Planning Intermediate Route Selection

When a PNNI network node receives a call request, there can be multiple routes available that meet the quality of service (QoS) requirements for the call. This chapter describes how PNNI selects a route from multiple acceptable routes, and it describes parameters that you can modify to control route selection.

How MGX and SES Nodes Select Routes

MGX and SES nodes provide for the following PNNI route sources:
- Manually defined preferred routes (Release 3 and later)
- Pre-calculated routing tables called shortest path tables
- On-demand routes calculated from PNNI database entries

The following sections describe the link and route metrics used during routing, how shortest path table routing works, and how on-demand routing works. At the end of this chapter is a section on additional routing feature provided by MGX and SES nodes.

Link and Route Metrics

Most route metrics are calculated based on the link metrics for each link along the route. Because of this, link and route metrics often use the same name or similar names. This can be confusing if you do not consider the context in which the terms are used. Link metrics apply when configuring an individual link or when choosing between two or more links. Route metrics apply when configuring a connection or choosing between two or more routes.

The following sections introduce some of the most common link and route metrics and explain the differences between their use as either link or route metrics.

Administrative Weight

Administrative weight (AW) is a configurable cost that can be defined for each link in a PNNI network. The default link AW is 5040. There is no significance to the cost units. What is important is how the cost relates to that for other links in the network. For example, if two parallel links between two nodes have different costs, and if the link selection criteria is set to use the link with the lowest AW, the link with the smallest AW is chosen.
You can change the AW on links to control network traffic. For example, you can reduce traffic on a backup link by increasing the AW to more than that on the desirable link. If the desirable link fails, the backup link becomes the lowest cost link and becomes available.

When AW is applied to a route, it is sometimes called the cumulative AW and is the sum of the AW values assigned to all links along a route. Some operations calculate the cumulative AW from the source to the destination, and other operations calculate the round trip cumulative AW. For example, if all links in a network use the default link AW, the source to destination AW for a route that uses two links is 10080. The round trip AW for the same route is 20160.

If you leave the AW set to the default value on all network links, routing using the lowest AW is the same as routing using the fewest hops. A hop is a connection segment through a node. Changing the AW on a link gives you the opportunity to make that link more or less desirable for routing.

**Cell Transfer Delay**

Cell transfer delay (CTD) is the measure of the delay an ATM cell encounters as it passes through an interface. Since each link has an interface at each end, each link CTD is the sum of the CTD at each end of the link.

The route CTD is the sum of the CTD values for all links through which the route passes and represents the time interval between a cell exiting the source node and entering the destination node.

The CTD used in MGX and SES nodes is a static value that is set by Cisco according to PNNI 1.0 standard and is based on the speed of the interface. Faster interfaces will have lower CTD values, and slower interfaces will have higher values.

---

**Note**

Because the CTD is defined according to the PNNI 1.0 standard, the CTD for any specific link speed should match the CTD assigned to third-party interfaces that use that link speed.

**Cell Delay Variation**

Cell delay variation (CDV) is a measurement of the variation in CTD over links and through nodes. The route CDV is equal to the largest CDV along a route.

The CDV used in MGX and SES nodes is a static value that is set by Cisco and is based on the type of interface and node.

**Available Cell Rate**

Available cell rate (AvCR) is a dynamically generated value that indicates how much of the link bandwidth is available for the requested service class. AvCR is measured in cells per second (cps).

You cannot configure the AvCR for a link, but you can configure a parameter called the overbooking factor, which can change how the AvCR is advertised for new calls. After the PNNI controller calculates the AvCR for a route, it applies the overbooking factor to the AvCR before advertising the AvCR. The purpose of the overbooking factor is to allow you to purposely under book or over book a link.

For example, if link users are reserving more bandwidth than they actually need, bandwidth is being wasted. Overbooking allows you to make the wasted bandwidth available to other users. For example, if you estimate that 30% of the link bandwidth is not being used, you can configure the overbooking factor so that the advertised AvCR is 30% higher than the actual value. This enables the PNNI controller to...
route more calls for the link. Of course, if link users suddenly start using all link resources, some user-compliant traffic may be discarded when congestion occurs. Bandwidth overbooking can be configured on a per-service-class-basis for each interface in the node.

**Note**

Beginning with Release 3.0, Cisco MGX and SES nodes also support connection overbooking, which is configured with the `addcon` command. When per-service-class overbooking and connection based overbooking are both configured, both are applied simultaneously to each affected connection.

For more information, see the *Cisco MGX 8800/8900 Series Configuration Guide, Release 5.2* or see the appropriate service module configuration guide.

For CBR, rtVBR, and nrtVBR traffic, the advertised AvCR represents the bandwidth available for calls. For ABR traffic, AvCR is the capacity available for minimum cell rate (MCR) reservation. AvCR does not apply to UBR traffic.

The AvCR for a route is equal to the lowest link AvCR along the route.

**Maximum Cell Rate**

The maximum cell rate (maxCR) is a static value that is configured for each logical interface and can be configured separately for each service class. The maxCR represents the maximum throughput available for PNNI connections and cannot be modified by the overbooking factor. To block traffic for a particular service class over a link, set the maxCR for that service class to 0.

The maxCR for a route is equal to the lowest link maxCR along the route.

**Shortest Path Table Routing**

Most routing attempts begin with a search for a route in the shortest path tables. The following sections introduce the shortest path tables and explain how the tables are used by SVCs, SVPs, SPVCs, and SPVPs.

**The Shortest Path Tables**

The PNNI routing protocol automatically builds shortest path tables (SPTs) that list optimized routes for each destination address. When an MGX or SES node receives a call request, it compares the destination ATM address with the addresses and address prefixes in the node’s routing tables. The node looks for a match between the first 19 bytes of the destination address and the address prefixes in its database. The longest match determines the routes that are eligible. If there is just one route for the longest matching entry, and if that route meets the QoS requirements for the call, that is the route selected.

When multiple routes are available for the longest match, other route selection parameters are used to determine the optimum route.

**Note**

Border nodes can be configured with a 0-length prefix, which matches all ATM addresses. This 0-length prefix serves as a default destination or route for all calls that do not match up to a longer ATM address or prefix within the PNNI network. When a border nodes uses AINI or IISP links to communicate with an external network, the use of the 0-length prefix allows the administrator to specify that all calls that do not match a longer prefix should be routed to the external network. If the 0-length prefix is not used, the administrator must manually configure static addresses for all external destinations.
MGX and SES nodes generate routing tables using PTSE information from other nodes and the Dijkstra SPF Algorithm. The pre-computed routing tables are derived by applying the following information from the PTSEs:

- Destination address
- AW
- CTD
- CDV
- Available bandwidth
- Available logical connection numbers (LCNs)
- Port ID

The end result is the set of SPTs shown in Table 4-1.

### Table 4-1 Pre-calculated Routing Tables

<table>
<thead>
<tr>
<th>Traffic Metric</th>
<th>Class of Service Tables</th>
</tr>
</thead>
<tbody>
<tr>
<td>AW</td>
<td>CBR, rt-VBR, nrt-VBR, ABR, UBR</td>
</tr>
<tr>
<td>CTD</td>
<td>CBR, rt-VBR, nrt-VBR</td>
</tr>
<tr>
<td>CDV</td>
<td>CBR, rt-VBR</td>
</tr>
</tbody>
</table>

The SPTs can be divided into the three groups listed in the Traffic Metric column in Table 4-1. For each traffic metric, a class of service SPT is created for each class of service listed in the Class of Service Tables column. The service classes are defined in Table 4-2.

### Table 4-2 Supported Service Classes for MGX and SES Nodes

<table>
<thead>
<tr>
<th>Service Class</th>
<th>Acronym Definition</th>
<th>Guidelines</th>
</tr>
</thead>
<tbody>
<tr>
<td>CBR</td>
<td>Constant bit rate</td>
<td>Use to limit connections to a static amount of bandwidth that is continuously available until the connection is torn down. The amount of bandwidth is characterized by the peak cell rate (PCR) value.</td>
</tr>
<tr>
<td>rt-VBR</td>
<td>Real-time variable bit rate</td>
<td>Use for real-time applications that require tightly constrained delay and delay variation (voice/video applications). Category characterized in terms of a PCR, sustainable cell rate (SCR), and maximum burst size (MBS).</td>
</tr>
<tr>
<td>nrt-VBR</td>
<td>Non-real-time variable bit rate</td>
<td>Use for non-real-time applications with bursty traffic. Category is characterized in terms of a PCR, SCR, and MBS.</td>
</tr>
<tr>
<td>ABR</td>
<td>Available bit rate</td>
<td>Use to allow ATM layer transfer characteristics provided by the network to change after the connection is established. Flow control mechanism is specified.</td>
</tr>
<tr>
<td>UBR</td>
<td>Unspecified bit rate</td>
<td>Use for unspecified bit-rate ranges. This setting provides only maximum bit-rate configuration—no bit rate is guaranteed.</td>
</tr>
</tbody>
</table>

Each class of service SPT is simply a list of the shortest paths for a particular routing metric to all known destinations. AW SPTs list the shortest paths or routes based on the lowest cumulative AW, and CTD SPTs list the shortest routes based on the lowest cumulative CTD.
How MGX and SES Nodes Select Routes

The number of shortest paths stored in a SPT for any destination depends on whether there are multiple routes with the lowest routing metric value. For example, if three routes to a destination all have the same minimum CDV value, all three routes are listed in the CDV table for the appropriate class of service. There is also a range option that you can use to make the SPTs store routes with similar values. For example, you can configure the switch to store routes that are within 5 percent of the shortest route in the table. Up to five routes can be listed in a SPT for a destination.

The default configuration of MGX and SES nodes creates all 10 class of service tables. If you do not plan to use the routing tables for a particular routing metric, you can save processor resources by disabling the construction and maintenance of the appropriate routing metric SPTs (using the `cnfpnni-routing-policy` command).

### How SVCs and SVPs use the SPTs

SVCs and SVPs are initiated by CPE using UNI connections to the switch. UNI versions 3.0 and 3.1 cannot request a CTD or CDV value for a connection, so all UNI 3.0 and 3.1 connections are routed using the AW SPTs.

UNI 4.0 connections can request CTD and CDV values for a connection. UNI 4.0 connections use the SPTs in one of the following ways:

- If no CTD or CDV value is requested for the connection, the connection uses a route from the AW SPT for the appropriate class of service.
- If a CTD or CDV value is requested for the connection, the connection uses a route from the appropriate CTD or CDV SPT for the appropriate class of service.
- If both a CTD and a CDV value is requested for the connection, the connection uses a route from the CTD SPT for the appropriate class of service. The route chosen is a route that conforms to the CTD and CDV values requested. If a conforming route is not available in the SPT, on demand routing is used to find a conforming route.

### How SPVCs and SPVPs use the SPTs

The default configuration for each SPVC and SPVP uses the appropriate AW SPT for each class of service. However, you can configure requested values for AW, CTD, and CDV for each connection using the `addcon` and `cnfcon` commands.

If multiple routing metrics are specified for an SPVC or SPVP, the switch searches the SPTs for conforming routes according to following priorities:

1. AW
2. CTD
3. CDV

For example if all three routing metrics are specified, the switch searches for conforming routes in the AW SPTs. If CTD and CDV are specified, the switch searches the CTD SPTs.

Any route selected from the SPTs must conform to all specified metrics. If a conforming route is not available in chosen SPT, on demand routing is used to find a conforming route.

On PXM1E cards and service modules, you can change this with the `addcon` command.
How MGX and SES Nodes Select Links

On-Demand Routing

When the SPTs cannot produce a route for a connection, the switch performs on-demand routing. A SPT can fail to produce a route because the shortest route or routes in the table have failed. On-demand routing is also required when a connection specifies multiple routing metrics and the SPT routes do not conform to all of the metrics.

During on-demand routing, the switch searches the PNNI database for routes that match the specified criteria. On demand routing takes more time than SPT routing. However, on-demand routing can access more of the PNNI database and select better routes.

As a switch administrator, you can choose what action the controller takes when it discovers the first acceptable on-demand route. You can configure the controller for first fit, which produces the fastest route selection, or you can configure the switch for best fit. When the controller is configured for first fit on-demand route selection, it selects the first route that satisfies all connection requirements. When the controller is configured for best fit on-demand route selection, it identifies all routes that meet the call requirements, and then it chooses the route based on the setting of the load balancing option.

Load Balancing for SPT and On-Demand Routing

The load balancing option, which applies to SPT routing and on-demand routing, is a configurable parameter that allows you to control how a route is chosen when multiple routes offer the same level of service. You can configure the load balancing option to choose randomly from multiple routes or choose according to the best AvCR. If you select the random method, the PNNI controller considers the conforming routes equal and balances the load by randomly assigning calls to each. If you choose the route based on the AvCR, the route with the highest available cell rate is chosen.

How MGX and SES Nodes Select Links

The SPTs are built by calculating routes that are optimized for lowest AW, CTD, or CDV. However, for most service classes, each link along a route must conform to additional parameters. If no route is found in the SPTs, on-demand routing must be used to calculate a conforming route from the PNNI database.

The link parameter requirements for a service class establishes the quality of service (QoS) required for a call. Table 4-3 shows the link parameters that must be satisfied for each service class.

Table 4-3 Link Selection Parameters Required for Various Classes of Service

<table>
<thead>
<tr>
<th>Service Class</th>
<th>Address</th>
<th>AW</th>
<th>maxCR</th>
<th>AvCR</th>
<th>CTD</th>
<th>CDV</th>
<th>CLR&lt;sub&gt;0&lt;/sub&gt;</th>
<th>CLR&lt;sub&gt;0+1&lt;/sub&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>CBR</td>
<td>Required</td>
<td>Required</td>
<td>Required</td>
<td>Required</td>
<td>Required</td>
<td>Required</td>
<td>Required</td>
<td>Required</td>
</tr>
<tr>
<td>rt-VBR</td>
<td>Required</td>
<td>Required</td>
<td>Required</td>
<td>Required</td>
<td>Required</td>
<td>Required</td>
<td>Required</td>
<td>Required</td>
</tr>
<tr>
<td>nrt-VBR</td>
<td>Required</td>
<td>Required</td>
<td>Required</td>
<td>Required</td>
<td>Required</td>
<td>Required</td>
<td>Required</td>
<td>Required</td>
</tr>
<tr>
<td>ABR</td>
<td>Required</td>
<td>Required</td>
<td>Required</td>
<td>Required</td>
<td>——</td>
<td>——</td>
<td>——</td>
<td>——</td>
</tr>
<tr>
<td>UBR</td>
<td>Required</td>
<td>Required</td>
<td>Required</td>
<td>——</td>
<td>——</td>
<td>——</td>
<td>——</td>
<td>——</td>
</tr>
</tbody>
</table>

1. CLR<sub>0</sub> is the cell loss ratio for cells with the Cell Loss Priority bit set to 0.
2. CLR<sub>0+1</sub> is the cell loss ratio for all cells with the Cell Loss Priority bit set to either 0 or 1.
When two parallel links are available along the route, the controller chooses a link based on the configuration of the switch. The link selection options are:

- **AW**—Selects the link with the least AW in the egress direction. This is the default.
- **AvCR**—Selects the link with the largest AvCR in the egress direction.
- **maxCR**—Selects the link with the largest maxCR in the egress direction.
- **loadbalance**—Selects links randomly so that one link does not become overburdened while the other is idle.

## Additional Routing Features in MGX and SES Nodes

The following sections describe additional routing features you might want to consider when planning a PNNI network.

### Preferred Routing

*Preferred* routes allow the switch administrator to define a specific route between the source and destination nodes, and then specify this route as the preferred route when defining SPVCs. If the connection is configured as a *directed* route, no other route is allowed, even if the route fails. If the connection is not configured as a directed route, other routes are considered when the preferred route is not available. When other routes are required, the switch can use the pre-calculated routing tables or on-demand routing.

Preferred routing was introduced in Release 3.0.00 and is supported on the MGX 8830, MGX 8850 (PXM1E), MGX 8850 (PXM45), and MGX 8950 switches and the MGX 8880 Media Gateway. Release 3.0.20 and later support preferred routing on SES nodes.

**Note**

Preferred routes created with Release 3 software cannot be gracefully upgraded to Release 4 or later preferred routes.

**Note**

In all Release 3 software, the preferred routing feature specifies a route within a single peer group. Release 3 software does not support preferred routes that span multiple peer groups. Release 4 and later software does support preferred routes that span multiple peer groups.

The preferred route and directed route for an SPVC or SPVP is defined when the connection is created. Although you can change the preferred route configuration after a connection is created, you can eliminate reconfiguration by planning for preferred routes before creating connections.

### Priority Routing

Priority based routing allows you to specify a priority for each SPVC or SPVP connection. High priority connections are established before low priority connections. During failures, the high priority connections are also released and reestablished before low priority connections.

Priority routing was introduced in Release 3.0.00 and is supported on the MGX 8830, MGX 8850 (PXM1E), MGX 8850 (PXM45), and MGX 8950 switches and the MGX 8880 Media Gateway. Release 3.0.20 and later support priority routing on SES nodes.
The routing priority for an SPVC or SPVP can be defined with either the `addcon` or the `cnfcon` command. For SVCs and SVPs, the routing priority is assigned using the `cnfpnportsig` command. This routing priority also applies to the priority bumping feature. Although you can change the routing priority after a connection is created, you can eliminate reconfiguration by planning for priority routing before creating connections.

**Tip**
The priority routing feature allows administrators to influence the order in which connections are routed or rerouted when network events require connection rerouting. The priority routing feature does not change the criteria for selecting routes. It controls the sequence in which connections are routed or rerouted.

### Grooming

Connection grooming is the process of checking each connection to determine if a more efficient route is available. If a prospective new route is significantly better than the incumbent route, the connection is rerouted.

Grooming is also used to return a connection to its non-directed preferred route (if configured) after it has been rerouted due to failure along its preferred route. Connections will only return to their non-directed preferred routes when one of the following occurs:

- The connection is manually groomed.
- Automatic grooming is enabled and the grooming operation completes.
- The current connection route experiences a failure.

Grooming may be needed, for example, if a link fails along the most desirable route, and then returns to service. When the link fails, the connection is rerouted to another route, which may be a less desirable route. To return the connection to the more desirable route, you can use manual grooming or scheduled grooming. The advantage to scheduled grooming is that it can occur automatically at times when the network is not busy.

The grooming feature can be implemented at any time. Grooming is not configured at the same time as connections, so there is no penalty if you do not include grooming in the initial plan for a PNNI network.

### Soft Rerouting

The soft reroute feature is new in Release 5.0 and minimizes reroute times by establishing a new connection before releasing the rerouted connection. This feature requires no prior planning and can be implemented at any time. For more information, refer to the *Cisco MGX 8800/8900 Series Configuration Guide, Release 5.2*.

### Priority Bumping

Priority bumping is a new feature in Release 5.0. When enabled, priority bumping can be used to release lower priority connections to make room for an incoming, higher priority connection.
The priority bumping feature can be implemented at any time. However, the routing priority used for priority bumping is the same as used for priority routing. Because the routing priority is configured while creating and configuring connections, you might want to review the priority bumping feature details before configuring connections and interfaces. You can find more information on priority bumping in the *Cisco MGX 8800/8900 Series Configuration Guide, Release 5.2*.

**Blocking Pass-Through Connections**

As a switch administrator, you can configure MGX and SES nodes to support or deny connections that pass through the node. If you chose to deny transit or pass-through connections, the node will only accept calls that terminate on one of the node’s interfaces. Other nodes will not be able to establish routes through the blocked node to other nodes. This feature is called the Nodal Transit Restriction feature.

**Nodal Point-to-Multipoint Branch Restriction**

The point-to-multipoint (P2MP) feature enables select Cisco MGX switches to support PNNI network applications such as data and video broadcast and LAN emulation. P2MP branching is a feature that allows a switch to accept one incoming connection and produce multiple outgoing connections. This enables basic P2MP connectivity. For the nodes that support P2MP branching, branching can be enabled or disabled.

---

**Note**

Cisco SES equipped BPX switches can serve as the source or destination of a P2MP connection, but these switches cannot perform branching.

*Figure 4-1* shows the data flow in a P2MP connection and introduces the root, leaf, and party terms, which apply to the interfaces that support P2MP connections.

*Figure 4-1  P2MP Root, Leaf, and Party Components*

The simplest P2MP connection takes place through a single node. One endpoint serves as the root of a simple tree topology and is labeled the **root** end point. The data traffic is uni-directional. All data flows from the root endpoint to the destination endpoints.

A destination end point is called a **party**. A party is an ATM end station that connects to an edge switch and receives data from the connection root. Each party is identified by an ATM End Station Address (AESA) and an **end point reference**, which is a number that uniquely identifies the party. The endpoint reference is critical when multiple parties connect through the same AESA.
A leaf is a connection endpoint on an outgoing switch interface. At the edge of the network, the leaf represents the connection between the network and the party. Connections (SVCs or SPVCs) are established between the root and each leaf. At the interface that hosts the leaf, the received data is forwarded to each party using the AESA. As shown in Figure 4-1, each leaf can support multiple parties.

The current release of the P2MP feature on Cisco MGX switches operates on the service modules listed in Table 4-4.

### Table 4-4  MGX Service Module Support for P2MP Branching

<table>
<thead>
<tr>
<th>Service Module</th>
<th>Slot Multicasting Supported</th>
<th>Port Multicasting Supported</th>
</tr>
</thead>
<tbody>
<tr>
<td>AXSM/A&lt;sup&gt;1&lt;/sup&gt;</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>AXSM/B</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>AXSM-E</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>AXSM-XG</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>BPX/SES</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>PXM1E</td>
<td>Yes&lt;sup&gt;2&lt;/sup&gt;</td>
<td>Yes&lt;sup&gt;3&lt;/sup&gt;</td>
</tr>
</tbody>
</table>

<sup>1</sup> The AXSM/A term refers to the first release of the AXSM card, which is named AXSM. The AXSM/A term is often used to clarify that the reference is to the AXSM card and not the AXSM/B card.

<sup>2</sup> Slot multicasting is supported in Release 4.0.15 and later.

<sup>3</sup> Port multicasting is supported in Release 5.0 and later.

Slot multicasting or branching enables the PXM to branch an incoming P2MP connection to multiple service modules within the switch. Port multicasting allows a service module to branch an outgoing P2MP connection to multiple egress interfaces on that service module or to multiple channels on a port. When a service module does not support branching, the branching must be done at an upstream node that does support branching. To show how this works, this section introduces the PNNI farthest node branching feature.

The farthest-node branch option is a PNNI enhancement that allows PNNI to use network links more efficiently. Figure 4-2 demonstrates farthest-node branching.

![Figure 4-2 Farthest Node Branching](image)

In Figure 4-2, Switch 2 supports branching and Switch 3 does not. When PNNI sets up the P2MP connection to parties 1 through 4, it determines that the Switch 2 outgoing interface supports branching, so PNNI establishes 1 SVC between Switch 1 and 2.

Switch 3 in Figure 4-2 does not support branching, so Switch 1, which does support branching, establishes 4 SVCs to Switch 3. The destination for each SVC must be a leaf, so four leaf end points are established on Switch 3, one for each party. The four leaf end points can be on one interface, or they can be spread out on multiple interfaces. A leaf end point is an SVC endpoint.
Farthest-node branching is a PNNI feature that takes advantage of branching when it is available. Switch 1 could have originated a separate connection for each party on Switch 2, but this would have required 4 SVCs instead of 1, and all four would be carrying the same data. Farthest node branching improves network efficiency by reducing the number of SVCs required for P2MP connections and by reducing the bandwidth requirements for P2MP connections.

Previous software releases disabled branching with the branching restricted option. This option is now set by default to enable branching. If the network includes a node that does not support branching, the farthest branching node from the root creates an SVC for every downstream party.

**Note**

This release does not support the P2MP leaf-initiated join feature, and leaf endpoints cannot use a P2MP connection to communicate with other leaf endpoints.