CHAPTER 2

ATM Routing

The BPX SES PNNI Controller provides two types of routing protocols for an ATM SVC network:

- Dynamic routing with Private Network-to-Network Interface (PNNI) protocol, which enables provides Quality of Service (QoS) routes to signaling based on QoS requirements specified in the call request.
- Static routing, as configured with network management tools.

The ATM routing protocol is necessary to establish switched virtual circuits between ATM end users, (ATM CPE). The PNNI routing protocol is used to route SPVCs as well as SVCs. See Chapter 5, "ATM Soft Permanent Virtual Circuits" for more information about SPVCs.

This chapter includes descriptions of the following ATM routing functions:

- Private Network-to-Network Interface
- Interim Inter-switch Signaling Protocol (static routing)
- PNNI-IISP Interworking

Private Network-to-Network Interface

Private Network-to-Network Interface (PNNI) is a link-state routing protocol that provides dynamic ATM routing with QoS support as defined by the ATM Forum. PNNI supports aggregation for private ATM addresses and links between switches, and can scale the network and its performance by means of configuring PNNI peer groups and hierarchical levels.

A key feature of the PNNI hierarchy mechanism is its ability to automatically configure itself in networks in which the address structure reflects the topology.

PNNI routing functions include the following:

- Hello protocol
- Neighbor FSM
- PTSE database synchronization and management
- PTSE flooding
- Address summarization and advertisement
- Link and nodal aggregation
- Pre-computation of routing tables

- On demand routing
- Actions to request from commands issued using the CLI and/or SNMP

The SES PNNI node's implementation complies with the standard PNNI VI specification by the ATM Forum, which includes:

- Support for inside links in a multi-level PNNI hierarchy
- Quality of Service
- Multiple Routing Metrics
- Discovery of neighbors and link status
- Synchronization of topology databases
- Flooding of PTSEs (PNNI Topology State Elements)

PNNI software, which serves to create a routing database for use by the SVC software, runs on the PXM of the BPX SES PNNI Controller (Figure 1-7). The PNNI routing entity is relatively independent in the integrated system.

The PNNI routing entity maintains routing tables for use by the SVC processing function (that is, the Call Control Block) of the BPX SES PNNI Controller. The PNNI routing entity calculates the routing tables (as based on the most recent network resource availability information), tracks changes to local resources, and floods this information to other SES PNNI nodes in the network.

Route Agent

The Route Agent of the Call Control Block provides an interface to the PNNI routing tables. When the ATM CPE requests an SVC (for example, when a Q.2931 Call Setup message is received through the UNI signaling channel from ATM CPE 1), the SVC processing function uses the Route Agent to determine the route to the destination (ATM CPE 2). The destination ATM CPE is identified by its ATM address. The call is either cleared or forwarded:

- If no route can be found, the call is then cleared.
- If a route is found, the source SES PNNI node forwards the call to the next SES PNNI node along the route, and local resources are then programmed on trunks and lines as the call progresses.

  With each route request, the PNNI controller looks up the pre-calculated routing database for the given destination along with the requested service class and traffic metrics parameters. If a satisfactory route exists, the associated Designated Transit List (DTL) is returned as a response. The DTL is a list of nodes, and optionally link IDs, that completely specify a path across a single PNNI peer group.

PNNI Operation and Parameters

This section describes the configuration elements (Table 2-1) associated with PNNI network operations.

Table 2-1 PNNI Operation Elements

• PNNI Topology Database	• Border Node
• Hello Protocol	• Summary Addresses
• Database Synchronization	• Internal Reachable ATM Addresses
• ATM Service Categories	• Exterior Reachable Addresses
• ATM Metrics, Attributes, and Policies	• Logical Links
• ATM End System Address	• PNNI Hierarchy
• Address Prefix	• Hierarchical Level Indicator
• Load Balancing	• Crankback
• Node ID	• PNNI Statistics and Diagnostics
• Peer Group ID	• PNNI Redundancy

Default settings for these configuration elements are shown in Table 2-3 on page 2-18.

PNNI Topology Database

The PNNI entity maintains a topology database which is a collection of PNNI Topology State Elements (PTSEs). Each PTSE describes a piece of topology information. A PNNI node originates one or more PTSEs which describe its own environment, and it also learns PTSEs originated and advertised from all the other PNNI nodes in the network. A SES PNNI node generates a PTSE that describes its identity and capabilities.

PTSPs containing one or more PTSEs are used to disseminate information in the ATM network. PTSPs contain reachability, link and node status information necessary for PNNI to calculate QoS paths in an ATM network.

The PTSEs contain information such as:

- ATM address
- Node ID
- Peer Group ID
- Logical link to a neighboring node in the same Peer Group
- Internal Reachable Addresses
- Exterior Reachable Address

PTSEs are re-originated periodically by the originating nodes. They are also re-originated when triggered updates occur. Each PTSE also has a holddown timer which assures that the origination does not happen too frequently.The PNNI protocol uses a reliable flooding mechanism to exchange topology information with other PNNI nodes. Each PNNI node has a default address. Additional address prefixes can be obtained via ILMI address registration and PXM command line interface.

The PNNI flooding advertising mechanism provides for reliable distribution of PTSEs throughout a peer group. It ensures that each node in the peer group has a synchronized topology database. A node periodically issues an update to its PTSEs. Other nodes replace their copy with this PTSE if they recognize a change. Other nodes age and remove PTSEs belonging to a node if they have not received an update from it for a while. The default periodic flooding interval is 30 minutes. The minimum interval between updates to the same PTSE will be between 0.1 and 1 seconds. The minimum interval will prevent excessive flooding due to link attributes and metrics changing their values beyond their established thresholds very frequently.

Hello Protocol

Hello protocol is used in PNNI to detect and maintain adjacency between PNNI nodes. The Hello protocol is used to discover the identity of the adjacent neighbor node. The PNNI Hello protocol was modeled on the Open Shortest Path First (OSPF) protocol with appropriate extensions to support hierarchical organization of the topological database. Discovering the identity of the neighbor is done via an exchange of hello packets containing appropriate information. When the SES PNNI nodes discover they are members of the same peer group, they form an inside link.

Database Synchronization

When the Hello protocol has declared the link as functional, the adjacent SES PNNI nodes exchange a summary of their database contents. This mechanism is similar to the OSPF database synchronization procedures. The synchronization is governed by a a master and slave relationship of the SES PNNI nodes. Nodes exchange database summary packets which contain header information of all PNNI PTSEs in a node database. After such an exchange, differences in the topological databases are updated. When completed, both SES PNNI nodes have consistent topological databases.

A newly initialized SES PNNI node connecting to an existing PNNI network copies its database from its immediate neighbor.

ATM Service Categories

CBR	constant bit rate	Used by connections that request a static amount of bandwidth that is continuously available during the connection lifetime. The amount of bandwidth is characterized by Peak Cell Rate (PCR) value.
rtVBR	real-time variable bit rate	Intended for real-time applications that require tightly constrained delay and delay variation (voice/video applications). It is characterized in terms of a PCR, Sustainable Cell Rate (SCR), and Maximum Burst Size (MBS).
nrtVBR	non-real-time variable bit rate	Intended for non-real-time applications that have bursty traffic characteristics and are characterized in terms of a PCR, SCR, and MBS.
ABR	available bit rate	An ATM layer service category for which the limiting ATM layer transfer characteristics provided by the network may change subsequent to connection establishment. Flow control mechanism is specified.
UBR	unspecified bit rate	Intended for non-real-time applications, such as those not requiring tightly constrained delay and delay variation.

ATM Metrics, Attributes, and Policies

As a topology state routing protocol, PNNI advertises detailed information about the status of the links and nodes. The status of topological entities (links and nodes) is described via metrics and attributes that can be used to tune network behavior.

Metrics are added along a path. Each link contributes a value. The simplest example of a metric is the administrative weight (AW). The AW of a path is the sum of the weights of links and nodes along the path.

Attributes are treated by PNNI in a different way. If an attribute value for a parameter violates the QoS constraint, the PNNI excludes that topological entity from consideration while making a path selection.

During or/and after the network design, a number of parameters can be provisioned on a per-node or per-link basis to enhance overall network performance. The tunable parameters described in this section are as follows:

- Metrics
  - Administrative Weight (AW)
  - Maximum Cell Transfer Delay (MaxCTD)
  - Peak-to-Peak Cell Delay Variation (CDV)
- Attributes
  - Available Cell Rate (AvCR)
  - Maximum Cell Rate (maxCR)
  - Cell Loss Ratio for CLP= 0 traffic (CLR0)
  - Cell Loss Ratio for CLP= 0+1 traffic (CLR0+1)
- Policies
  - Bandwidth Overbooking
  - Link Selection

Administrative Weight (AW)

The AW can be configured at a per-interface and per-service class basis at lowest level nodes, and it is associated with each link and each complex node at all hierarchical levels. The AW is one of the key matrixes used in route selection procedure for optimization. The assignment of administrative weights to links and nodes influences the manner by which PNNI selects paths in the SES PNNI node network.

The default AW on all interfaces is the standard 5040. Tuning the AW on interfaces impacts the distribution of SPVC/SVC over the network.

Administrative weight can also be used to exclude certain links from routing, such as a backup link that needs to be used only when the primary link is full. The administrative weight for a path is simply the sum of the individual weights of the links on the path.

Available Cell rate (AvCR)

AvCR is a measure of effective available capacity for CBR, rtVBR, and nrtVBR service categories. For ABR service, AvCR is a measure of capacity available for minimum cell rate (MCR) reservation.

Maximum Cell Rate (maxCR)

MaxCR is the maximum capacity usable by connections belonging to the specified service category. MaxCR=0 is a distinguished value used to indicate inability to accept new connections in UBR and ABR service categories.

Bandwidth Overbooking

The per-service class based Available Cell Rate (AvCR) on each link is advertised by PNNI routing protocol after the overbooking factor applied. Tuning bandwidth overbooking impacts the results of the G-CAC during the routing path selection procedure. A more aggressive overbooking approach allows more SVC/SPVC to share the same set of network resources more efficiently, with trade-off of possible more signaling crankbacks in the network. A more conservative approach tends to have the opposite result.

The bandwidth overbooking factor may be configured at a per-interface and per-service class basis at lowest level nodes. The setting is established by configuring AvCR and reflects the amount of equivalent bandwidth that is available on the link or node. AvCR is the most dynamic attribute in PNNI. AvCR depends on the calls traversing the links and is viewed as the residual capacity left for use by additional calls. PNNI needs the knowledge of AvCR to decide whether a given link or node is suitable to carry a given call, as measured in calls per second.

Not all AvCR changes are advertised in the network by PNNI. Only significant changes, as defined by the ATM Forum Specification, are advertised in the network.

Maximum Cell Transfer Delay (MaxCTD)

The CTD, which is the quantile of the elapsed time for transmission of cells across a link or node, can be configured on a per-interface and per-service class basis at lowest level nodes, and is associated with each link and each complex node at all hierarchical levels. The CTD is one of the key matrixes used in route selection procedure for optimization as well as meeting QoS based signaling requirements. Tuning the CTD on interfaces impacts distribution of SPVC/SVC over the network.

CTD includes processing and queueing delays plus propagation delay, as measured in calls per second.

Peak-to-Peak Cell Delay Variation (CDV)

Currently, the CDV, which is the quantile of the cell transfer delay minus the fixed delay experienced by all cells crossing the link or node, is statically defined at a per-interface and per- service class basis at lowest level nodes, and it is associated with each link and each complex node at all hierarchical levels. The CDV is one of the key matrixes used in route selection procedure for optimization as well as meeting QoS based signalling requirements. Tuning the CDV on interfaces will impact on the distributions of SPVC/SVC over the network. In the future, the CDV may be measured dynamically. CDV is measured in micro-seconds.

Link Selection

When multiple links connecting two PNNI nodes or peer groups, preferences can be applied as which link is used for any given SVC/SPVC based on configured policies, which include the AW, AvCR, maximum cell rate (maxCR) or randomizing. This mechanism may be used in the network design process to layout the traffic behavior on parallel links between nodes or peer groups.

Maximum cell rate is the maximum capacity usable by connections belonging to the specific service category, measured in calls per second.

Cell Loss Ratio for CLP= 0 traffic (CLR0)

CLR0 the maximum cell loss ratio for CLP_0 traffic over a link or node.

Cell Loss Ratio for CLP= 0+1 traffic (CLR0+1)

CLR0+1 is the maximum cell loss ratio for CLP_{0+1} traffic over a link or node.

In the PNNI protocol, not every change of parameter value is substantial enough to generate an advertisement. The network would be overwhelmed with PNNI advertisement packets if frequently changing parameters were to generate advertisements every time any change in their value occurred. Changes in CDV, MaxCDT or AvCR are measured in terms of a proportional difference from the last value advertised. A proportional multiplier threshold expressed as a percentage provides flexible control over the definition of significant change.

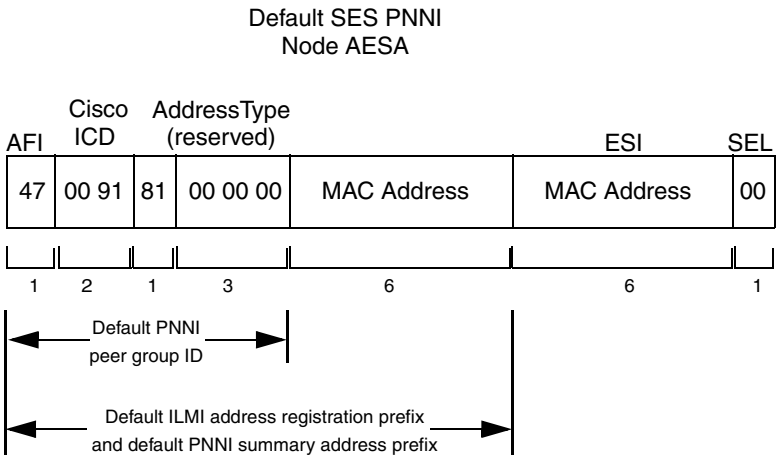
For some parameters, such as administrative weight, any change in value is considered significant.

ATM End System Address

UNI ports on a SES PNNI node will be identified by an End System Address (that is, ATM address) within the network. The PNNI entity within a SES PNNI node also is associated with an End System Address that is unique within a network.

The default PNNI Node ATM address is based on the SES PNNI node's MAC address, as shown in Figure 2-1.

Figure 2-1 Default PNNI Node ATM Address

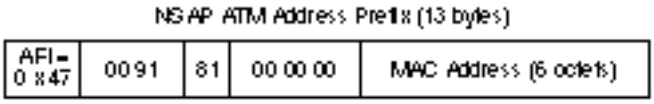


The 13-octet prefix of the PNNI Node ATM Address is used for the default UNI port prefix. All SES PNNI nodes use the first 7 bytes of this prefix (0x47 0091 8100 0000), the default PNNI peer group ID. The default PNNI Node ATM address is also used to form the PNNI Node ID as described in the section on Node ID.

Address Prefix

One or more address prefixes may be configured for each SES PNNI node on a per-port basis. The prefix is unique for each SES PNNI node in the network. ATM end point addresses (UNI ports) or host addresses attached to the SES PNNI node will usually bear this prefix. The default ATM UNI port address prefix, which is 104 bits long and configurable, is shown in Figure 2-2.

Figure 2-2 Default UNI Port Network Prefix



This prefix includes the 7 byte PNNI peer group ID (0x47 0091 8100 0000), plus a unique 6 byte MAC address. This is the prefix used for ATM ILMI address registration with ATM UNI ports.

Hierarchical Level Indicator

Each node in the PNNI network operates at the PNNI hierarchy configured in the level indicator, which range from 0 to 104. The hierarchical level indicator specifies the number of significant bits used for the peer group ID.

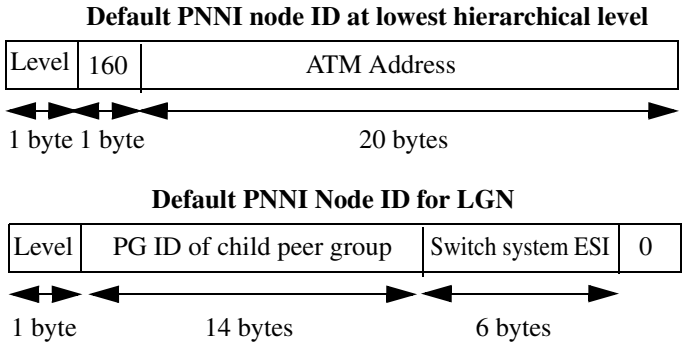
At the lowest hierarchical level, the standard default is 96. The Cisco ATM switch defaults to 56. Larger numbers equate to lower positioning in the hierarchy, therefore the higher the number configured as the level indicator, the lower in the hierarchy is the associated switch.

The hierarchical level indicator is configurable by using the network management interface.

Node ID

A SES PNNI node is identified by a unique identifier called PNNI node ID. A node ID consists of 22 bytes, including the SES PNNI node's 20-byte default ATM address described in the section, ATM End System Address, and with two prepended bytes (Figure 2-3).

Figure 2-3 22-byte Node ID



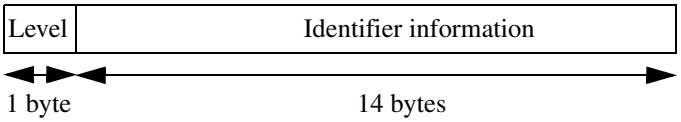
The first prepended byte corresponds to the PNNI level indicator byte (38 hex, or 56). The second prepended byte is decimal 160 (A0 hex). The reminder of the node ID address may be configured independently if the default node ID is not used.

The node ID is configurable by using the network management interface.

Peer Group ID

Each peer group has a 14-byte peer group ID.

Figure 2-4 14-Byte Peer Group ID



Nodes with the same peer group ID are within the same peer group. the Peer Group ID at a child peer group is the default summary address of the logical group node at the parent peer group.

The configuration of peer group IDs determines the structure of the PNNI network hierarchy.

Peer group ID is configuration by using the network management interface.

Border Node

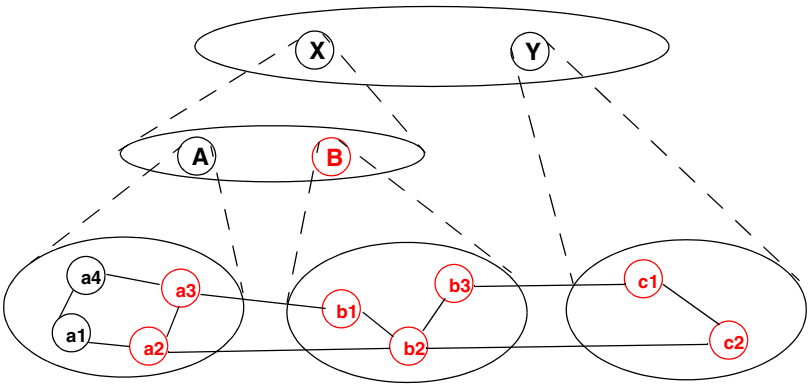
Border nodes are PNNI nodes that originate and advertise PNNI uplink PTSEs at the highest and lowest hierarchical levels of the PNNI network.

At the lowest hierarchical level, border nodes also execute Hello protocol exchanging information with nodes in other peer groups. For entry border nodes or peer groups, border nodes must construct new DTLs for routing paths traversing the new peer group for SVC/SPVC call requests. As a result PNNI

border nodes are busier than others. It also requires border nodes to be as stable as possible, since malfunctions or outages of a border node may cause problems for SVC/SPVC that traverse the two peer groups.

Figure 2-5 shows a high-level border node (B), and border nodes at the lowest level (a3, a2, b1, b2, b3, c1, and c2).

Figure 2-5 Border Nodes



Considerations for selecting border nodes include:

- Switches that have more memory and CPU processing power.
- Switches that have more stable software release.
- Switches that have less traffic load, especially in regard to VC/VP connections and bandwidth usage.
- Switches that are not PGL nodes.
- Switches that have enough bandwidth on outside links that connect to other peer groups.

Summary Addresses

PNNI allows summarization of multiple ATM addresses into a single summary address prefix. Address summarization and the hierarchical organization of the topology enables PNNI to scale to very large networks.

Reachability information is used as the first step in routing a PNNI signaling request for a virtual connection. The Call Setup packet will be directed to a SES PNNI node advertising a prefix which matches the leading portion of the destination address. The longest matching reachable prefix is always used.

Host IDs are associated with the UNI ports of the SES PNNI node. A host ID usually consists of the prefix value that has been established for the SES PNNI node. In this case the remainder of the host ID will be determined through ILMI address registration procedure. Host IDs determined through the ILMI address registration procedure will be known to the PNNI element.

A SES PNNI node may consist of many ATM end point addresses. It may be desirable to reduce the amount of addressing information which needs to be distributed in a network. A summary address may be used for the purpose of optimizing the amount of messages. Summary addresses are configurable at each SES PNNI node. A summary address is a collective representation of multiple ATM end point

addresses sharing a prefix. Each SES PNNI node has a default summary address which is equivalent to its nodal address prefix. Individual addresses will be summarized. Additional summary addresses may be configured through network management action.

Internal Reachable ATM Addresses

This information group describes internal reachable ATM destinations. Internal means known to PNNI to be local. For a node representing a single switch, an internally reachable address represents a summary of end systems (ATM CPE) attached to the SES PNNI node, for example discovered ILMI address registration. Internal reachable ATM addresses can also summarize information provided by members of the same peer group.

Exterior Reachable Addresses

Exterior reachable addresses may be configured at the PNNI element. A NSAP address format may be used for an exterior reachable entity. Addresses belonging to other independent PNNI domains are also established using this mechanism. Manual configuration commands are available to establish external reachable addresses. PNNI protocol will not run over logical links that reach outside the PNNI domain.

Logical Links

A trunk on the SES PNNI node appears to the PNNI process as a logical link. In a single peer group, a physical link connecting two SES PNNI nodes corresponds to a logical link. 255 logical links may be defined per SES PNNI node. The PNNI process utilizes an ATM/PVC socket to communicate over a logical link. VPI=0 and VCI=18 will be reserved for use by PNNI on each logical link.

The PNNI Hello protocol is used to track the operational status of a logical link. The operational status reflects the reachability of the neighbor attached to a logical link. The logical link's operational status may also change due to a failure at the physical layer. A physical layer failure may be trunk failure on a BPX 8620 (a BPX switch). **Upcd** and **Dncd** commands issued on BPX switches will result in change of operational status as well. A failed or downed trunk appears as a deleted trunk for PNNI.

Every logical link has a dedicated set of resources for use by SVCs. The Maximum Cell Rate bandwidth (MaxCR) will be applicable to all service classes. MaxCR will be configurable per logical link and is the partition size for SVCs for a logical link. Available Cell Rate (AvCR) changes due to the addition or a deletion of an SVC. A significant change in AvCR will trigger the PNNI element to flood the network.

There are a limited number of connection channels per SES PNNI node port. A trunk may have sufficient capacity to support a given QoS with respective to the other metrics but it may still be limited from carrying more connections beyond its channel capacity limit. The free channel capacity must be tracked as an additional resource. If number of local resources like LCNs become zero, then the BPX SES PNNI Controller will inform the PNNI process that the AvCR for this link is 0.

The BPX switch will detect link-state changes at the physical layer and will pass them along to the PNNI protocol entity.

Peer Group Leader Elections

Peer group leader election is a dynamic process that occurs within each peer group at all hierarchical levels. Only one peer group leader may exist in a single peer group. All nodes in a peer group execute the election algorithm except those nodes that are configured for non-transit peer group election. The node with the highest leadership priority, using node ID as the tie breaker, is elected as the preferred peer group leader for the peer group. The range for the election algorithm is 0-205.

The election is the result of the comparison of all nodes except for those nodes that are not reachable from the calculating node, and those nodes with leadership priority configured for 0. A preferred node becomes a peer group node is the vote count is at least 2/3 majority. A peer group node that demonstrates network stability gets a “bonus” priority increment of 50.

To increase the network stability, the re-election of PGL in any peer group must be avoided to the extent, since any re-election might result to a different PGL node, which will in turn cause PNNI logical links going down unnecessarily. The network stability in this sense can be achieved by the provisioning of proper PNNI leadership priority. The following table shows the leadership priority values based on varies of PGL preferences on PNNI nodes.

No.	Preferences on Nodes to be PGL	Leadership Priority Values
1	Nodes unwilling or unable to become PGL at any time	0
2	No preference, allowing PGL randomly elected	1 <= priority <= 49
3	With preference in a small groups of nodes	100<= priority<=149
4	Preference is on a single node	200<= priority<=205

The peer group leader is responsible for the following:

- Activates or deactivates a parent logical group node.
- Presents default and other summary, individual internal, exterior reachable addresses to the parent logical group node.
- Presents uplinks to the parent logical group node.
- Inherits PTSEs from the parent logical group node and distributes them within its own peer group.
- Calculate path information between border nodes and presents them to the parent logical group node, using complex node representation.

Logical Group Node

The logical group node (LGN) is activated on the switching system if the node at the next lower level becomes a peer group leader in the child peer group, and the logical group node is administratively configured.

Responsibilities of the logical group node are as follows:

- Aggregation of uplinks from the child peer group into horizontal links.
- Setup and maintenance of SVCC-based RCC with each of the peer logical group nodes.
- Advertisement and origination of nodal state parameter PTSEs (for logical group node with complex node representation).
- Downstream feeder of PTSEs from its own and its ancestors' peer groups to its child peer group.

PNNI Hierarchy

The level of hierarchy of a PNNI-based network can be from one to ten inclusively. When the level is one, all nodes are in the same peer group, and it is referred to as a non-hierarchical or flat network.

In a flat network, the network topology database in all nodes is synchronized, so that the routing decisions are most likely very accurate, optimized and effective. The problem for a flat network is poor scalability. A hierarchical PNNI network with multi-level and multi- peer group can scale to very large size, sufficient for nearly all world-wide public or private networks for many years to come. In general, the larger the total number of levels in hierarchy, the larger the network can be. However, the secret of the scaling ability in a hierarchical network is the topology abstraction; for example, topology information in each peer group is aggregated before presented to the next higher level peer group, and nodes in a peer group does not know the topology details of other peer groups. As a result, the routing paths may not be as accurate nor as optimized as in a flat network in general. When the total number of hierarchical levels increases, the degree of abstraction goes up, and the routing efficiency tends to suffer more.

The trade-off between routing efficiency and scalability must be recognized and managed properly. In most of the large-scale public and private networks, hierarchical network infrastructure is inevitable, but on the other hand, a hierarchy of up to 3-4 levels may be sufficient for a very large network today or in the future.

Even in a hierarchical PNNI network, varies of network structures can be evaluated and chosen, along with different means that may be used to tune the network behavior, and the purpose is to improve the overall network performance in a customized manner.

PNNI Topology Database

Memory consumptions in PNNI are dominated by the size of the topology database. In a large scale public ATM network consists of switches with high port density, the most of the advertisements in the database are PNNI links. Further, in a SPVC-centric ATM network, the advertisements for reachable addresses will be much less than that of SVC-centric ATM network, such as a large ATM/LANE network.

Insufficient system resources will cause network instabilities. The consequences of delayed processing, memory shortage, etc. include the following for a PNNI node:

- link flapping
- loss of neighbors
- topology database out-of-sync
- high crankback rate

- stale routing tables
- very frequent re-retransmission (PTSPs and PTSE acknowledgment packets)

Routing

The PTSEs on the SES PNNI node are used to calculate the PNNI routing tables. Route calculation is a CPU-intensive, background activity that is triggered by significant changes in network topology, Significant changes include the appearance and/or disappearance of nodes, operational variations in link availability, and changes to link state parameters on links and complex nodes.

Route selection is a two-part process that consists of searching for the called party address and the determining the routing path to the destination node.

1. Searching for the called party address in the system address table and PNNI network address table

   If the called party address is locally attached, data is then routed to the local port. If not locally attached, and advertising nodes are resident, a different path to the node is selected. If not locally attached and no advertising nodes are resident, the call is dropped.

2. Determining routing path to the destination node.

   Pre-computed tables are searched first. Routing paths are based on on-demand criteria if the search in the pre-computed table fails.

   Routing based on pre-computed tables. The number of pre-computed routing tables is dependent on the combined number of class-of-service PTSEs that a node generates and receives from the network. The minimum number of tables generated is three (the default), when each of the these traffic metrics is identical for all applicable class of service types.These tables are:

   - a table optimized on AW for all QoS
   - a table optimized on CTD for CBR, rtVBR, and nrtVBR
   - a table optimized on CDV for CBR and rtVBR

   After the calculating node generates or receives class-of-service based traffic metrics, the tables split accordingly, up to a maximum of 10. These tables are:

   - a table optimized on AW for CBR
   - a table optimized on AW for rtVBR
   - a table optimized on AW for nrtVBR
   - a table optimized on AW for UBR
   - a table optimized on AW for ABR
   - a table optimized on CTD for CBR
   - a table optimized on CTD for rtVBR
   - a table optimized on CTD for nrtVBR
   - a table optimized on CDV for CBR
   - a table optimized on CDV for rtVBR

   Each routing table is maintained as a shortest path tree (SPT) and each SPT is constructed based on the optimization of a particular traffic metric. An SPT maintains the shortest path tree to each reachable destination node located in the same PNNI routing domain as based on the traffic metric.

On Demand Routing. The routing path is selected directly from the PNNI topology database on the basis of the currently configured mode—either first fit or best fit. If first fit is configured, the route consists of finding a single path that satisfies the call request as quickly as possible. If best fit is configured, the route consists of finding a single path that satisfies the call request at the least cost. Other characteristics of on-demand routing are:

- link verification and checks for path constraints
- avoidance of blocked nodes and links in the event of crankbacks
- DTL limit checking

An originating SES PNNI node looks up the routing tables to determine the DTL for an SVC. For a local/DACS connection, the call to the Route Agent returns the egress port. Nodes in route do not call the Route Agent. The port ID in DTL is used to forward the call, and the DTL pointer is advanced. If the port ID is zero, the Route Agent is called to get the egress port to the next node in the DTL. This is the responsibility of Call Control Block function. At the far-end SES PNNI node of the connection, the Route Agent provides the egress port.

PNNI Routing and Other Resources

When PNNI detects the system is out of the resource to handle the call of a specific service category in an interface, it will advertise maxCR=AvCR=0 for this interface in that service category. No more calls will be routed to this interface for that service category. The resources include the Logical Channel Number (LCN), VPI/VCI range.

When the resources resumed back (more than a threshold), PNNI will advertise the real maxCR and AvCR again. The threshold is hardcode to 8. It is recommended that the total LCN numbers and VPI/VCI ranges shall be more than 10.

Load Balancing

The SES PNNI node provides configuration for load balancing—in one of three levels—to ensure equal distribution of the load across all trunks.

- Load balancing on equal cost paths, as configured to within an exact match, ± 1.5%, ±3.1%, ±6.2%, ±12.5%, and ±25%.
- Load balancing on parallel links (to the same peer neighbor)
- Load balancing with redundant addresses

Alternate Paths - Network Wide Load Balancing

When the Route Agent discovers that there are more than one equal paths qualified for a given call, a path is selected as follows:

1. Randomize— for all service categories, select a path randomly among equal paths.
2. Bandwidth-maximize:
   a. For UBR, select a path randomly among equal paths.
   b. For CBR/VBR/ABR, select a path that has max(AvPB0,AvPB1,... AvPBj,... AvPBm), where AvPBj(Available Path Bandwidth) = min(AvCR0j,AvCRj,... AvCRij,... AvCRnj):

      i - link index
      j - path index
      n - total number of links per path
      m - total number of paths

Parallel Links - Local Load Balancing

The Route Agent returns a port list for parallel links to Call Control. The port list is sorted depending on service category and/or link-selection option—either CBR and VBR, or UBR and ABR.

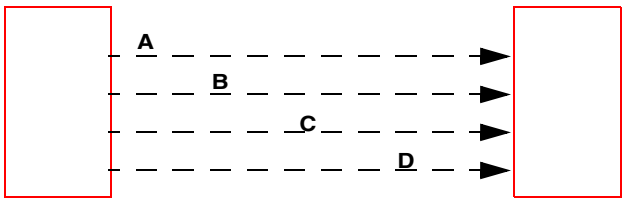
1. CBR and VBR
   a. admin-weight-minimize—sorted on AW
   b. blocking-minimize (default)—sorted on AvCR
   c. transmit-speed-maximize—sorted on MaxCR
   d. load-balance—randomize

   Preference is also in the order of the above in case the configuration options on links disagree.

   Link selection requires two steps (see example in Table 2-2):

   - Where configuration options are different on the four links, the option with the highest preference is used., as in Admin-weight-minimize (AW on link A)
   - Admin-weight-minimize—Link with minimum AW to be selected, as in link A (AW=2000).

Table 2-2 Example: Parallel Link Selection in Load Balancing



Links	Link Select	AW	AvCR	MaxCR
Link A	AW	2000	6000	300000
Link B	AvCR	5040	100000	200000
Link C	Balance	10000	500	200000
Link D	Balance	3500	100000	300000

2. UBR and ABR

   sorted on (UBR AvCR + ABR AvCR)/(UBR conns + ABR conns)

Redundant Addresses

If identical addresses are provisioned or learned from the network, and if redundant addresses are in different address tables, precedence is taken as follows:

1. Addresses in the System Address Table
2. Addresses in the Network Reachable Address Table

If redundant addresses are in the same table, the target is selected as follow:

1. In the System Address Table, randomly select a routing target.
2. In the Network Reachable Address Table, addresses are prioritized by:
   a. Resource Availability Information (IISP trunk) if available.
   b. Cost.
   c. In the event of identical priorities, select a target as follows:

      —For CBR/VBR/ABR calls, select the IISP links with the highest AvCR.

      —For UBR calls, select the IISP links randomly.

Crankback

A crankback is a call that has been diverted back to the originating DTL due to inaccurate resource and connectivity information from the DTL originator.

Crankback happens when there are changes in the availability of resources, topology, or hierarchical aggregation. Depending on the nature of the crankback, one of the following determines the best alternate route:

- pre-computed routing tables
- on-demand route
- the next longest matching target in the reachable address table

PNNI Statistics and Diagnostics

The PNNI entity maintains statistical counts for the normal and error activities at the packet layer, at the PTSE layer for advertisements, for the PTSE age out discards, for the PTSE flood procedures, for route table calculations and for route table look-up activities.

PNNI Redundancy

The PNNI process initializes its local node information from disk storage. It derives the remote node information through flooding. Therefore, when a standby BPX SES PNNI Controller takes over, the PNNI information takes some time to be gathered from the network. During this period, all calls to the Route Agent will return ‘No Route Available’. PNNI is transparent to the other redundancy aspects of the SES PNNI node such as a BPX switch trunk card switchover, a BCC switchover and a UNI card switchover.

As a dynamic routing protocol with built-in mechanism of robustness, PNNI should be able to re-build its PTSE database, pre-computed routing tables, and other internal data structures in a reasonably short period of time, after the BPX SES PNNI Controller switches over from the active to the standby.

The actual re-build time depends on a number of factors that include:

1. Size of the PTSE database in the peer group where the node belongs.
2. Traffic contract on the RCC on which the node performs database synchronization with the neighboring peer.
3. Type of the node (such as a PGL node, border node, with hierarchy, or similar).

During the re-build time, the PTSE database is re-synchronized in an incremental manner, which causes the pre-computed routing tables are re-generated repeatedly (subject to the holddown timer) also in the same manner. The routes in the routing tables as well as those found directly in the on-demand search procedure become increasingly available in the re-build time interval. Therefore, the success of the call routing follows the same curve statistically, and the calls, if any, are not necessarily all blocked.

All PNNI processes on the active platform are also spawned on the standby platform, where, however, all PNNI node(s) are put into "administratively down" status. All PNNI provisioning data on the active platform are also stored in the associated data spaces that belonged to PNNI as well as on the disk. This arrangement speeds up the recovery procedure when the switch-over occurs.

Default PNNI Configuration

The default configuration (Table 2-3) is used if no associated provision data is received through the Control Point.

Table 2-3 Default PNNI Configuration Values

Configuration	Default Values	Range
AESA of lowest level node	Cisco's standard prefix[1]	N/A
level	56	1-104
PG id	derived automatically from AESA	N/A
node id	derived automatically from AESA	N/A
complex node representation	off	off, on
restricted transit	off	off, on
restricted branching	off	off, on
PTSE holddown timer value	1 second	0.1-10 seconds
Hello holddown timer value	1 second	0.1-10 seconds
Hello interval	15 second	1-300 seconds
Hello inactivity factor	5	1-50
Horizontal logical link inactive timer value	120 second	30-600 seconds
PTSE refresh interval	1800 seconds	30-1800 seconds
PTSE lifetime factor	200%	101-1000%
Retransmit interval	5 seconds	5-60 seconds

Table 2-3 Default PNNI Configuration Values (continued)

Configuration	Default Values	Range
PTSE delayed interval	1 second	1-10 seconds
AvCR proportional multiplier	50%	1-99%
AvCR minimum threshold	3%	1-99%
maxCTD proportional multiplier	50%	1-99%
CDV proportional multiplier	25%	1-99%
PGL election priority	0	0-205
PGL init time	15 seconds	1-120 seconds
PGL override delay	30 seconds	1-120 seconds
PGL re-elect time	15 seconds	1-120 seconds
SVCC-based RCC init time	4 seconds	1-10 seconds
SVCC-based RCC retry time	30 seconds	10-60 seconds
SVCC-based RCC calling integrity time	35 seconds	5-300 seconds
SVCC-based RCC called integrity time	50 seconds	10-300 seconds
default summary address	derived automatically from the node AESA	N/A
configured summary address	none	N/A
PNNI links	none	N/A
locally reachable addresses	none	N/A
parent node configuration	none	N/A
routing tables generation holddown time	1 second	1-600 seconds
border node bypass holddown time	2 second	2-600 seconds
routing tables equal-cost path *epsilon*	0	0-20
network-wide load-balance policy	randomized	CBR/VBR/ABR links with bandwidth
link selection on parallel links	link with the largest AvCR	based on AvCR, max CR, AW, or randomized
scope map	96, 96, 96, 80, 80, 72, 72, 64, 64, 64, 48, 48, 32, 32, 0	0-104, inclusively

PNNI SNMP Support

The PNNI parameters can be configured through the SES command line interface (CLI), CiscoView, or another SNMP manager. The PNNI SNMP Management Information Base is described in Appendix E, "SNMP Management Information Base".

Interim Inter-switch Signaling Protocol

The SES PNNI node also supports the Interim Inter-Switch Signaling Protocol (IISP) defined by ATM Forum Specification 1.0. IISP is a static routing protocol built upon ATM Forum User to Network (UNI) Specification 3.1, with optional support for UNI 3.0. IISP assumes no exchange of routing information between switching systems, and allows you to connect to networks that do not support PNNI. It uses a fixed routing algorithm with static routes. Routing is done on a hop-by-hop basis by making a best match of the destination address in the call Setup with address entries in the next hop routing table at a given switching system (such as in the case of a SES PNNI node or foreign switch).

Note IISP was specified by the ATM Forum to provide a base-level capability between ATM networks in the interim during finalization of PNNI Phase 1 specifications.

IISP Links

IISP links are Network to Network Interface (NNI) trunks between SES PNNI nodes. Each link is provisioned for the following:

- protocol version
  - IISP 3.0 to comply with ATM Forum 3.0 UNI signaling standards.
  - IISP 3.1 to comply with ATM Forum 3.1 UNI signaling standards.
- NNI Role
  - One side performs the user-side (NNI/user) of the UNI signaling standard.
  - The other side performs the network-side (NNI/network) of the UNI signaling standard.
- Weight

Note IISP links should only be used at border nodes in a PNNI network.

IISP Signaling

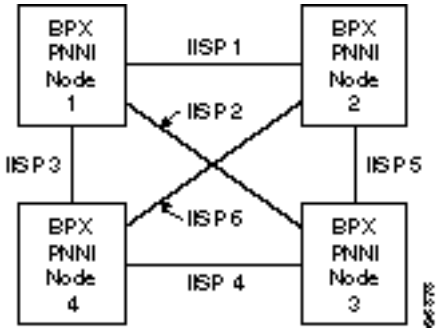
IISP signaling is based on the UNI 3.1 signaling specification. At a specific IISP link, one of the switching systems (a SES PNNI node) plays the role of the user side, and the other switching system assumes the network side role. Only the network side is allowed to assign VPI/VCI values to a call to avoid call collisions on the IISP link. These roles are assigned manually through the CWM or CLI.

Note The SES PNNI node does not support Transit Network Selection as defined in the ATM Forum IISP specification.

IISP Network

To be fully reliable, an IISP network composed of SES PNNI nodes would need to have full-mesh of IISP links. Obviously, this becomes impractical for a network of more than a few nodes. Figure 2-6 illustrates a simple IISP network with four SES PNNI nodes. Each SES PNNI node has three IISP links, one to each of the other three other SES PNNI nodes.

Figure 2-6 IISP Full-Mesh Network



PNNI-IISP Interworking

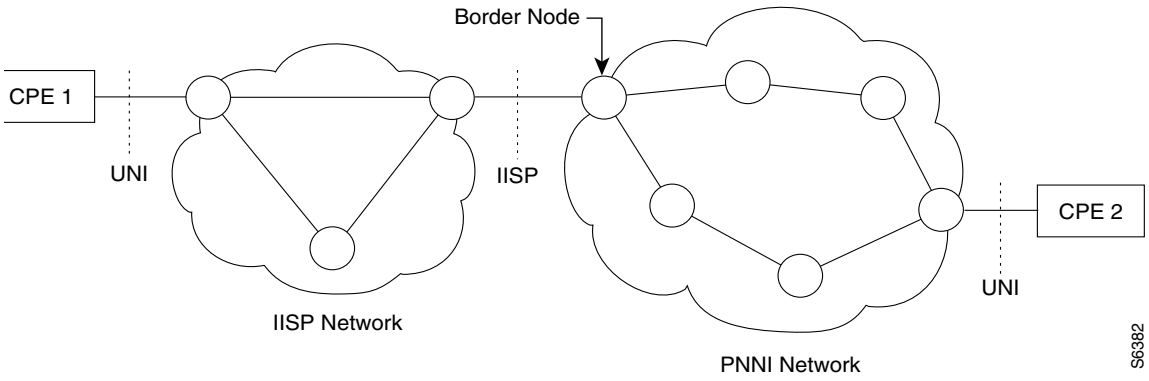
A SES PNNI node may be configured to have both PNNI and IISP links. A SES PNNI node that has both PNNI and IISP links is effectively a PNNI border node for the PNNI group.

Note IISP links should only be used at border nodes within a PNNI network.

The IISP static routes at the border node are PNNI external reachable addresses. When IISP static routes are configured on an IISP port, the PNNI Controller's system manager sends them as external reachable addresses to the PNNI protocol entity (Figure 2-7). The PNNI protocol entity advertises them to other PNNI nodes if they do not match any summary addresses. These IISP static route addresses are included in the routing tables maintained by the PNNI Route Agent.

The example in Figure 2-7 shows a border node that is configured with a set of static routes that represent reachable end-points in the IISP network. These routes are broadcast to all nodes in the PNNI network.

Figure 2-7 PNNI to IISP Interworking

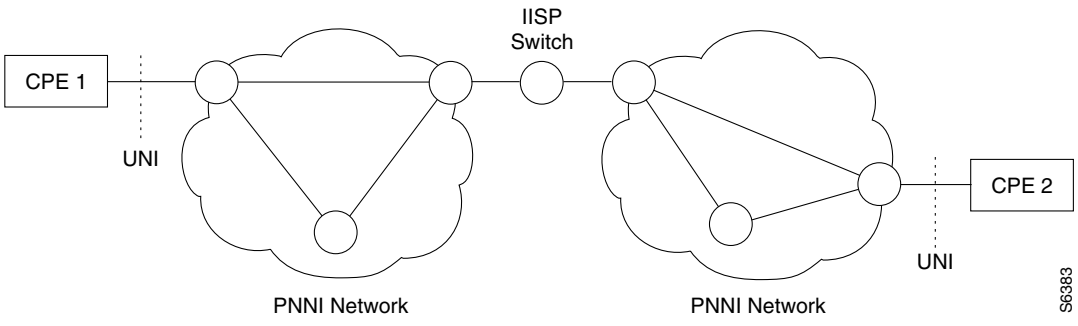


A call that originates in the PNNI network from CPE 2, for example, can place a call to one of these IISP addresses. At the originating PNNI node, a DTL is created representing PNNI nodes from the originating PNNI node to the border node. The Setup message traverses the PNNI network using the DTL list. When the Setup message reaches the border node, the Route Agent supplies the egress port, which is an IISP port. The IISP Setup message without the DTL list now traverses the IISP network using static routes configured at each node until it reaches the destination UNI, such as CPE 1.

A call originating in the IISP network, from CPE 1, can place a call to one of its statically configured routes which may be a destination address in the PNNI network. The IISP Setup message traverses the IISP network using static routes configured at each node until it reached the PNNI border node. At the entry border node a DTL list is created representing PNNI nodes from the entry border node to the terminating node. The Setup message traverses the PNNI network using the DTL list. When the Setup message reaches the terminating node, the Route Agent supplies the egress port which is the UNI port to the destination, CPE 2.

The example in Figure 2-8 illustrates a typical case whereby two PNNI networks are connected through an IISP switch. In this case, the border nodes (namely, the nodes that are connected to the IISP switch) in both networks must have IISP trunks configured to the IISP switch.

Figure 2-8 Typical Use of IISP



IISP Enhancements

The BPX SES PNNI Controller enhances IISP beyond the version 1.0 standard with the features described in the following topics:

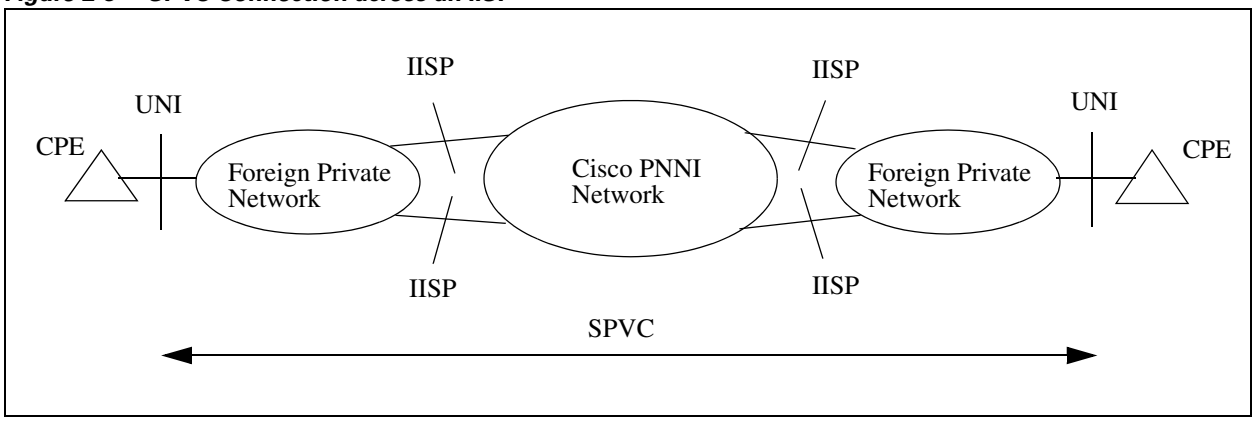
- SPVC (Transport SPVC IE)
- Signaling of VBR-rt
- Signaling of Transparent VP Services
- Signaling of Soft-PVPC
- Transport of Generic Identifier Transport IE
- SPVC Termination
- Static Routing
- IISP/UNI Partitioning
- IISP CAC and Overbooking
- IISP CAC and Overbooking
- Policing Enable/Disable on IISP
- Frame Discard
- IISP Crankback

Note The default configuration for IISP Crankback is "ON". Turn IISP Crankback "OFF" for an interoperable version of ISP.

SPVC (Transport SPVC IE)

SPVC is supported between the Cisco PNNI network and the foreign private network when these two domains are connected by IISP trunks (Figure 2-9). The IISP support is implemented accordingly to IISP 1.0 based on UNI 3.1; Available Bit Rate (ABR) connection will not be supported for SPVC across IISP trunk.

Figure 2-9 SPVC Connection across an IISP



To support soft-PVPC and soft-PVCC across the IISP trunk, two information elements (IEs) from PNNI 1.0 signaling are supported in the IISP.

- Called party soft PVPC or PVCC IE
- Calling party soft PVPC or PVCC IE

These two IEs are located in the SETUP and CONNECT messages, and are transported transparently over an IISP interface.

Signaling of VBR-rt

The BPX SES PNNI Controller provides differentiation of VBR-rt from VBR-nrt service category. The VBR-rt service category uses an illegal combination of traffic parameters according to the IISP 1.0, as based on UNI 3.1 specification.

The combination of traffic parameters for VBR-rt are specified in the Broadband Bearer Capability (BBC) IE, as follows:

- Bearer Class—**X**
- Traffic Type—**Not CBR**
- Timing Requirements—**end-to-end**

Because this illegal combination is not allowed on a standard IISP (based on UNI 3.1) interface, a conversion to an acceptable value:

Traffic Type = no indication and Timing Requirements = end-to-end

takes place before the BBC IE is forwarded to the IISP interface. If the illegal combination must be transported without any modification, the IISP interface must be specified to accept the illegal combination by using the **cnfenhiisp** command. (See Appendix B, “SVC, SPVC, and PNNI Commands”, for command syntax.)

Invalid traffic parameters is the default setting.

Signaling of Transparent VP Services

The BPX SES PNNI Controller provides signaling of transparent VP services based on UNI 4.0. The combined parameters in the Connection Identifier IE and the Broadband Bearer Capability IE are supported and transported transparently over an IISP interface.

According to UNI 4.0 Section 4.5.16/Q.2931, the preferred/exclusive field (bits 3-1 of Octet 5) in the Connection Identifier IE is set to 100 for PVP.

Bits

3 2 1

1 0 0 Exclusive VPCI; no VCI (used for switched VPCs)

According to UNI 4.0 Table 4-8/Q.2931, the Bearer Class field (bits 5-1 of Octet 5) in the Broadband Bearer Capability IE is set to 11000 for transparent VP service.

Bits

5 4 3 2 1

1 1 0 0 0 Transparent VP service

Signaling of Soft-PVPC

The BPX SES PNNI Controller supports varied bearer capabilities, Traffic Parameters and QoS as specified in the UNI 4.0 specification to support VP services.This feature enables switched virtual paths to use the UNI 3.1 conformance definitions. (See UNI 4.0 specification, NOTE 5, p108, A9.3 “Allowed Combination of Bearer Capabilities, Traffic Parameters, and QoS”, p105.)

Transport of Generic Identifier Transport IE

The Generic Identifier Transport IE (GITIE) is transported transparently across the Cisco PNNI network. The necessary IE mapping between IISP based on UNI 3.1 and UNI 4.0/PNNI 1.0 is supported at the edge node of the PNNI network.

Static Routing

Static routes are used to forward SVC/SPVC connection requests to pre-determined egress ports. Call forwarding using static routes is with hop-by-hop in nature. Static routes are stored in a separate table accessible by the Route Agent.

Note The Static Routes table must not be shared with the PNNI address table to eliminate unnecessary dependency as well as route leakage management

IISP/UNI Partitioning

The BPX SES PNNI Controller supports resource partitioning between UNI and IISP (in particular. between PVC/PVP with AutoRoute and SPVC/SVPV with PNNI) on a line interface. Configuration of the partition is done on the BPX switch.

Partitioning can only be done on a non-PNNI link. On a UNI link, provisioning can be done for AutoRoute UNI, PNNI UNI, and IISP on single port. A static route is associated with a set of attributes that include:

- the address prefix
- the egress port id
- the flag to indicate if need to be advertised by PNNI routing protocol
- the advertising scope if is advertised by PNNI

IISP CAC and Overbooking

The CAC and overbooking functions on and IISP link operate the same way they would on any other link.

The utilization percentage (overbooking) factor for each service category can be configured on any interface (including UNI, PNNI and IISP).

Policing Enable/Disable on IISP

You can enable or disable policing on an IISP interface. This feature is supported on the BPX switch but not on the controller.

Policing disabled is the default setting.

Frame Discard

Prior to UNI 4.0, the presence of the AAL5-AAL parameters IE in the SETUP message was used to indicate frame discard in both the forward and backward directions by using **cnffdonaal5** command.

This IE is transported transparently across the Cisco PNNI network.

IISP Crankback

A crankback is a call that has been diverted back to the originating DTL due to inaccurate resource and connectivity information from the DTL originator. Inaccurate information may be due to changes in the availability of resources, changes in topology, and hierarchical aggregation. Proper cause code is required for all instances of crankback.

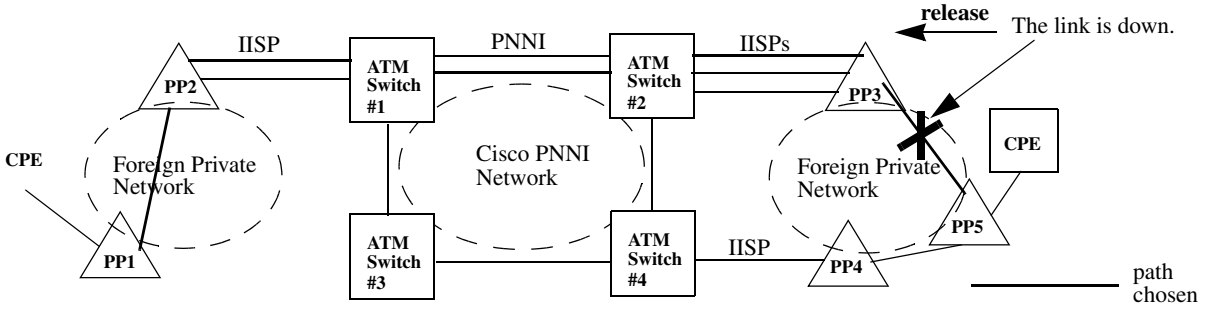
At the DTL originator, follow-up rules for handling the crankback are as follows:

- Select an alternate path for the call, if available.
- Be sure that the alternate path obeys all higher-level DTLs.
- Be sure that the alternate path avoids blocked node(s) and link(s).
- Continue to crankback the call, if necessary.

For example (Figure 2-10) a call from *PP1* to *PP5* fails due to the link down between *PP3* and *PP5*, a RELEASE message is sent back from *PP3* to ATM Switch *#2*. IISP crankback will be supported as follows:

- ATM Switch *#2* receives the RELEASE message from *PP3* and does a local crankback by trying out the other IISP trunks.
- If all local crankbacks fail on Switch *#2*, Switch *#2* generates a crankback to Switch *#1*. Switch *#1* then performs the standard PNNI entry node crankback.

Figure 2-10 IISP Crankback



PNNI and IISP Compliance

See Appendix A, “Technical Specifications”, for the PNNI and IISP compliance information