber, all links that connect to the same LAN switch, all equipment sharing the same power source, and so on. All objects are treated as /32 host addresses.

For a group to be meaningful, it should contain at least two objects. You can configure groups with zero or one object; these groups are ignored during processing.

An object can be in any number of groups, and a group can contain any number of objects. Each group has a configurable cost attributed to it, which represents the level of impact this group has on CSPF computations. The higher the cost, the less likely a backup path will share with the primary path any objects in the group. The cost is directly comparable to traffic engineering metrics. By default, the cost is 1. Changing the fate-sharing database does not affect existing established LSPs until the next reoptimization of CSPF. The fate-sharing database does influence fast-reroute computations.

Implications to CSPF

When CSPF computes the primary paths of an LSP (or secondary paths when the primary path is not active), it ignores the fate-sharing information. You always want to find the best possible path (least IGP cost) for the primary path.

When CSPF computes a secondary path while the primary path (of the same LSP) is active, the following occurs:

1. CSPF identifies all fate-sharing groups that are associated with the primary path. CSPF does this by identifying all links and nodes that the primary path traverses and compiling group lists that contain at least one of the links or nodes. CSPF ignores the ingress and egress nodes in the search.
2. CSPF checks each link in the TED against the compiled group list. If the link is a member of a group, the cost of the link is increased by the cost of the group. If a link is a member of multiple groups, all group costs are added together.
3. CSPF performs the check for every node in the TED, except the ingress and egress node. Again, a node can belong to multiple groups, so costs are additive.
4. The router performs regular CSPF computation with the adjusted topology.

Example: Configure Fate Sharing

Configure fate-sharing groups thunder and shadow. Because shadow has no objects, it is ignored during processing.

```
[edit routing-options]
fate-sharing {
  group thunder {
    cost 20;                # Optional, default value is 1
    from 1.2.3.4 to 1.2.3.5; # A point-to-point link
    from 192.168.200.1;     # LAN interface
    from 192.168.200.2;     # LAN interface
    from 192.168.200.3;     # LAN interface
    from 192.168.200.4;     # LAN interface
    from 10.168.1.220;      # Router ID of a router node
    from 10.168.1.221;      # Router ID of a router node
  }
  group shadow {
    .....
  }
}
```

Configure All Other MPLS Routers for Signaled LSPs

To configure signaled LSPs on all MPLS routers that should participate in MPLS, you need to enable MPLS and RSVP on these routers, as described in “Minimum MPLS Configuration” on page 44 and “Enable RSVP” on page 79.

Enable RSVP

For all routers that should participate in signaled LSPs, you must enable RSVP because it is used to set up LSPs. To do this, include the following statements in the configuration. In general, we recommend that you enable RSVP on all router interfaces, except for those on the AS border:

```
[edit]
interfaces {
  interface-name {
    unit logical-unit-number {
      family mpls;
    }
  }
}
protocols {
  mpls {
    interface all;
  }
  rsvp {
    interface all;
  }
}
```

For more information about RSVP, see “RSVP Configuration Guidelines” on page 161.

Improve TED Accuracy with RSVP PathErr Messages

An essential element of RSVP-based traffic engineering is the TED. The TED contains a complete list of all network nodes and links participating in traffic engineering, and a set of attributes each of those links can hold. (For more information about the TED, see “Constrained-Path LSP Computation” on page 25.) One of the most important link attributes is bandwidth.

Bandwidth availability on links changes quickly as RSVP LSPs are established and terminated. It is likely that the TED will develop inconsistencies relative to the real network. These inconsistencies cannot be fixed by increasing the rate of Interior Gateway Protocol (IGP) updates.

Link availability can share the same inconsistency problem. A link that becomes unavailable can break all existing RSVP LSPs. However, its unavailability might not readily be known by the network.

When you configure the `rsvp-error-hold-time` statement, a source node (ingress of the RSVP LSPs) learns from the failures of its LSP by monitoring PathErr messages transmitted from downstream nodes. Information from the PathErr messages is incorporated into subsequent LSP computations, which can improve the accuracy and speed of LSP setup. Some PathErr messages are also used to update TED bandwidth information, reducing inconsistencies between the TED and the network.

You can control the frequency of IGP updates by using the `update-threshold` statement, which you configure at the `[edit protocols rsvp]` hierarchy level. See “Configure the RSVP Update Threshold on an Interface” on page 168.

PathErr Messages

PathErr messages report a wide variety of problems by means of different code and subcode numbers. You can find a complete list of these PathErr messages in RFC 2205 and RFC 3209.

When you configure the `rsvp-error-hold-time` statement, two categories of PathErr messages, which specifically represent link failures, are examined:

Link bandwidth is low for this LSP:

Requested bandwidth unavailable—code 1, subcode 2

These types of PathErr messages represent a global problem that affects all LSPs transiting the link. They indicate that the actual link bandwidth is lower than that required by the LSP, and that it is likely that the bandwidth information in the TED is an overestimate.

When this type of error is received, the available link bandwidth is reduced in the local TED database. This affects all future LSP computations.

Link unavailable for this LSP:

Admission Control failure—code 1, any subcode except 2

Policy Control failures—code 2

Service Preempted—code 12

Routing problem—no route available toward destination—code 24, subcode 5

These types of PathErr messages are generally pertinent to the specified LSP. The failure of this LSP does not necessarily imply that other LSPs could also fail. These errors can indicate maximum transfer unit (MTU) problems, service preemption (either manually initiated by the operator or by another LSP with a higher priority), that a next-hop link is down, that a next-hop neighbor is down, or service rejection because of policy considerations. It is best to route this particular LSP away from the link.

Identify the Problem Link

Each PathErr message includes the sender's IP address. This information is propagated unchanged toward the ingress router. A lookup in the TED can identify the node that originated the PathErr message.

Each PathErr message carries enough information to identify the RSVP session that triggered the message. If this is a transit router, it simply forwards the message. If this router is the ingress router (for this RSVP session), it has the complete list of all nodes and links the session should traverse. Coupled with the originating node information, the link can be uniquely identified.

Configure the Router to Improve TED Database Accuracy

To improve the accuracy of the TED database, configure the `rsvp-error-hold-time` statement. When this statement is configured, a source node (ingress of the RSVP LSPs) learns from the failures of its LSP by monitoring PathErr messages transmitted from downstream nodes. Information from the PathErr messages is incorporated into subsequent LSP computations, which can improve the accuracy and speed of LSP setup. Some PathErr messages also are used to update TED bandwidth information, reducing inconsistencies between the TED and the network.

To configure how long MPLS should remember RSVP PathErr messages and consider them in CSPF computation, include the `rsvp-error-hold-time` statement at the `[edit protocols mpls]` hierarchy level:

```
[edit protocols mpls]
rsvp-error-hold-time seconds;
```

The time can be a value from 1 to 240 seconds. The default is 25 seconds. Configuring a value 0 disables the monitoring of PathErr messages.

Examples: Configure Signaled LSPs

On the ingress router, create a constrained path LSP in which the JUNOS software makes all the forwarding decisions. When the LSP is successfully set up, a route toward 11.1.1.1/32 is installed in the `inet.3` table so that all BGP routes with matching BGP next-hop addresses can be forwarded through the LSP.

```
[edit]
interfaces {
  so-0/0/0 {
    unit 0 {
      family mpls;
    }
  }
}
protocols {
  rsvp {
    interface so-0/0/0;
  }
  mpls {
    label-switched-path to-hastings {
      to 11.1.1.1;
    }
    interface so-0/0/0;
  }
}
```

On the ingress router, create an explicit-path LSP and specify the transit routers between the ingress and egress routers. In this configuration, no constrained-path computation is performed. For the primary path, all intermediate hops are strictly specified so that its route cannot change. The secondary path must travel through router 14.1.1.1 first, then take whatever route is available to reach the destination. The remaining route taken by the secondary path is typically the shortest path computed by the IGP.

```
[edit]
interfaces {
  so-0/0/0 {
    unit 0 {
      family mpls;
    }
  }
}
protocols {
  rsvp {
    interface so-0/0/0;
  }
  mpls {
    path to-hastings {
      14.1.1.1 strict;
      13.1.1.1 strict;
      12.1.1.1 strict;
      11.1.1.1 strict;
    }
    path alt-hastings {
      14.1.1.1 strict;
      11.1.1.1 loose; # Any IGP route is acceptable
    }
    label-switched-path hastings {
      to 11.1.1.1;
      hop-limit 32;
      bandwidth 10m; # Reserve 10 mbps
      no-cspf;       # do not perform constrained-path computation
      primary to-hastings;
      secondary alt-hastings;
    }
  }
  interface so-0/0/0;
}
}
```

On the ingress router, create a constrained-path LSP in which the JUNOS software makes most of the forwarding decisions, taking into account the hop constraints listed in the path statements. The LSP is adaptive so that no bandwidth double-counting occurs on links shared by primary and secondary paths. To acquire the necessary link bandwidth, this LSP is allowed to preempt lower priority sessions. Finally, this path always keeps the secondary path in hot-standby state for quick failover.

```
[edit protocols]
mpls {
  path to-hastings {
    14.1.1.1 loose;
  }
  path alt-hastings {
    12.1.1.1 loose;
    11.1.1.1 strict;
  }
  label-switched-path hastings {
    to 11.1.1.1;
    bandwidth 10m;      # Reserve 10 mbps
    priority 0 0;       # Preemptive, but not preemptable
    adaptive;           # Set adaptivity
    primary to-hastings;
    secondary alt-hastings {
      standby;
      bandwidth 1m;     # Reserve only 1 Mbps for the secondary path
    }
  }
}
interface all;
}
```

On the ingress router, create a constrained-path LSP in which the JUNOS software makes most of the forwarding decisions for the primary path, subject to constraints of the path to-hastings, and in which the secondary path is an explicit path. The primary path must transit green or yellow links and must stay away from red links. The primary path is periodically recomputed and reoptimized. Finally, this path always keeps the secondary path in hot-standby state for quick failover.

When the LSP is up—either because the primary or secondary path is up, or because both paths are up—the prefix 16.0.0.0/8 is installed in the inet.3 table so that all BGP routes whose BGP next-hop falls within that range can use the LSP. Also, the prefix 17/8 is installed in the inet.0 table so that BGP can only resolve its next-hop through that prefix. The route also can be reached using traceroute or ping. These two routes are in addition to the 11.1.1.1/32 route.

```
[edit protocols]
mpls {
  admin-groups {
    green 1;
    yellow 2;
    red 3;
  }
  path to-hastings {
    14.1.1.1 loose;
  }
  path alt-hastings {
    14.1.1.1 strict;
    13.1.1.1 strict;
    12.1.1.1 strict;
    11.1.1.1 strict;
  }
  label-switched-path hastings {
    to 11.1.1.1;
    bandwidth 100m;
    install 16.0.0.0/8;           # in inet.3; cannot use to traceroute or ping
    install 17.0.0.0/8 active;   # installed in inet.0; can use to traceroute or ping
    primary to-hastings {
      admin-group {
        include [ green yellow ]; # further constraints for path computation
        exclude red;
      }
      optimize-timer 3600;       # reoptimize every hour
    }
    secondary alt-hastings {
      standby;
      no-cspf;                   # do not perform constrained-path computation
    }
  }
  interface all;
}
```

Configure MPLS over GRE Tunnels

MPLS LSPs can use generic routing encapsulation (GRE) tunnels to cross routing areas, Autonomous Systems, and ISPs. Bridging MPLS LSPs over an intervening IP domain is possible without disrupting the outlying MPLS domain.

LSPs can reach any destination that the GRE tunnels can reach. MPLS applications can be deployed without requiring all transit nodes to support MPLS, or requiring all transit nodes to support the same label distribution protocols (LDP or RSVP). If you use CSPF, you must configure OSPF or IS-IS through the GRE tunnel. Traffic engineering is not supported over GRE tunnels; for example, you cannot reserve bandwidth or set priority or preemption.

For more information about GRE tunnels, see the *JUNOS Internet Software Configuration Guide: Services Interfaces*.

Example: Configure MPLS over GRE Tunnels

To configure MPLS over GRE tunnels:

1. Enable family mpls under the GRE interface configuration.

```
[edit interfaces]
interface gr-1/2/0 {
  unit 0 {
    tunnel {
      source 192.168.1.1;
      destination 192.168.1.2;
    }
    family inet {
      address 5.1.1.1/30;
    }
    family iso;
    family mpls;
  }
}
```

2. Enable RSVP and MPLS over the GRE tunnel.

```
[edit protocols]
rsvp {
  interface gr-1/2/0.0;
}
mpls {
  .....
  interface gr-1/2/0.0;
}
```

3. Configure LSPs to travel through the GRE tunnel endpoint address.

```
[edit protocols]
mpls {
  label-switched-path gre-tunnel {
    to 5.1.1.2;
    .....
  }
}
```

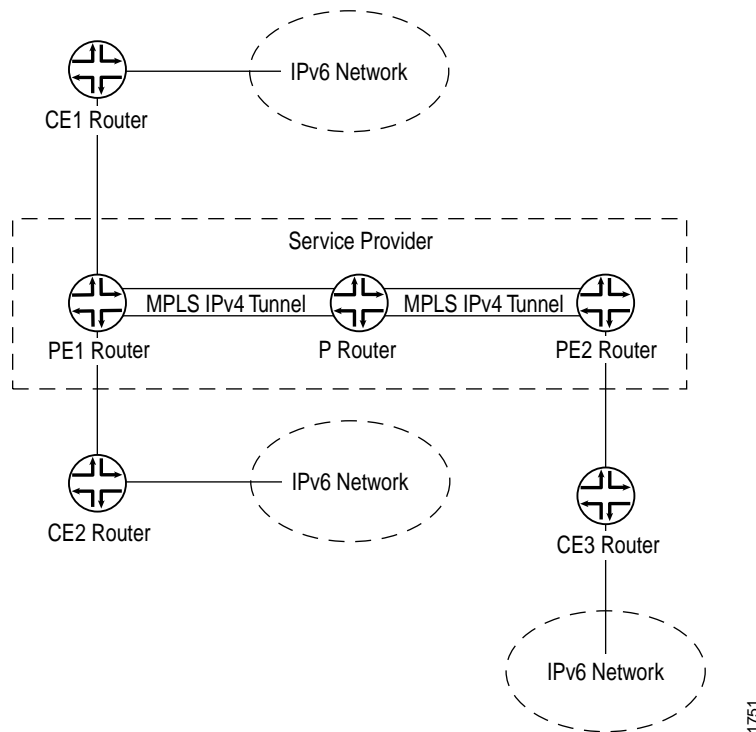
Standard LSP configuration options apply. If the routing table specifies that a particular route will traverse a GRE tunnel, the RSVP packets will traverse the tunnel as well.

Configure IPv6 Tunnels over MPLS

You can configure the JUNOS software to tunnel IPv6 over an MPLS-based IPv4 network. This allows you to interconnect a number of smaller IPv6 networks over an IPv4-based network core, giving you the ability to provide IPv6 service without having to upgrade the routers in your core network. Multiprotocol Border Gateway Protocol (MP-BGP) is configured to exchange routes between the IPv6 networks, and data is tunneled between these IPv6 networks by means of IPv4-based MPLS.

In Figure 18, Routers PE1 and PE2 are dual-stack BGP routers, meaning they have both IPv4 and IPv6 stacks. The PE routers link the IPv6 networks through the CE routers to the IPv4 core network. The CE routers and the PE routers connect through a link-layer that can carry IPv6 traffic. The PE routers use IPv6 on the CE router-facing interfaces and use IPv4 and MPLS on the core-facing interfaces. Note that one of the connected IPv6 networks could be the global IPv6 Internet.

Figure 18: IPv6 Networks Linked by MPLS IPv4 Tunnels



The two PE routers are linked through a MP-BGP session using IPv4 addresses. They use the session to exchange IPv6 routes with an IPv6 (label value 2) Address Family Identifier (AFI) and a Subsequence AFI (SAFI) (label value 4). Each PE router sets the next-hop for the IPv6 routes advertised on this session to its own IPv4 address. Because MP-BGP requires the BGP next hop to correspond to the same address family as the Network Layer Reachability Information (NLRI), this IPv4 address needs to be embedded within an IPv6 format.

1751

The PE routers can learn the IPv6 routes from the CE routers connected to them by means of the routing protocols RIPNG or MP-BGP, or through static configuration. Note that if BGP is used as the PE-router-to-CE-router protocol, the MP-BGP session between the PE router and CE router could occur over an IPv4 or IPv6 TCP session. Also, the BGP routes exchanged on that session would have SAFI unicast. You must configure an export policy to pass routes between IBGP and EBGP, and between BGP and any other protocol.

The PE routers have MPLS LSPs routed to each others' IPv4 addresses. IPv4 provides signalling for the LSPs using either LDP or RSVP. These LSPs are used to resolve the next-hop addresses of the IPv6 routes learned from MP-BGP. The next hops use IPv4-mapped IPv6 addresses, while the LSPs use IPv4 addresses.

The PE routers always advertise IPv6 routes to each other using a label value of 2, the explicit null label for IPv6 as defined in Internet RFC 3032. As a consequence, the forwarding next-hop(s) for the IPv6 routes learned from remote PE routers normally push two labels. The inner label is 2 (this label could be different if the advertising PE router is a non-Juniper Networks router) and the outer label is the LSP label. If the LSP is a single-hop LSP, then only Label 2 is pushed.

It is also possible for the PE routers to exchange plain IPv6 routes using SAFI unicast. However, there is one major advantage in exchanging labeled IPv6 routes. The penultimate-hop router for an MPLS LSP can pop the outer label and send the packet with the inner label as an MPLS packet. Without the inner label, the penultimate-hop router would need to discover whether the packet is an IPv4 or IPv6 packet in order to set the protocol field in the layer 2 header correctly.

When the PE1 router (in Figure 18 on page 86) receives an IPv6 packet from the CE1 router, it performs a lookup in the IPv6 forwarding table. If the destination matches a prefix learned from the CE2 router, then no labels need to be pushed and the packet is simply sent to the CE2 router. If the destination matches a prefix that was learned from the PE2 router, then the PE1 router pushes two labels onto the packet and sends it to the Provider router. The inner label is 2 and the outer label is the LSP label for the PE2 router.

Each Provider router in the service provider's network handles the packet as it would any MPLS packet, swapping labels as it passes from Provider router to Provider router. The penultimate-hop Provider router for the LSP pops the outer label and sends the packet to the PE2 router. When the PE2 router receives the packet, it recognizes the IPv6 explicit null label on the packet (Label 2). It pops this label and treats it as an IPv6 packet, performing a lookup in the IPv6 forwarding table and forwarding the packet to the CE3 router.

IPv6 over MPLS Documentation

Detailed information about the Juniper Networks implementation of IPv6 over MPLS is described in the following Internet drafts:

`draft-ietf-ngtrans-bgp-tunnel-04.txt`

`draft-ietf-ipngwg-addr-arch-v3-07.txt`

These Internet drafts are available on the IETF Web site at <http://www.ietf.org/>.

Configure an IPv4 MPLS Tunnel to Carry IPv6 Traffic

You must perform the following tasks to allow IPv6 to be carried over an IPv4 MPLS tunnel:

Configure IPv6 on Both Core and CE Router Facing Interfaces on page 88

Configure MPLS and RSVP between PE Routers on page 88

Enable IPv6 Tunneling in MPLS on page 89

Configure Multiprotocol BGP to Carry IPv6 Traffic on page 89

Configure IPv6 on Both Core and CE Router Facing Interfaces

In addition to configuring the family inet6 statement on all the CE-router-facing interfaces, you must also configure the statement on all the core-facing interfaces running MPLS. This is necessary because the router must be able to process any IPv6 packets it receives on these interfaces. You should not see any regular IPv6 traffic arrive on these interfaces, but you will receive MPLS packets tagged with Label 2. Even though Label 2 MPLS packets are sent in IPv4, these packets are treated as native IPv6 packets.

Configure the family inet6 statement at the [edit interfaces *interface-name*] hierarchy level:

```
[edit]
interfaces {
  interface-name {
    unit unit-number {
      family inet6 {
        address inet6-address;
      }
    }
  }
}
```

Configure MPLS and RSVP between PE Routers

For information about how to configure MPLS and RSVP, see the following sections:

Configure the Ingress Router for Signaled LSPs on page 45

Configure All Other MPLS Routers for Signaled LSPs on page 78

Enable RSVP on page 79

Enable IPv6 Tunneling in MPLS

When you enable IPv6 tunneling, an MPLS network can handle IPv6 routes.

Configure the `ipv6-tunneling` statement on the PE routers at the `[edit protocols mpls]` hierarchy level:

```
[edit]
protocols {
  mpls {
    ipv6-tunneling;
  }
}
```

You also need to configure IPv6 tunneling when you configure IPv6 VPNs. For more information, see the *JUNOS Internet Software Configuration Guide: VPNs*.

Configure Multiprotocol BGP to Carry IPv6 Traffic

At the `[family inet6]` hierarchy level in BGP, configure the `labeled-unicast` statement with the `explicit-null` option. As with regular BGP configuration, the family statement can be specified on a per-neighbor, per-group, or global basis, so it can be configured at the following hierarchy levels:

```
[edit protocols bgp]
```

```
[edit protocols bgp group group-name]
```

```
[edit protocols bgp group group-name neighbor neighbor-name]
```

Configuring these statements enables the IPv4 MPLS label to be removed at the destination PE router. The remaining label-less IPv6 packet can then be forwarded to the IPv6 network.

Configure the `labeled-unicast` statement as follows:

```
[edit]
family inet6 {
  labeled-unicast {
    explicit-null;
  }
}
```

LSP Attributes for GMPLS

When configuring GMPLS, use the LSP attributes statements, configured at the `[edit protocols mpls label-switched-path lsp-path-name lsp-attributes]` hierarchy level. For information, see “Configure MPLS Label-Switched Paths for GMPLS” on page 270.

