

Chapter 5

MPLS-Signaled LSPs Configuration Guidelines

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 “LDP Configuration Guidelines” on page 309.) 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:

- Configuring the Ingress Router for Signaled LSPs on page 64

- Configuring All Other MPLS Routers for Signaled LSPs on page 106

- Enabling RSVP on page 107

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

- Configuring Point-to-Multipoint LSPs on page 107

- Configuring MPLS Exception Monitoring on page 111

- Improving TED Accuracy with RSVP PathErr Messages on page 111

- Configuring MPLS over GRE Tunnels on page 118

- Configuring IPv6 Tunnels over MPLS on page 120

- Configuring ICMP Message Tunneling on page 124

- LSP Attributes for GMPLS on page 124

For configuration examples, see “Examples: Configuring Signaled LSPs” on page 113.

Configuring the Ingress Router for Signaled LSPs

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

Creating a Named Path on page 64

Creating an LSP on page 66

Configuring Alternate Backup Paths Using Fate Sharing on page 104

Creating 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 pathname 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 Protocols, Class of Service, and System Basics Command Reference*.

To create an empty path, create a named path by including the following form of the path statement. 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.

```
path path-name;
```

You can include this statement at the following hierarchy levels:

```
[edit protocols mpls]
```

```
[edit logical-routers logical-router-name protocols mpls]
```

To create a path in which you specify some or all transit routers in the path, include the following form of the path statement, specifying one address for each transit router:

```
path path-name {
  address | host name <strict | loose>;
}
```

You can include this statement at the following hierarchy levels:

```
[edit protocols mpls]
```

```
[edit logical-routers logical-router-name protocols mpls]
```

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 hostname 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. Ensuring that the incoming interface is the one specified is important 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, 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: Creating 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;
}
```

Creating 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:

```
label-switched-path lsp-path-name {
  disable;
  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 text;
  fast-reroute {
    bandwidth bps;
    (exclude group-names | no-exclude);
    hop-limit number;
    (include group-names | no-include);
  }
  from address;
  hop-limit number;
  install {
    destination-prefix/prefix-length <active>;
  }
  ldp-tunneling;
  link-protection;
  lsp-attributes {
    gpid (ethernet | hdlc | ipv4 | ppp);
    signal-bandwidth type;
    switching-type type;
  }
  metric number;
  no-cspf;
  no-decrement-ttl;
  node-link-protection;
  optimize-timer seconds;
  p2mp path-name;
  policing filter-name;
  preference preference;
  priority setup-priority hold-priority;
}
```

```

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;
}
(random | least-fill | most-fill);
(record | no-record);
retry-limit number;
retry-timer seconds;
secondary path-name {
  adaptive;
  admin-group {
    exclude group-names;
    include group-names;
  }
  bandwidth bps;
  class-of-service 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;
}
soft-preemption {
  cleanup-timer seconds;
}
standby;
to address;
traceoptions {
  file filename <replace> <size size> <files number> <no-stamp>
  <(world-readable | no-world-readable)>;
  flag flag <flag-modifier> <disable>;
}
}

```

You can include this statement at the following hierarchy levels:

[edit protocols mpls]

[edit logical-routers *logical-router-name* protocols mpls]

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. 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:

- 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:

- Configuring the Address of the Egress and Ingress Routers on page 70
- Configuring the Primary and Secondary LSPs on page 72
- Configuring the Description on page 76
- Configuring Fast Reroute on page 77
- Configuring Addresses to Associate with the LSP on page 78
- Configuring Path Connection Retry Information on page 79
- Configuring the LSP Metric on page 80
- Configuring CSPF Tie Breaking on page 82
- Configuring Load Balancing for MPLS LSPs on page 83
- Disabling Normal TTL Decrementing on page 85
- Configuring MPLS Soft Preemption on page 87
- Configuring Automatic Bandwidth Allocation on page 88

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

Disabling Constrained-Path LSP Computation on page 92

Configuring Administrative Groups on page 93

Configuring the LSP Preference on page 95

Configuring Path Route Recording on page 95

Configuring Class of Service for MPLS on page 95

Configuring Adaptive LSPs on page 98

Configuring Priority and Preemption on page 99

Optimizing Signaled LSPs on page 100

Configuring the Maximum Path Length on page 102

Configuring the Path Bandwidth on page 102

Configuring the Standby State on page 102

Configuring LSP Hold Time on page 104

Configuring LDP Tunneling on page 104

Configuring the Address of the Egress and Ingress Routers

The following sections describe how to configure the address for the egress and ingress routers:

Configuring the Address of the Egress Router on page 71

Configuring the Address of the Ingress Router on page 72

Configuring the Address of the Egress Router

When configuring an LSP, you must specify the address of the egress router by including the `to` statement:

```
to address;
```

You can include this statement at the following hierarchy levels:

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

```
[edit logical-routers logical-router-name protocols mpls label-switched-path
lsp-path-name]
```

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. This route can then be used by Border Gateway Protocol (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 to which the route leads. 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.

Configuring 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:

```
from address;
```

You can include this statement at the following hierarchy levels:

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

```
[edit logical-routers logical-router-name protocols mpls label-switched-path lsp-path-name]
```

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

Configuring 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 “Creating a Named Path” on page 64. Then apply the named path by including the primary or secondary statement. A named path can be referenced by any number of LSPs.

To configure primary and secondary paths for an LSP, complete the steps in the following sections:

Configuring Primary and Secondary Paths for an LSP on page 73

Configuring the Revert Timer on page 74

Specifying Path Selection on page 75

Configuring Primary and Secondary Paths for an LSP

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.

To configure primary and secondary paths, include the primary and secondary statements:

```

primary path-name {
    ...
}
secondary path-name {
    ...
}

```

You can include these statements at the following hierarchy levels:

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

```
[edit logical-routers logical-router-name protocols mpls label-switched-path
lsp-path-name]
```

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 “Configuring Path Connection Retry Information” on page 79.)

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.

Configuring the Revert Timer

For LSPs configured with both primary and secondary paths, it is possible to configure the revert timer. If a primary path goes down and traffic is switched to the secondary path, the revert timer specifies the amount of time (in seconds) that the LSP must wait before it can revert traffic back to a primary path. If during this time, the primary path experiences any connectivity problems or stability problems, the timer is restarted.

The range of values you can configure for the revert timer is 0 through 65,535 seconds. The default value is 60 seconds.

If you configure a value of 0 seconds, the traffic on the LSP, once switched from the primary path to the secondary path, remains on the secondary path permanently (until the network operator intervenes or until the secondary path goes down).

You can configure the revert timer for all LSPs on the router at the [edit protocols mpls] hierarchy level or for a specific LSP at the [edit protocols mpls label-switched-path *lsp-path-name*] hierarchy level.

To configure the revert timer, include the revert-timer statement:

```
revert-timer seconds;
```

For a list of hierarchy levels at which you can include this statement, see the summary section for this statement.

Specifying Path Selection

When you have configured both primary and secondary paths for an LSP, it might be necessary to ensure that only a specific path is used.

The select statement is optional. If you do not configure this statement, an automatic path selection algorithm is used. If you configure the select statement, you must specify either the manual option or the unconditional option. You cannot specify both.

The manual and unconditional options do the following:

manual—The path is immediately selected for carrying traffic as long as it is up and stable. Traffic is sent to other working paths if the current path is down or degraded (receiving errors). This parameter overrides all other path attributes except the select unconditional statement.

unconditional—The path is selected for carrying traffic unconditionally, regardless of whether the path is currently down or degraded (receiving errors). This parameter overrides all other path attributes.

Because the unconditional option switches to a path without regard to its current status, be aware of the following potential consequences when using this parameter:

If a path is not currently up when you enable the unconditional option, it can cause traffic disruptions. Ensuring that a path is functional before specifying the unconditional option helps to avoid this issue.

Once a path is selected because it has the unconditional option enabled, all other paths for the LSP are gradually cleared, including the primary and standby paths. No path can act as a standby to an unconditional path, so signaling those paths serves no purpose.

You can configure only one path for an LSP with the select manual statement and one path for an LSP with the select unconditional statement. These statements cannot be configured on two or more paths used by the same LSP.

You can enable and disable the manual and unconditional options for the select statement while LSPs and their paths are up and running without disrupting traffic.

To modify the behavior of the LSP such that a path is selected for carrying traffic if it is up and stable for at least the revert timer window, include the select statement with the manual option:

```
select {
    manual;
}
```

For a list of hierarchy levels at which you can include this statement, see the summary section for this statement.

To modify the behavior of the LSP such that a path is selected for carrying traffic unconditionally, regardless of whether the path is currently down or degraded, include the select statement with the unconditional option:

```
select {
    unconditional;
}
```

For a list of hierarchy levels at which you can include this statement, see the summary section for this statement.

Configuring the Description

You can provide a textual description for the LSP. 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.

To provide a textual description for the LSP, include the description statement:

```
description text;
```

You can include this statement at the following hierarchy levels:

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

```
[edit logical-routers logical-router-name protocols mpls label-switched-path
lsp-path-name]
```

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

Configuring 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.

To configure fast reroute on an LSP, include the `fast-reroute` statement on the ingress router:

```
fast-reroute {
    bandwidth bps;
    (exclude group-names | no-exclude);
    hop-limit number;
    (include group-names | no-include);
}
```

You can include this statement at the following hierarchy levels:

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

```
[edit logical-routers logical-router-name protocols mpls label-switched-path
lsp-path-name]
```

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 with the LSP itself. By default, the hop limit is set to 6. For example, if an LSP traverses 4 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 "Configuring Administrative Groups" on page 93.

Configuring 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 use the ping or traceroute command 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:

```
install {
    destination/mask <active>;
}
```

You can include this statement at the following hierarchy levels:

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

```
[edit logical-routers logical-router-name protocols mpls label-switched-path
lsp-path-name]
```

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 trace 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 point of presence (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 interior BGP (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.

Configuring 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:

```
retry-timer seconds;
```

For a list of hierarchy levels at which you can include this statement, see the summary section for this statement.

By default, no limit is set to the number of times an ingress router attempts to establish or reestablish a connection to the egress router using the primary path. To limit the number of attempts, include the `retry-limit` statement:

```
retry-limit number;
```

For a list of hierarchy levels at which you can include this statement, see the summary section for this statement.

The limit can be a value up to 10,000. 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.

Configuring 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:

Configuring a Dynamic LSP Metric on page 80

Configuring a Static LSP Metric on page 80

Configuring 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 Open Shortest Path First (OSPF), Intermediate System-to-Intermediate System (IS-IS), Routing Information Protocol (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.

Configuring a Static LSP Metric

You can manually assign a fixed metric value to an LSP. Once configured with the metric statement, the LSP metric is fixed and cannot change:

```
metric number;
```

You can include this statement at the following hierarchy levels:

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

```
[edit logical-routers logical-router-name protocols mpls label-switched-path lsp-path-name]
```

The LSP metric has several uses:

When there are parallel LSPs with the same egress router, the metrics are compared to determine 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 37), 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 by means of the metric value to determine which path is preferred.

By adjusting the LSP metric, you can force traffic to prefer LSPs, prefer the IGP path, or 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 a BGP multiple exit discriminator (MED) a stable metric to downstream neighbors. As long as Y remains reachable through the LSP, no changes are visible to downstream BGP neighbors.

Configuring 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 33.

You can configure one of the following statements (you can only configure one of these statements at a time) to alter the behavior of CSPF tie-breaking:

To configure a random tie-breaking rule for CSPF to use to choose among equal-cost paths, include the `random` statement:

```
random;
```

To prefer the path with the least-utilized links, include the `least-fill` statement:

```
least-fill;
```

To prefer the path with the most-utilized links, include the `most-fill` statement:

```
most-fill;
```

You can include each of these statements at the following hierarchy levels:

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

```
[edit logical-routers logical-router-name protocols mpls label-switched-path  
lsp-path-name]
```

Configuring Load Balancing for MPLS LSPs

Load balancing is used to evenly distribute traffic when:

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

There is a single next hop 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 next hops for a destination changes in any way, the next-hop address is reselected by means of 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, the IP payload, or the first and second MPLS labels and the IP payload. These configurations are described in the following sections:

Using the First MPLS Label in the Hash Key on page 83

Using the Second MPLS Label in the Hash Key on page 83

Using the IP Payload in the Hash Key on page 84

Using the First Two Labels and the IP Payload in the Hash Key on page 84

Configuring Load Balancing for MPLS LSPs without CSPF on page 84

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

Using the First MPLS Label in the Hash Key

To use the first MPLS label in 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;
```

Using the Second MPLS Label in the Hash Key

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;
```

Using the IP Payload in the Hash Key

To use the MPLS packet's IP payload (IP version 4 [IPv4] or IP version 6 [IPv6]) in the hash key, include the `label-1` and `payload` statements at the `[edit forwarding-options hash-key family mpls]` hierarchy level. Specify the `ip` option to the `payload` statement at the `[edit forwarding-options hash-key family mpls payload]` hierarchy level:

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



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), the packet is assumed to be an IP packet.

Using the First Two Labels and the IP Payload in the Hash Key

To use the first and second MPLS labels and the MPLS packet's IP payload in the hash key, include the `label-1`, `label-2`, and `payload` statements at the `[edit forwarding-options hash-key family mpls]` hierarchy level. Specify the `ip` option to the `payload` statement at the `[edit forwarding-options hash-key family mpls payload]` hierarchy level:

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



NOTE: This feature can be used on T-series and M320 routers only. If you configure this feature on an M-series router, only `label-1` and the IP payload are used in the hash key.

Configuring Load Balancing for MPLS 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, which compares IGP metrics alone. No consideration is given to bandwidth or congestion levels.

Disabling Normal TTL Decrementing

By default, the time to live (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 Internet Control Message Protocol (ICMP) error packet can 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 on transiting the LSP, making the LSP appear as a one-hop router to diagnostic tools, such as traceroute. Decrementing the TTL field by 1 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 from outside that LSP, traceroute sees the ingress router, although the egress router performs the TTL decrement. 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, 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.

```
no-decrement-ttl;
```

You can include this statement at the following hierarchy levels:

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

```
[edit logical-routers logical-router-name protocols mpls label-switched-path lsp-path-name]
```

Note that the RSVP object is proprietary to the JUNOS software and might not work with other software. This potential incompatibility applies only 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. 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.

```
no-propagate-ttl;
```

You can include this statement at the following hierarchy levels:

```
[edit protocols mpls]
```

```
[edit logical-routers logical-router-name protocols mpls]
```

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 that all routers are configured identically.

Configuring MPLS Soft Preemption

Soft preemption attempts to establish a new path for a preempted LSP before tearing down the original LSP. The default behavior is to tear down a preempted LSP first, signal a new path, and then reestablish the LSP over the new path. In the interval between when the path is taken down and the new LSP is established, any traffic attempting to use the LSP is lost. Soft preemption prevents this type of traffic loss. The trade-off is that during the time when an LSP is being soft preempted, two LSPs with their corresponding bandwidth requirements are used until the original path is torn down.

MPLS soft preemption is useful for network maintenance. For example, you can move all LSPs away from a particular interface, then take the interface down for maintenance without interrupting traffic. MPLS soft preemption is described in detail in Internet draft draft-ietf-mpls-soft-preemption-02.txt, *MPLS Traffic Engineering Soft Preemption*.

Soft preemption is a property of the LSP and is disabled by default. You configure it at the ingress of an LSP by including the soft-preemption statement:

```
soft-preemption;
```

You can include this statement at the following hierarchy levels:

```
[edit logical-routers logical-router-name protocols mpls label-switched-path  
lsp-path-name]
```

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

You can also configure a timer for soft preemption. The timer designates the length of time the router should wait before initiating a hard preemption of the LSP. At the end of the time specified, the LSP is torn down and resignaled. The soft-preemption cleanup timer has a default value of 30 seconds; the range of permissible values is 0 through 180 seconds. A value of 0 means that soft preemption is disabled. The soft-preemption cleanup timer is global for all LSPs.

Configure the timer by including the cleanup-timer statement:

```
cleanup-timer seconds;
```

You can include this statement at the following hierarchy levels:

```
[edit logical-routers logical-router-name protocols rsvp preemption soft-preemption]
```

```
[edit protocols rsvp preemption soft-preemption]
```



NOTE: Soft preemption cannot be configured on LSPs for which secondary paths or fast reroute has been configured. The configuration fails to commit.

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

To configure automatic bandwidth allocation, complete the steps in the following sections:

Configuring MPLS Statistics on page 89

Configuring the Maximum and Minimum Bounds of the LSP's Bandwidth on page 89

Configuring the Threshold for Automatic Bandwidth Adjustment on page 90

Configuring Passive Bandwidth Utilization Monitoring on page 90

Requesting an Automatic Bandwidth Allocation Adjustment on page 91

Configuring MPLS Statistics

To enable automatic bandwidth allocation, you first need to configure MPLS statistics. As part of this configuration, include the auto-bandwidth statement. 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 on specific LSPs.

To configure the MPLS and automatic bandwidth allocation statistics, include the statistics statement:

```
statistics {
  auto-bandwidth;
  file filename size size files number <no-stamp>;
  interval seconds;
}
```

You can include this statement at the following hierarchy levels:

```
[edit protocols mpls]
```

```
[edit logical-routers logical-router-name protocols mpls]
```

Configuring 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. Specify the bandwidth reallocation interval in seconds using the adjust-interval statement.

To configure automatic bandwidth allocation, include the auto-bandwidth statement:

```
auto-bandwidth {
  adjust-interval seconds;
  adjust-threshold percent;
  minimum-bandwidth bps;
  maximum-bandwidth bps;
}
```

You can include this statement at the following hierarchy levels:

```
[edit protocols mpls label-switched-path label-switched-path-name]
```

```
[edit logical-routers logical-router-name protocols mpls label-switched-path
label-switched-path-name]
```

Configuring 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 megabits per second (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.

To configure the threshold for automatic bandwidth adjustment, include the `adjust-threshold` statement:

```
adjust-threshold percent;
```

You can include this statement at the following hierarchy levels:

```
[edit protocols mpls label-switched-path lsp-name auto-bandwidth]
```

```
[edit logical-routers logical-router-name protocols mpls label-switched-path  
lsp-name auto-bandwidth]
```

Configuring Passive Bandwidth Utilization Monitoring

Use the `monitor-bandwidth` statement 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.

To configure passive bandwidth utilization monitoring, include the `monitor-bandwidth` statement:

```
monitor-bandwidth;
```

You can include this statement at the following hierarchy levels:

```
[edit protocols mpls label-switched-path lsp-name auto-bandwidth]
```

```
[edit logical-routers logical-router-name protocols mpls label-switched-path  
lsp-name auto-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.

Requesting an Automatic Bandwidth Allocation Adjustment

For MPLS LSP automatic bandwidth allocation adjustment, the minimum value for the adjustment interval is 5 minutes (300s). You might find it necessary to trigger a bandwidth allocation adjustment manually. The following describe a couple of scenarios when you might want to do this:

Testing automatic bandwidth allocation in a network lab.

When you configure the LSP for automatic bandwidth allocation in monitor mode (you have configured the monitor-bandwidth statement), you might want to initiate a bandwidth adjustment at your own discretion.

To use the request mpls lsp adjust-autobandwidth command, the following must be true:

Automatic bandwidth allocation must be enabled on the LSP.

The criteria required to trigger a bandwidth adjustment have been met (the difference between the adjust bandwidth and the current LSP path bandwidth is greater than the threshold limit).

A manually triggered bandwidth adjustment operates only on the active LSP path. Also, if you have enabled periodic automatic bandwidth adjustment, the periodic automatic bandwidth adjustment parameters (the adjustment interval and the maximum average bandwidth) are not reset after a manual adjustment.

For example, suppose the periodic adjust interval is 10 hours and there are currently 5 hours remaining before an automatic bandwidth adjustment is triggered. If you initiate a manual adjustment with the request mpls lsp adjust-autobandwidth command, the adjust timer is not reset and still has 5 hours remaining.

To manually trigger a bandwidth allocation adjustment, you need to use the request mpls lsp adjust-autobandwidth command. You can trigger the command for all affected LSPs on the router, or you can specify a particular LSP:

```
user@host> request mpls lsp adjust-autobandwidth
```

Once you execute this command, the automatic bandwidth adjustment validation process is triggered. If all the criteria for adjustment are met, the LSP's active path bandwidth is adjusted to the adjusted bandwidth value determined during the validation process.

Disabling Constrained-Path LSP Computation

If the IGP is a link-state protocol (such as IS-IS or OSPF) and 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 Routing Protocols Configuration Guide*.

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 Routing Protocols Configuration Guide*.

If IS-IS is enabled on a router or you enable OSPF traffic engineering extensions, MPLS performs the constrained-path LSP computation by default. For information on how constrained-path LSP computation works, see “Constrained-Path LSP Computation” on page 32.

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 “Optimizing Signaled LSPs” on page 100.

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 “Configuring Path Connection Retry Information” on page 79.

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:

```
no-cspf;
```

For a list of hierarchy levels at which you can include this statement, see the statement summary section for this statement.

Configuring 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 by including the `admin-groups` statement:

```
group-name group-value;
```

You can include this statement at the following hierarchy levels:

```
[edit protocols mpls]
```

```
[edit logical-routers logical-router-name protocols mpls]
```

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 by including the `admin-groups` statement:

```
admin-groups {
  best-effort 1;
  copper 2;
  silver 3;
  gold 4;
  violet 5;
}
```

You can include this statement at the following hierarchy levels:

```
[edit protocols mpls]
```

```
[edit logical-routers logical-router-name protocols mpls]
```

2. Define the administrative groups to which an interface belongs. You can assign multiple groups to an interface. Include the interface statement:

```
interface interface-name {
  admin-group [ group-names ];
}
```

You can include this statement at the following hierarchy levels:

```
[edit protocols mpls]
```

```
[edit logical-routers logical-router-name protocols mpls]
```

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, is 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. Include the `label-switched-path` statement:

```
label-switched-path lsp-path-name {
  to address;
  ...
  admin-group {
    exclude [ group-name group-name ... ];
    include [ group-name group-name ... ];
  }
  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 ... ];
    }
  }
}
```

You can include the `label-switched-path` statement at the following hierarchy levels:

```
[edit protocols mpls]
```

```
[edit logical-routers logical-router-name protocols mpls]
```

If you omit the `include` or `exclude` statements, the path computation proceeds unchanged. The path computation is based on the constrained-path LSP computation. For information on how the constrained-path LSP computation is calculated, see “How CSPF Selects a Path” on page 33.



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

Configuring 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 for RSVP LSPs is 7 and for LDP LSPs is 9. These preference values are lower (more preferred) than all learned routes except direct interface routes.

To change the default preference value, include the preference statement:

```
preference preference;
```

For a list of hierarchy levels at which you can include this statement, see the statement summary section for this statement.

Configuring Path Route Recording

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:

```
no-record;
```

For a list of hierarchy levels at which you can include the record (no-record) statement, see the statement summary section for the statement.

Configuring Class of Service for MPLS

The following sections provide an overview of MPLS class of service (CoS) and describe how to configure the MPLS CoS value:

Class of Service for MPLS Overview on page 96

Configuring the MPLS CoS Bits on page 96

Rewriting IEEE 802.1p Packet Headers with the MPLS CoS Value on page 97

Class of Service for MPLS Overview

When IP traffic enters an LSP tunnel, the ingress router marks all packets with a 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 by means of the CoS bits (also known as the EXP or experimental bits). For more information, see “Label Allocation” on page 28.

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 Network Interfaces and Class of Service Configuration Guide*.

Configuring the MPLS CoS Bits

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 Network Interfaces and Class of Service Configuration Guide* 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. A fixed CoS value means that all packets entering the LSP receive the same class of service.

To set a fixed CoS value, include the class-of-service statement:

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

For a list of hierarchy levels at which you can include this statement, see the statement summary section for this statement.

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 Network Interfaces and Class of Service Configuration Guide*.



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 Network Interfaces and Class of Service Configuration Guide*.

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.

Rewriting IEEE 802.1p Packet Headers with the MPLS CoS Value

For Ethernet interfaces installed on a T-series platform or M320 router 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). Rewriting these bits 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 about how to configure the EXP and IEEE 802.1p rewrite rules, see the *JUNOS Network Interfaces and Class of Service Configuration Guide*.

For information on the CoS bits, see “Label Allocation” on page 28 and “Configuring Class of Service for MPLS” on page 95.

Configuring Adaptive LSPs

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 levels, the 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.

To configure adaptive behavior for all LSP paths, include the `adaptive` statement in the LSP configuration:

```
adaptive;
```

You can include this statement at the following hierarchy levels:

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

```
[edit logical-routers logical-router-name protocols mpls label-switched-path lsp-path-name]
```

If you specify the adaptive statement at the primary or secondary hierarchy level, adaptive behavior is enabled only on the path on which it is specified. Bandwidth double-counting occurs between different paths.

To configure adaptive behavior for either the primary or secondary level, include the adaptive statement:

```
adaptive;
```

You can include this statement at the following hierarchy levels:

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

```
[edit logical-routers logical-router-name protocols mpls label-switched-path  
lsp-path-name (primary | secondary)]
```

Configuring 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:

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

For a list of hierarchy levels at which you can include this statement, see the statement summary section for this statement.

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.

Optimizing 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 “Disabling Constrained-Path LSP Computation” on page 92.

To enable path reoptimization, include the *optimize-timer* statement:

```
optimize-timer seconds;
```

For a list of hierarchy levels at which you can include this statement, see the statement summary section for this statement.

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. The effect on congestion is determined by a comparison of the percentage of available bandwidth on each link traversed by the new paths to the 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 include the `optimize-aggressive` statement:

```
optimize-aggressive;
```

You can include this statement at the following hierarchy levels:

```
[edit protocols mpls]
```

```
[edit logical-routers logical-router-name protocols mpls]
```

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.

Configuring 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:

```
hop-limit number;
```

For a list of hierarchy levels at which you can include this statement, see the statement summary section for this statement.

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

Configuring 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:

```
bandwidth bps;
```

For a list of hierarchy levels at which you can include this statement, see the statement summary section for this statement.

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. RSVP's reservation scheme is used to reserve this capacity. Any failure in bandwidth reservation (such as failures at RSVP policy control or admission control) might cause the LSP setup to fail.

Configuring 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:

```
standby;
```

You can include this statement at the following hierarchy levels:

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

```
[edit logical-routers logical-router-name protocols mpls label-switched-path  
lsp-path-name secondary]
```

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. Although 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 following hierarchy levels, the hot-standby state is activated on all secondary paths configured beneath that hierarchy level:

```
[edit protocols mpls]
```

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

```
[edit logical-routers logical-router-name protocols mpls]
```

```
[edit logical-routers logical-router-name protocols mpls label-switched-path lsp-path-name]
```

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.

Configuring 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 affects only IS-IS advertisements of the LSP; other routing software and hardware react immediately to LSP transitions.

To damp LSP transitions, include the `advertise-hold-time` statement:

```
advertise-hold-time seconds;
```

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

You can include this statement at the following hierarchy levels:

```
[edit protocols mpls]
```

```
[edit logical-routers logical-router-name protocols mpls]
```

Configuring 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.

Configuring Alternate Backup Paths Using Fate Sharing

You can create a database of information that CSPF uses to compute one or more backup paths 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 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:

```
fate-sharing {
  group group-name {
    cost value;
    from address <to address>;
  }
}
```

For a list of hierarchy levels at which you can include this statement, see the statement summary section for this statement.

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.

Non-point-to-point links—Include links on a LAN interface (such as Gigabit Ethernet interfaces) or nonbroadcast multiaccess (NBMA) interfaces (such as Asynchronous Transfer Mode [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 that share the same fiber conduit, all optical channels that share the same fiber, all links that connect to the same LAN switch, all equipment that shares 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 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: Configuring Fate Sharing

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

```
[edit routing-options]
fate-sharing {
  group east {
    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 west {
    .....
  }
}
```

Configuring 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 62 and “Enabling RSVP” on page 107.

Enabling RSVP

For all routers that should participate in signaled LSPs, you must enable RSVP because it is used to set up LSPs. To enable RSVP, include the following statements in the configuration. In general, we recommend that you enable RSVP on all router interfaces, except those on the autonomous system (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 255.

Configuring Point-to-Multipoint LSPs

A point-to-multipoint MPLS LSP is an RSVP LSP with multiple destinations. By taking advantage of the MPLS packet replication capability of the network, point-to-multipoint LSPs avoid unnecessary packet replication at the ingress router.

The following sections describe how to configure point-to-multipoint LSPs:

Configuring Point-to-Multipoint LSPs on page 107

Example: Configuring Point-to-Multipoint LSPs on page 109

Configuring Link Protection for Point-to-Multipoint LSPs on page 110

Configuring Point-to-Multipoint LSPs

To configure point-to-multipoint LSPs, you begin by configuring the primary point-to-multipoint LSP, which carries traffic from the ingress router. The configuration of the primary point-to-multipoint LSP is similar to a signaled LSP. See “Configuring the Ingress Router for Signaled LSPs” on page 64 for more information. In addition to the conventional LSP configuration, you need to specify a path name for the primary point-to-multipoint LSP by including the `p2mp` statement:

```
p2mp path-name;
```

You can include this statement at the following hierarchy levels:

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

```
[edit logical-routers logical-router-name protocols mpls label-switched-path
lsp-name]
```

The primary point-to-multipoint LSP sends traffic to two or more sub-LSPs carrying traffic to each of the egress provider edge (PE) routers. In the configuration for each of these sub-LSPs, the point-to-multipoint LSP path name you specify must be identical to the path name configured for the primary point-to-multipoint LSP.

To associate a sub-LSP with the primary point-to-multipoint LSP, specify the point-to-multipoint pathname by including the `p2mp` statement:

```
p2mp path-name;
```

You can include this statement at the following hierarchy levels:

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

```
[edit logical-routers logical-router-name protocols mpls label-switched-path
lsp-name]
```

Example: Configuring Point-to-Multipoint LSPs

The following example shows a configuration based on the topology shown in Figure 18 on page 54. There are four sub-LSPs, each belonging to a single point-to-multipoint LSP called p2mp-lsp-sample.

```
[edit protocols mpls]
label-switched-path sub-LSP-to-PE2 {
  to 10.255.235.25;
  p2mp p2mp-lsp-sample;
  primary path1;
}
label-switched-path sub-LSP-to-PE2 {
  to 10.255.235.25;
  p2mp p2mp-lsp-sample;
  primary path2;
}
label-switched-path sub-LSP-to-PE3 {
  to 10.255.241.34;
  p2mp p2mp-lsp-sample;
  primary path3;
}
label-switched-path sub-LSP-to-CE4 {
  to 10.255.244.125;
  p2mp p2mp-lsp-sample;
  primary path4;
}
```

Configuring Link Protection for Point-to-Multipoint LSPs

Link protection helps to ensure that traffic going over a specific interface to a neighboring router can continue to reach this router if that interface fails. When link protection is configured for an interface and a point-to-multipoint LSP that traverses this interface, a bypass LSP is created that will handle this traffic if the interface fails. The bypass LSP uses a different interface and path to reach the same destination.

To extend link protection to all of the paths used by a point-to-multipoint LSP, link protection must be configured on each router that each sub-LSP traverses.

To enable link protection on point-to-multipoint LSPs, complete the following steps:

1. Configure link protection on each sub-LSP. To configure link protection, include the link-protection statement:

```
link-protection;
```

You can include this statement at the following hierarchy levels:

```
[edit protocols mpls lsp sub-lsp-name]
```

```
[edit logical-routers logical-router-name protocols mpls lsp sub-lsp-name]
```

2. Configure link protection for each RSVP interface on each router that the sub-LSP traverses. For information on how to configure link protection on the RSVP interfaces, see “Configuring Link Protection on the Interfaces Used by the LSPs” on page 266.

For more information on how to configure link protection, see “Configuring Node Protection or Link Protection” on page 265.

Configuring MPLS Exception Monitoring

You can process MPLS packets that have not been assigned label values and have no corresponding entry in the mpls.0 routing table. This allows you to assign a default route to unlabeled MPLS packets.

To configure a default label value for MPLS packets, include the default-route statement:

```
default-route {
  (next-hop (address | interface-name | address/interface-name) |
  (reject | discard);
  (pop | (swap <out-label>);
  class-of-service value;
  preference preference;
  type type;
}
```

You can include this statement at the following hierarchy levels:

```
[edit protocols mpls interface interface-name label-map]
```

```
[edit logical-routers logical-router-name protocols mpls interface interface-name
label-map]
```

Improving 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 32.) 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 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. See “Configuring the RSVP Update Threshold on an Interface” on page 264.

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, *Resource Reservation Protocol (RSVP), Version 1, Functional Specification* and RFC 3209, *RSVP-TE: Extensions to RSVP for LSP Tunnels*.

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, affecting 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.

Identifying 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.

Configuring the Router to Improve TED Accuracy

To improve the accuracy of the TED, 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:

```
rsvp-error-hold-time seconds;
```

You can include this statement at the following hierarchy levels:

```
[edit protocols mpls]
```

```
[edit logical-routers logical-router-name protocols mpls]
```

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: Configuring Signaled LSPs

The following examples illustrate how to configure signaled LSPs:

Example: Constrained-Path LSP, JUNOS Makes All Forwarding Decisions on page 114

Example: Explicit-Path LSP on page 115

Example: Constrained-Path LSP, JUNOS Makes Most Forwarding Decisions, Hop Constraints Accounted For on page 116

Example: Constrained-Path LSP, JUNOS Makes Most Forwarding Decisions, Secondary Path Is Explicit on page 117

Example: Constrained-Path LSP, JUNOS Makes All Forwarding Decisions

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;
  }
}
```

Example: Explicit-Path LSP

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;
  }
}
```

Example: Constrained-Path LSP, JUNOS Makes Most Forwarding Decisions, Hop Constraints Accounted For

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;
```

Example: Constrained-Path LSP, JUNOS Makes Most Forwarding Decisions, Secondary Path Is Explicit

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 resolve only its next hop through that prefix. The route also can be reached with 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 {          # further constraints for path computation
        include [ green yellow ];
        exclude red;
      }
      optimize-timer 3600;  # reoptimize every hour
    }
    secondary alt-hastings {
      standby;
      no-cspf;              # do not perform constrained-path computation
    }
  }
  interface all;
}
```

Configuring 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 Services Interfaces Configuration Guide*.

Example: Configuring 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;
    ...
  }
}
```

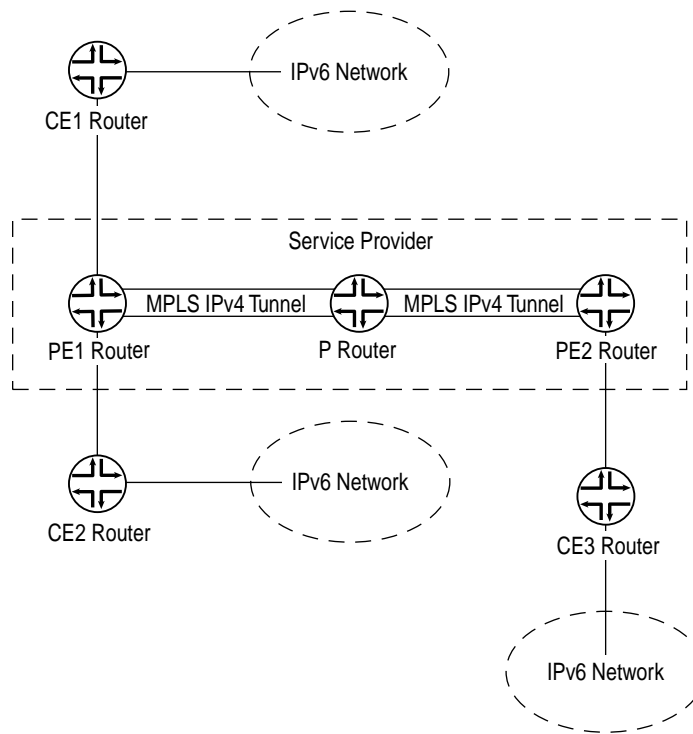
Standard LSP configuration options apply. If the routing table specifies that a particular route will traverse a GRE tunnel, the RSVP packets will traverse the tunnel as well.

Configuring IPv6 Tunnels over MPLS

You can configure the JUNOS software to tunnel IPv6 over an MPLS-based IPv4 network. This configuration 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 19, 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 customer edge (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 19: IPv6 Networks Linked by MPLS IPv4 Tunnels



1751

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 (value 2) address family indicator (AFI) and a Subsequent AFI (SAFI) (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.

The PE routers can learn the IPv6 routes from the CE routers connected to them by means of the routing protocols RIP next generation (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 Transmission Control Protocol (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 external BGP (EBGP), and between BGP and any other protocol.

The PE routers have MPLS LSPs routed to each others' IPv4 addresses. IPv4 provides signaling for the LSPs by means of 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 RFC 3032, *MPLS Label Stack Encoding*. As a consequence, each of the forwarding next hops 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 not a Juniper Networks routing platform), 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 then 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 to set the protocol field in the Layer 2 header correctly.

When the PE1 router in Figure 19 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 Standards

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

Internet draft draft-ietf-ngtrans-bgp-tunnel-04.txt, *Connecting IPv6 Islands Across IPv4 Clouds with BGP* (expires July 2002)

Internet draft draft-ietf-ipngwg-addr-arch-v3-07.txt, *IP Version 6 Addressing Architecture* (expires April 2002)

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

Configuring 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:

Configuring IPv6 on Both Core- and CE Router-Facing Interfaces on page 122

Configuring MPLS and RSVP Between PE Routers on page 123

Enabling IPv6 Tunneling in MPLS on page 123

Configuring Multiprotocol BGP to Carry IPv6 Traffic on page 123

Configuring 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. Both configurations are 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:

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

You can configure these statements at the following hierarchy levels:

```
[edit logical-routers logical-router-name interfaces interface-name]
```

```
[edit interfaces interface-name]
```

Configuring MPLS and RSVP Between PE Routers

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

Configuring the Ingress Router for Signaled LSPs on page 64

Configuring All Other MPLS Routers for Signaled LSPs on page 106

Enabling RSVP on page 107

Enabling IPv6 Tunneling in MPLS

You enable IPv6 tunneling by including the `ipv6-tunneling` statement in the configuration for the PE routers. This statement allows IPv6 routes to be resolved over an MPLS network by converting all routes stored in the `inet.3` routing table to IPv4-compatible IPv6 addresses and then copying them into the `inet6.3` routing table. This routing table can be used to resolve next hops for both `inet6` and `inet6-vpn` routes.

To configure IPv6 tunneling, include the `ipv6-tunneling` statement on the PE routers:

```
ipv6-tunneling;
```

You can include this statement at the following hierarchy levels:

```
[edit protocols mpls]
```

```
[edit logical-routers logical-router-name protocols mpls]
```

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

Configuring 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]
```

```
[edit logical-routers logical-router-name protocols bgp]
```

```
[edit logical-routers logical-router-name protocols bgp group group-name]
```

```
[edit logical-routers logical-router-name 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 IPv6 packet without a label can then be forwarded to the IPv6 network.

Configure the labeled-unicast statement as follows:

```
family inet6 {
    labeled-unicast {
        explicit-null;
    }
}
```

Configuring ICMP Message Tunneling

When you configure MPLS to tunnel through a routing domain, it is difficult to route a fragmented packet to its source address; for example, when the IP addresses carried in a packet are private (not globally unique) and MPLS is used to tunnel the packets through a public backbone.

When you configure ICMP message tunneling, an Internet Control Message Protocol (ICMP) message is sent to the source of a packet. The label stack is copied from the original packet to the ICMP message. The ICMP message is then label switched across the network. This causes the message to go to the original packet's destination, rather than its source. Unless the message is label switched all the way to the destination host, it ends up unlabeled in a router that does know the source of the original packet, at which point the message is sent in the proper direction.

ICMP message tunneling can be useful for debugging and tracing purposes if the ICMP message is one of either of the following types of messages:

Time exceeded

Destination unreachable because fragmentation needed and DF set

To configure ICMP message tunneling, include the `icmp-tunneling` statement:

```
icmp-tunneling;
```

You can configure these statements at the following hierarchy levels:

```
[edit logical-routers logical-router-name protocols mpls]
```

```
[edit protocols mpls]
```

LSP Attributes for GMPLS

When configuring GMPLS, use the LSP attributes statements. For information, see "Configuring MPLS LSPs for GMPLS" on page 412.