3

Theory of Operation: Protocols

Protocol Server

Most ports in an Ixia chassis operate a Protocol Server. The Protocol Server includes a complete TCP/IP stack, allowing different forms of high-level DUT testing. The Protocol Server can be configured to test a set of provided Level 2 and Level 3 protocols, which include MAC and IP addressing and IP routing. The Protocol Server for Packet over SONET cards omits all MAC configuration items, since POS does not use a MAC layer. The information gathered by the Protocol Server is used within generated frame data, as well.

The Protocol Server can be accessed through the IxRouter Window. Each protocol must be individually enabled for a selected port in the IxRouter Window.

The protocols supported by the Ixia Protocol Server are described in the following sections in this chapter:

Table 3-1. Protocols Supported by Ixia Protocol Server

Address Resolution Protocol (non-POS only) **(includes IP to MAC addressing)**	See *ARP* on page 3-2
Internet Gateway Management Protocol	See *IGMP* on page 3-3
Open Shortest Path First Protocol	See *OSPF* on page 3-4
Open Shortest Path First Protocol Version 3 (for IPv6)	See *OSPFv3* on page 3-7
Border Gateway Protocol	See *BGP4/BGP+* on page 3-8
Routing Information Protocol	See *RIP* on page 3-14
Routing Information Protocol: Next Generation (for IPV6)	See *RIPng* on page 3-16
Intermediate System to Intermediate System (Dual Mode)	See *ISISv4/v6* on page 3-17

Table 3-1. Protocols Supported by Ixia Protocol Server

Resource ReSerVation Protocol: with Traffic Engineering Extensions	See *RSVP-TE* on page 3-20
Label Distribution Protocol	See *LDP* on page 3-27
Multicast Listener Discovery	See *MLD* on page 3-28
Protocol Independent Multicast: Sparse Mode	See *PIM-SM/SSM-v4/v6* on page 3-29
Multi-Protocol Label Switching	See *MPLS* on page 3-34
Bi-Directional Forwarding	See *BFD* on page 3-36
Connectivity Fault Management	See *CFM* on page 3-37
Fibre Channel over Ethernet (FCoE), FCoE Initialization Protocol (FIP) and NPIV	See *FCoE and NPIV* on page 3-39
Precision Time Protocol (PTP)	See *Precision Time Protocol (PTP) IEEE 1588v2* on page 3-41

There are additional sections on the following topics:

- *ATM Interfaces* on page 3-46
- *Generic Routing Encapsulation (GRE)* on page 3-53
- *DHCP Protocol* on page 3-57
- *Ethernet OAM* on page 3-59

ARP

The Address Resolution Protocol (ARP) facility controls the manner in which ARP requests are sent. This option is only available on Ethernet load modules. The resulting responses from ARP requests are held in the ARP Table, which is used to set MAC addresses for transmitted data. ARP'ing the Device Under Test (DUT) allows tests and generated frames to be configured with a specific IP address, which at run time is associated with the MAC address of that particular DUT.

IP

The IP table within the ARP window specifies a per-port correspondence between IP addresses, MAC addresses (for Ethernet ports only), and the Default

Gateway. IP addresses may be expressed as individual addresses or as a range of addresses.

All ARP requests (for Ethernet) are sent to the Default Gateway address. In most cases, the Default Gateway Address is the address of the DUT. When a gateway separates the Ixia port from the DUT, use the IP address of that gateway as the Default.

IGMP

The Internet Group Management Protocol (IGMP) is used with IPv4 to control the handling of group membership in the Internet. Version 3, specified in RFC 3376, is supported and is interoperable with Versions 1 and 2. Version 1 of the protocol is specified in RFC 1112, and Version 2 is specified in RFC 2236.

IGMP normally works in an environment in which there are a number of IGMP-capable hosts connected to one or more IGMP routers. The routers forward membership information and packets to other IGMP routers and receive group membership information and packets from other IGMP routers.

The Ixia hardware simulates one or more hosts, while the DUTs are assumed to be IGMP routers. The simulation calls for groups of simulated hosts to respond to IGMP router-generated queries and to automatically generate reports at regular intervals. A number of IGMP groups are randomly shared across a group of hosts.

Version 3 adds the concept of filtering, based on the IP source address, to cut down on the reception of unwanted multicast traffic. This filtering consists of limiting the receipt of packets to only those from specific sources (INCLUDE) or to those from all but specific sources (EXCLUDE). Refer to *MLD* on page 3-28 for information about similar functions for multicast traffic in IPv6 environments.

Compatibility with earlier versions of IGMP is an important part of IGMPv3. The Group Compatibility Modes for an IGMPv3 router are summarized as follows:

- IGMPv3 Compatibility Mode (default): An IGMPv2 and/or IGMPv1 Host is present, but NOT running.
- IGMPv2 Compatibility Mode: An IGMPv2 Host may be present and running. An IGMPv1 Host is present, but NOT running.
- IGMPv1 Compatibility Mode: An IGMPv1 Host is present and running.

OSPF

Note: See also *OSPFv3* on page 3-7.

Open Shortest Path First (OSPF) is a set of messaging protocols that are used by routers located within a single Autonomous System (AS). The Ixia hardware simulates one or more OSPF routers for the purpose of testing one or more DUT routers configured for OSPF. The OSPF version 2 specification (RFC 2328) details the message exchanges by OSPF routers, as well as the meanings and usage.

OSPF has the following three principal stages:

- The HELLO Protocol
- Database transfer
- HELLO Keepalive

When an OSPF router initializes, it sends out HELLO packets and learns of its neighboring routers by receiving their HELLO packets. If the router is on a Point-to-Point link, or on an Ethernet (transit network) link, these packets are addressed to the *AllSPFRouters* multicast address (224.0.0.5). In these types of networks, there is no need to manually configure any neighbor information for the routers.

Each router that is traversed on the path between neighbors is added to a list contained in the HELLO packet. In this way, each router discovers the shared set of neighbors and creates individual state machines corresponding to each of its neighbors.

If the network type is broadcast, then the process for selecting a Designated Router (DR) and Backup Designated Router (BDR) begins. A Designated Router is used to reduce the number of adjacencies required in a broadcast network. That is, if no Designated Router is used, then each router must pair (form an adjacency) with each of the other routers. In this case, the number of required adjacencies is equal to the square of the number of routers (N^2). If a DR and BDR are used, the number of required adjacencies drops to 2 times the number of routers (2N). Currently, the Ixia ports are unable to simulate a DR or BDR.

Once the routers have initialized their adjacency databases, they synchronize their databases. This process involves one router becoming the master and the other becoming the subordinate. On Ethernet networks, the DR is always the master; on point-to-point networks, the router with the highest Router ID is the master.

Link State Advertisements (LSAs) are OSPF messages that describe an OSPF router's local environment. The simplest LSA Type is the router-LSA (RouterLinks LSA). Each router is required to generate exactly one of these LSAs to describe its own attached interfaces. If a network that consists of a single OSPF area is being simulated with only point-to-point links and there are no

Autonomous System Border Routers (ASBR), then this is the only type of LSA that is sent.

The subordinate asks the master for its LSA (Link State Advertisement) headers, which enables the subordinate to determine the following information:

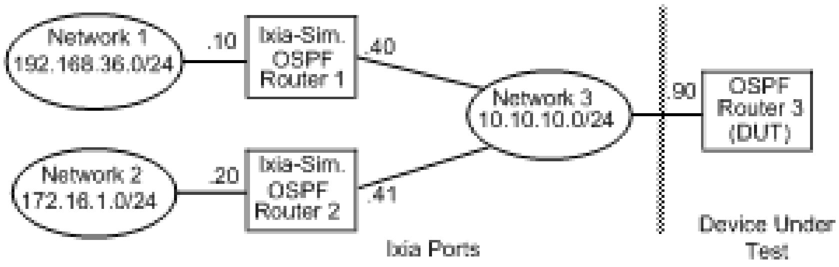
1. The subset of LSAs that the master holds, but that the subordinate does not have, **and**
2. The subset of LSAs that the master and subordinate both have, but which are more recent on the master.

The subordinate router then proceeds to explicitly query the master to send it each LSA from Steps (1) and (2). The subordinate sends an ACK to the master upon receipt of each LSA. The global Link State Database (LSDB) is constructed by each router, based on LSAs from all the other routers in the network.

Once this exchange process is complete, the routers are considered to have reached Full Adjacency, and each runs the link state algorithm to update its IP forwarding tables. The routers continue to exchange periodic HELLO packets, as keepalive messages, until a change occurs (for example, a link goes down or an LSA expires). OSPF routers continue to periodically exchange their LSAs every 30 minutes to ensure that they all hold identical LSDBs.

This section describes the programming of the Ixia hardware related to OSPF testing, as well as the theory of operation and protocol message formats. The Ixia hardware simulates multiple OSPF routers on multiple networks. For example, in Figure 3-1 on page 3-5 there are three networks and three routers.

Figure 3-1. Sample OSPF Network



The Protocol Server calls for the specification of router-network connections to be specified in a network-centric fashion. One specifies the network in terms of an Area ID and network mask. One specifies the routers in terms of the interface IP address on that network and Router ID, usually the lowest IP address for the router. For the sample OSPF network, in which Router 3 is the DUT, the three networks are specified by their significant characteristics as shown in Table 3-2 on page 3-6.

Table 3-2. Sample OSPF Network Assignments

Network	Area ID	Network Mask	Router ID	Router Interface IP Address
1	192.168.36.0	255.255.255.0	192.168.36.10	10.0.0.40
2	172.16.0.0	255.255.255.0	172.16.0.20	10.0.0.41
3	10.0.0.0	255.255.0.0	10.0.0.40 10.0.0.41 10.0.0.90	10.0.0.40 10.0.0.41 10.0.0.90

Within this framework, Link State Advertisements (LSAs) may be issued from the perspective of any interface on any router. Any OSPF messages from the DUT Routers may be captured and analyzed in the normal manner.

OSPFv3

Open Shortest Path First Protocol Version 3 supports Internet Protocol version 6 (IPv6), as defined in RFC 2740. The 128-bit IPv6 addressing scheme has been accommodated in OSPF through the use of new LSA types.

Some of the differences between OSPFv2 (for IPv4) and OSPFv3 (for IPv6) are listed as follows:

- Changes to adapt to the IPv6 128-bit address size. No addresses are carried in OSPF packets or basic LSAs, but addresses are carried in certain LSAs.
- OSPFv3 operation is per Link, with the IPv6 concept of ‘link’ replacing the ‘IP subnet’ and ‘network’ terminology of OSPFv2.
- OSPVv3 supports multiple instances of the protocol per link, through ‘Instance IDs.’
- LSA flooding scope is explicitly defined in the LS Type field of each LSA.
- Authentication is handled by the IPv6 protocol itself, rather than by the OSPF protocol. For this reason, Authentication information has been removed from the packet headers in OSPFv3.

Note: In OSPFv2, IPv4 addresses were used in many contexts besides IP source and destination addresses. For example, they were assigned as name identifiers for routers (RIDs). This naming convention for RIDs has been retained in OSPFv3.

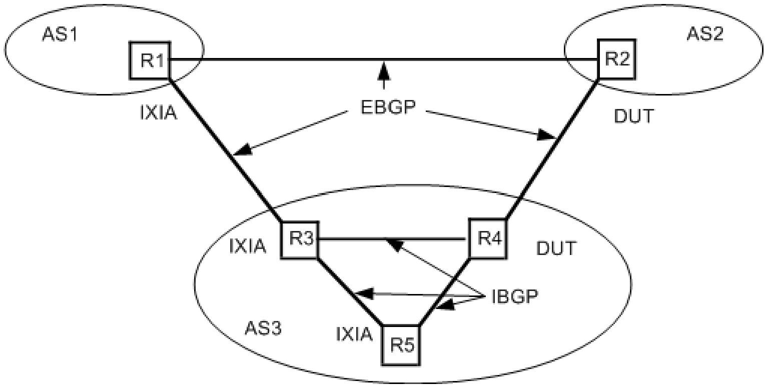
BGP4/BGP+

Border Gateway Protocol Version 4 (BGP-4) is the principal protocol used in the Internet backbone and in networks for large organizations. The BGP4 specification (RFC 1771) details the messages exchanged by BGP routers, as well as their meaning and usage. *BGP4 - Inter-Domain Routing in the Internet*, by John W. Stewart III is a descriptive reference on this protocol.

Internal Versus External BGP

The BGP4 protocol is used according to two sets of rules, depending on whether or not the two communicating BGP routers are within the same Autonomous System (AS). An AS is a collection of routers that implement the same routing policy and are typically administered by a single group of administrators. ASs connected to the Internet are assigned Autonomous System Numbers (ASNs) that are key to inter-domain routing. When BGP is used **between** two ASs, the protocol is referred to as EBGP (External BGP); when BGP is used **within** an AS it is referred to as IBGP (Internal BGP). Figure 3-2 on page 3-8 depicts the differences in topology between EBGP versus IBGP.

Figure 3-2. External BGP Versus Internal BGP



In the figure above, AS1, AS2, and AS3 are distinct Autonomous Systems. The R<u>n</u>s are routers in the various ASs. Routers on the links between ASs ‘speak’ EBGP, while the routers within AS3 ‘speak’ IBGP.

IBGP Extensions

In the original BGP4 specification (RFC 1771), all IBGP routers within an AS are required to establish a full mesh with each other. This leads to a lack of scalability which is solved by the introduction of two additional concepts: *route Reflection* and *Confederations.*

In route reflection, some routers in an AS are assigned the task of re-distributing internal routes to other internal AS routers. To prevent looping within an AS that uses route reflection, two concepts are important: the *originator-id* and *cluster-list* attributes.The originator-id is the identification of the router that originated a particular route. Routers within an AS propagate this information and refuse to send a route back to its originator. Even the use of route reflectors and originator-ids can lead to scalability problems in an AS. The cluster-list concept helps this problem. A cluster consists of a reflecting router and its clients. A Cluster ID is the IP address of the reflecting router if there is one, or a configured number otherwise. A cluster-list is a constructed list, consisting of the cluster IDs of all of the clusters that a route has passed through. Each router refuses to send a route back to a cluster that has seen the route already.

In a confederation, an AS is divided into multiple sub-confederation subsets. Each sub-confederation is defined in terms of its own ASN and a list of routers. Routers within a sub-confederation are expected to fully mesh using IGP. Sub-confederations within a confederation speak a variant of EGP, called EIGP. Additional path attributes are used with a confederation to indicate paths that should not be propagated outside the confederation.

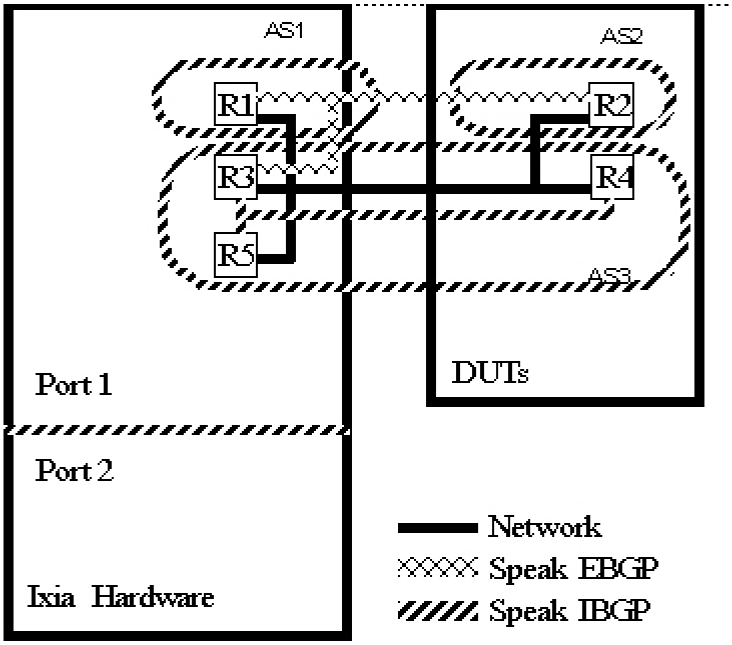
Communities

In deployment of BGP4 into a growing Internet environment, it became necessary to deal with certain routes in different manners not related to the strict routing of packets. The community attribute was invented to allow a route to be ‘tagged’ with multiple numbers, called communities. This is also referred to sometimes as *route coloring.*

BGP Router Test Configuration

The Ixia Protocol server implements an environment in which the Ixia hardware simulates multiple routers which speak IBGP and/or EBGP with one or more DUT routers. For example, in Figure 3-2 on page 3-8, the Ixia hardware emulates R1, R3, and R5 while the DUTs are R2 and R4. The following figure depicts the same setup based on the location of the simulated or actual router:

Figure 3-3. BGP Interconnection Environment

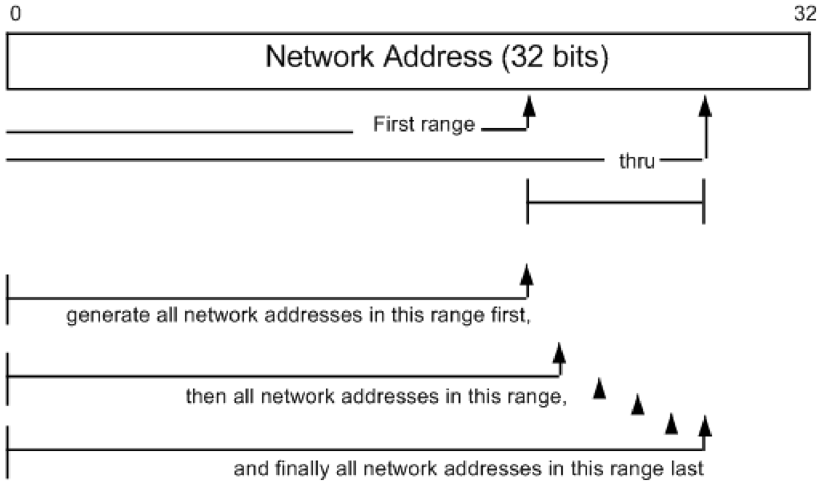


All of the routers are logically connected through appropriate networking hardware. The Ixia hardware is used to simulate three of the routers in two different ASs communicating with two routers being tested.

A single router emulated by the Ixia hardware is specified by a single IP address, and a number of emulated routers may be specified by a range of IP addresses. Each DUT router is identified by its IP address.

Messages may be sent between the emulated routers and the DUT routers when a connection is made and one of the two endpoints sends an OPEN message. Where the emulated routers and the DUT routers send their OPEN messages simultaneously, standard collision handling is applied. Thereafter, the emulated routers send a number of UPDATE messages to the DUT routers. The UPDATE messages contain a number of network address ranges (route ranges), also known as ranges of prefixes. The ranges of generated network addresses is illustrated in Figure 3-4 on page 3-10.

Figure 3-4. Generation of Network Addresses in BGP UPDATE Messages


0
32
Network Address (32 bits)
First range
thru
generate all network addresses in this range first,
then all network addresses in this range,
and finally all network addresses in this range last

A designated number of network addresses are generated with network Mask Width with the *From* through *To* values. Table 3-3 on page 3-11 shows some examples of generated addresses. Network Addresses are generated by starting with the First Route and *From* mask width up to, but not including 224.0.0.0. (127.*.*.* is also skipped). If the requested number of network addresses has not been generated before 224.0.0.0 is reached, then the next mask length is used with the First Route to generate network addresses.

Table 3-3. Examples of Generated BGP Routes (Network Addresses)

First Route	Mask Width From	Mask Width To	Iterator Step	Number of Routes	Generated BGP Routes (Network Addresses)
192.168.36.0	24	26	1	(14,378,756 Max.)	192.168.36.0/24 192.168.37.0/24 192.168.38.0/24 ... 223.255.255.0/24 (224.0.0.0+ skipped) 192.168.36.0/25 192.168.36.128/25 192.168.37.0/25 ... 223.255.255.128/25 (224.0.0.0+ skipped) 192.168.36.0/26 192.168.36.64/26 192.168.36.128/26 ... 223.255.255.192/26
204.197.56.0	24	24	10	4	204.197.56.0/24 204.197.66.0/24 204.197.76.0/24 204.197.86.0/24

All of the generated network addresses are associated with a set of attributes that describes routing to these generated network addresses and associated features.

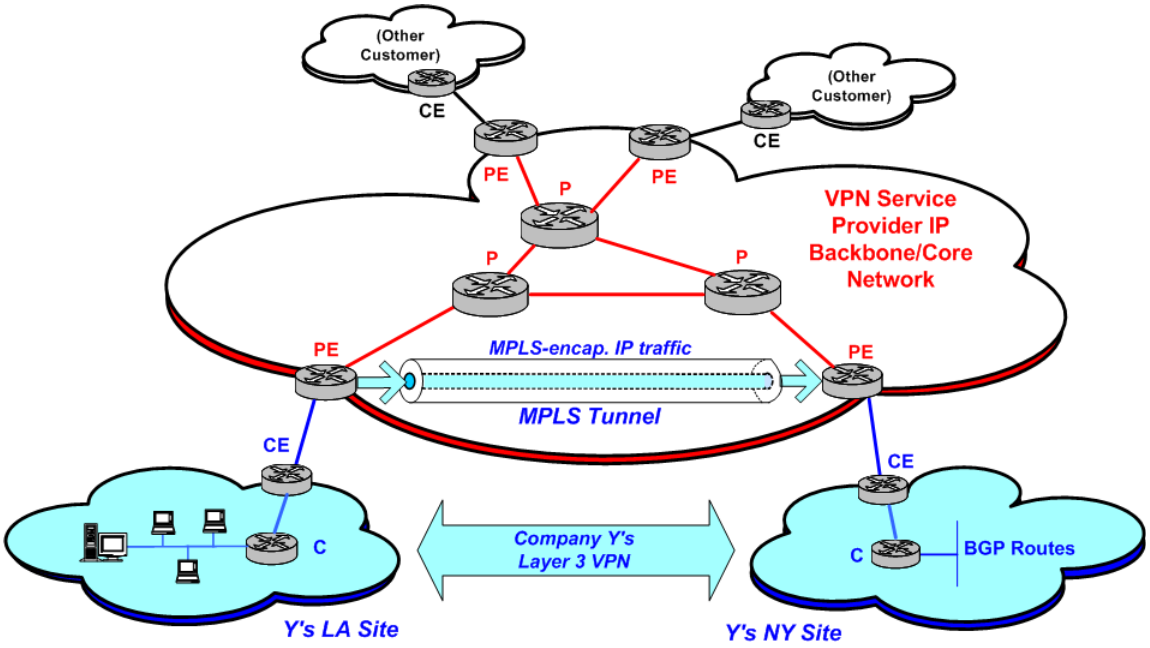
Only one route can be added per UPDATE message, but a variable number of *withdrawn* routes may be packed into each UPDATE message. The packing is randomly chosen across a range of a number of routes. The time interval between UPDATE messages is configurable, in units of milliseconds.

A BGP4 network condition called ‘flapping’ can be emulated by the protocol server on an Ixia port. In the Link flapping emulation, a peer BGP router appears to be going offline and online repeatedly, which is accomplished on the Ixia port by alternate disconnects and reconnects of the TCP/IP stack. In the Route flapping emulation, BGP routes are repeatedly withdrawn, and then re-advertised, in UPDATE messages.

BGP L3 VPNs

L3 Virtual Private Networks (VPNs) over an IP backbone (at Layer 3 of the OSI model), may be provided to the customers of a Service Provider (SP), providing connectivity between two or more sites owned by the customer. L3 VPNs are independent of the Layer 2 protocol. While MPLS handles the packet forwarding in the backbone/core, the BGP protocol provides a means of advertising external routes/network addresses across that backbone between sites. IETF Internet Draft 'draft-ietf-ppvpn-rfc2547bis-01.txt,' the proposed successor to RFC 2547, covers the VPN architecture designed for use by private service providers. A simplified example of a BGP L3 VPN topology is shown in Figure 3-5 on page 3-12.

Figure 3-5. Simplified BGP L3 VPN Diagram



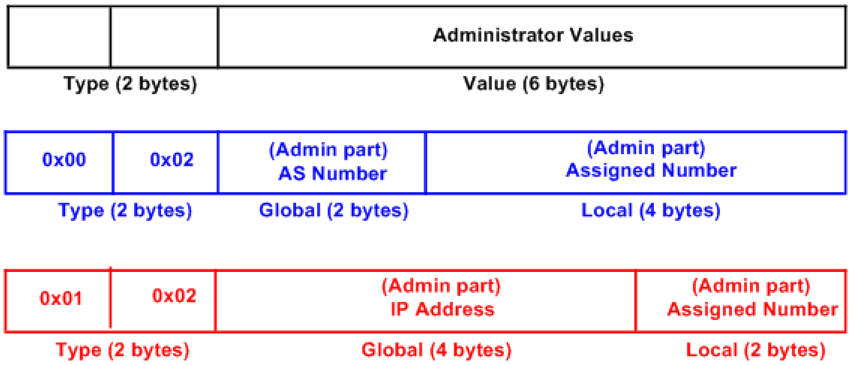
The term *site* refers to a customer/client site, which consists of a group of interconnected IP devices, usually in one geographic location. A Customer Edge (CE) device, typically a router, connects the site, through a data link connection, to a Provider Edge (PE) router—an entry point to the service provider's backbone. The PE-to-CE routing protocols may be static routing, or a dynamic protocol such as eBGP or RIPv2.

Provider (P) network core routers, 'transparently' carry the IP traffic across the internal core between CE routers. CEs and Ps are not 'VPN-aware' devices. CE devices are considered as belonging to a only one site, but that site may belong to multiple VPNs. A VPN Routing and Forwarding table (VRF) on a PE consists of an IP routing table, a forwarding table, and other information on the set of interfaces in the VPN. The VRF generally describes a VPN site's routing information, and a PE may maintain multiple VRFs, one for each connected customer site. See *L3 VPN VRFs* on page 3-14 for additional information on VRFs.

Layer 3 VPN sites are identified by a Route Target (RT). A route target is based on the mechanism proposed in the IETF draft for the 'BGP Extended

Communities Attribute.’ An 8-byte route target is common to all route ranges that belong to a single L3 site. Route targets are defined for individual VPN route ranges. The formats for Route Targets (RTs) are shown in Figure 3-6 on page 3-13.

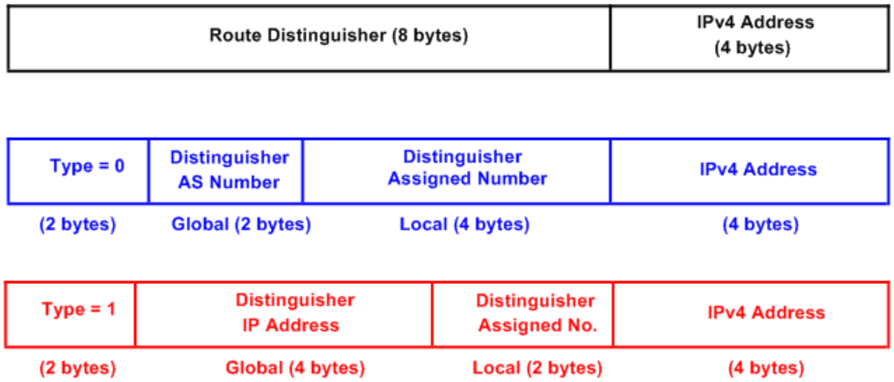
Figure 3-6. Route Target Formats (BGP Extended Community Types)



BGP VPN-IPv4 Address Formats

Globally unique 12-byte VPN-IPv4 prefixes are created by a PE router. This includes configuration of the 8-byte VPN Route Distinguishers (RDs). It should be noted that BGP IPv4 routes and VPN -IPv4 routes are considered noncomparable; VPN-IPV4 addresses can be used only within the VPN service provider network.The route distinguishers are used by PE routers to associate routes with the path to a particular CE site router in a VPN. Each route can only have one RD. The formats of the RDs are shown in Figure 3-7 on page 3-13.

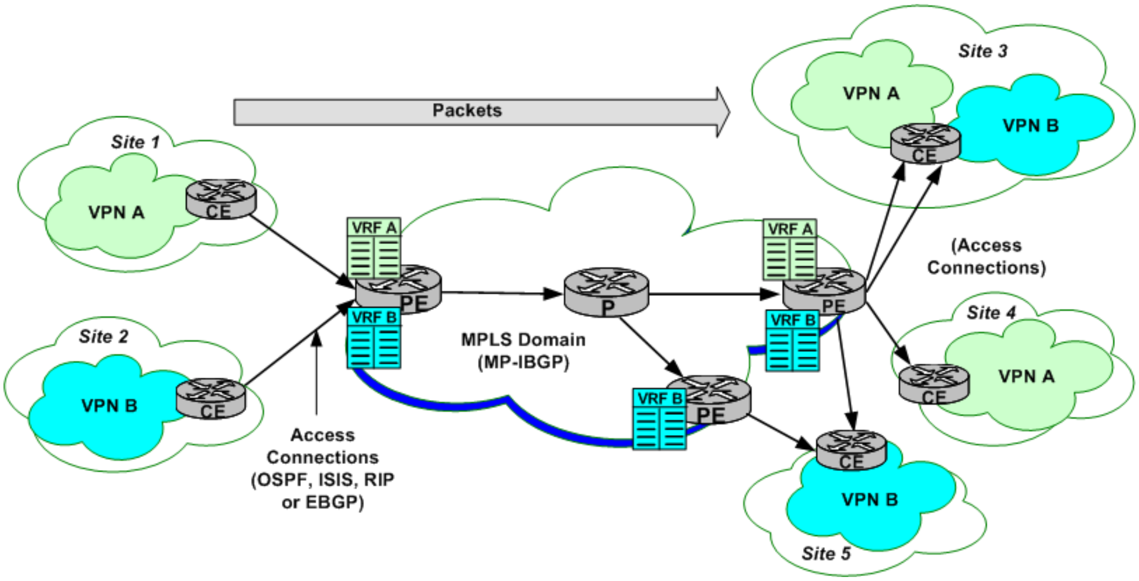
Figure 3-7. VPN-IPv4 Address Formats (with Route Distinguishers)



L3 VPN VRFs

For Layer 3 Virtual Private Network (L3 VPN) configurations, the Provider Edge (PE) routers maintain routing tables for each VPN that they participate in, termed VPN Routing and Forwarding tables (VRFs). The VRFs are populated with routes received from both the directly attached and remote Customer Edge (CE) routers. Each entry in the VRF is called a VPN Forwarding Instance (VPI). VRFs and CEs are not required to be configured on a one-to-one basis, although this is the typical situation. An example of the possible relationships between VRFs and CEs is shown in Figure 3-8 on page 3-14.

Figure 3-8. L3 VPN VRF Example



RIP

The Routing Information Protocol (RIP) is an interior routing protocol. It is the oldest and most frequently used of the LAN routing protocol. RIP routers broadcast or multicast to each other on a regular basis and in response to REQUEST packets. RIP routers incorporate routing information received from their neighbors into their own routing table and forward them on to other neighbors. Two distinct versions of RIP exist: version 1 and version 2. Both IPv4 and IPv6 are supported.

As implemented by the Protocol Server, each Ixia port is capable of simulating one or more routers at distinct addresses. Routing tables for the simulated routers are configured by you and sent out at regular intervals, with a configurable randomizing factor. Either version 1 or version 2 packet formats may be sent through multicast or broadcast (for compatibility with version 1 routers). Received packets may be filtered for version 1 and/or 2 compatibility.

The current implementation of the Protocol Server uses Split Horizon with Space Saver as its update mode, which receives, but not process RIP broadcasts heard from DUT routers. That is, it does **not** incorporate received information into its own table, but rather always broadcast the same routing table. Future versions will offer Split Horizon, Split Horizon with Poison Reverse, and Silent modes of update.

The Protocol Server, however, responds to REQUEST packets that it receives. Two types of requests are processed:

- Request for all routes: The Protocol Server sends the same routing table that it sends at regular intervals back to the requestor.
- Request for specific routes: The Protocol Server fills in the requested information in the received packet and send it back to the requestor.

RIP Overview

The Routing Information Protocol (RIP) is an interior gateway routing protocol (IGP) and uses a Distance Vector Algorithm. It is the oldest and most frequently used of the LAN routing protocols. RIP routers broadcast or multicast to each other on a regular basis and in response to REQUEST packets. RIP routers optionally incorporate routing information received from their neighbors into their own routing table and forward it on to other neighbors.

Note: For information on **RIPng** (RIP-Next Generation), based on IPv6, see *RIPng* on page 3-16.

As implemented by the Protocol Server, each Ixia port is capable of simulating one or more routers with separate addresses. Routing tables for the simulated routers are configured by you and sent out at regular intervals, with a configurable randomizing factor. Either Version 1 or Version 2 packet formats may be sent through multicast or broadcast (for compatibility with Version 1 routers). Received packets may be filtered for Version 1 or Version 2 compatibility.

The current implementation of the Protocol Server uses Split Horizon with Space Saver as its update mode, which receives, but not process, RIP broadcasts heard from DUT routers. That is, it does **not** incorporate received information into its own table, but rather always broadcast the same routing table. Future versions will offer Split Horizon, Split Horizon with Poison Reverse, and Silent modes of update.

The Protocol Server, however, responds to REQUEST packets that it receives. Two types of requests are processed:

- Request for all routes: The Protocol Server sends the same routing table that it sends at regular intervals back to the requestor.
- Request for specific routes: The Protocol Server fills in the requested information in the received packet and send it back to the requestor.

RIPng

Routing Information Protocol - Next Generation (RIPng) is specified for use with IPv6 in RFC 2080. Like the IPv4 version of RIP, this routing protocol is based on a Distance Vector algorithm. RIPng routers compare information for various routes through an IPv6 network, especially the information related to the RIPng metric. Due to the limited number of allowed hops, this protocol is used in small- to moderate-sized networks. The valid metric range is from 1 to 15 (hops). The metric values of 16 and above are defined as 'infinity' and are considered unreachable.

An RIPng router is assumed to have interfaces to one or more directly-connected networks. Each router maintains a routing table, with one entry for every reachable destination in the RIPng network. Each routing table entry contains a minimum of:

- IPv6 destination prefix(es)
- total metric cost for the path to the destination(s)
- IPv6 address of the next hop router
- a 'route change flag'
- timers

As a UDP-based protocol, the RIPng routing process functions on UDP well-known port number 521 (the 'RIPng port'), on which datagrams are sent and received. The RIPng port supports the following:

- Receives all communications received from another router's RIPng process.
- Sends all RIPng routing update messages.
- Unsolicited routing update messages specify this port as the source and destination.
- Responses to request messages are sent to the originating UDP port.
- Specific requests need not come from the RIPng port, but the destination on the targeted device must be the RIPng port.

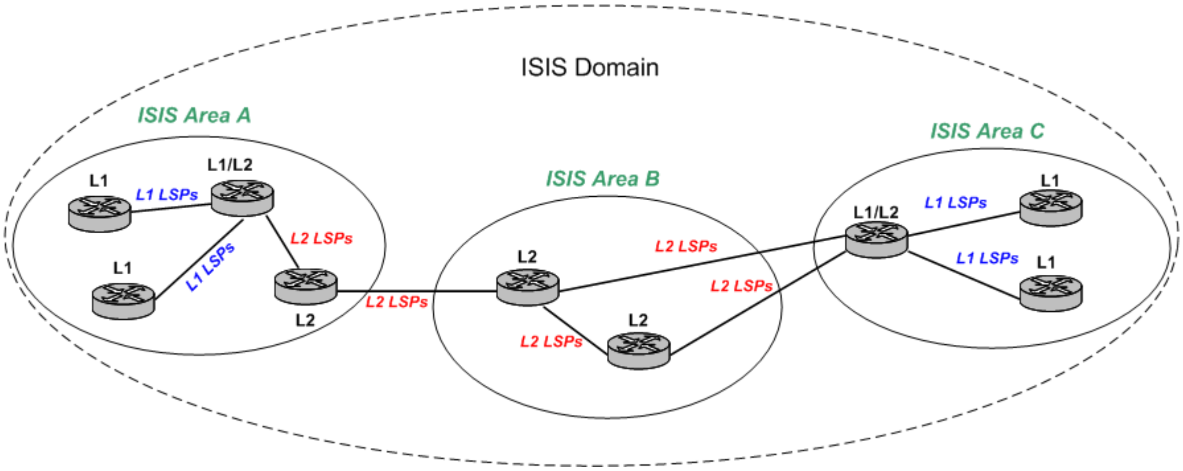
ISISv4/v6

The Intermediate System to Intermediate System (ISIS) routing protocol was originally designed for use with the OSI Connectionless Network Protocol (CLNP) and was defined in ISO DP 10589. It was later extended to include IP routing in IETF RFC 1195. When routing for OSI and IP packets (defined in ISO/IEC 10589:1992(E)) is combined in this way, the protocol is referred to as Integrated ISIS or Dual ISIS. In addition, RFC 2966 extends the distribution of routing prefixes among ISIS routers, and IETF DRAFT draft-ietf-isis-ipv6-05 adds IPv6 routing capability to the protocol.

ISIS Topology

ISIS areas are administrative domains which contain ISIS routers, have one or more private networks, and may share networks with other areas. The example shown in Figure 3-9 on page 3-17 consists of a theoretical ISIS topology. Note that, as shown in this diagram, all ISIS routers are considered to reside entirely **within** an area, unlike some other protocols such as OSPF, where routers can reside at the edges of areas and domains.

Figure 3-9. ISIS Topology



One or more Area IDs are associated with an area. Most areas only require one ID during steady state operation, but up to three IDs may be needed during the process of migrating a router from one area to another. In most cases, the maximum number of area IDs is set to three.

ISIS routers can be divided into three categories, as follows:

- **Level 1 (L1)**: These routers can connect only to L1 or L1/L2 routers within their own area (intra-area). They have no direct connection to any other ISIS area.
- **Level 2 (L2)**: These routers can connect only to other L2 routers outside their area, or to L1/L2 routers within their own area. They are used as backbone routers in the routing domain, to connect ISIS areas.

- **Level 1/2 (L1/L2)**: These routers have separate interfaces which can connect to both L1 routers within their own area and L2 routers in other areas.

Entirely separate routing tables are maintained for Level 1 and Level 2 ISIS information, even within L1/L2 routers. All L1s within an area maintain identical databases. All L2s within a domain maintain identical databases.

ISIS Processing

Many OSI concepts are necessary for describing ISIS. The following terms are important to the following discussion:

- IS - Intermediate System. An ISIS router is an IS.
- ES - End System. A host is an ES. (Note: The Ixia hardware does not currently simulate End Systems.)
- PDU - Protocol Data Unit. PDUs contain messages used for the ISIS protocol. The following PDUs are used in IS-IS communications:
  - IIH - IS-to-IS Hello PDU. This message is multicast over broadcast networks, or unicast on point-to-point links, between ISs to discover neighbors and maintain ISIS state.
  - LSP - Link State PDU. This message holds the significant part of the routing table sent between ISIS routers.
  - SNP - Sequence Number PDU. This message is used to request LSPs and acknowledge receipt of LSPs. Two types are used depending on the network type:
    - CSNP - Complete SNP. In broadcast networks, these are sent by the Designated Router in an area. On point-to-point connections, CSNPs are used for initialization. A CSNP contains a complete description of the LSPs in the sender's database.
    - PSNP - Partial SNP. On broadcast networks, PSNPs are used to request LSPs. On point-to-point connections, PSNPs are used to acknowledge receipt of LSPs. On both types of networks, PSNPs are used to advertise newly learned LSPs or purge LSPs. A PSNP contains a subset of the received records.

ISIS routers maintain knowledge of each other by exchanging Hello PDUs at regular, configured *Hello intervals*. A router is considered down if it does not respond within a separately configured *Dead interval.*

ISIS routers update each other using Link State PDUs (LSPs) at a regular interval of 30 minutes. The LSP header contains the Remaining Lifetime for the LSP, a Sequence Number, and a checksum. Each LSP contains information about a router's connection to local networks, plus a metric related to each network. ISO DP 10589 defines four types of metrics: default, delay, expense, and error.

In a Broadcast/LAN network, the Designated Router sends a Complete Sequence Number PDU (CSNP). In a Point-to-Point network, the receiving router sends a Partial Sequence Number PDU (PSNP).

In the ISIS protocol, for each of the levels (L1 or L2), one of the routers is elected as the Designated IS, based on priority values assigned to each interface as part of Hello PDU processing. The Ixia Protocol Server does not support the role of DR, so to ensure that it is not elected by its ISIS peers each Ixia-simulated ISIS router has a default priority of ‘0,’ indicating its unwillingness to be the Designated IS.

ISIS Addresses

Due to the OSI derivation of the ISIS protocol, each ISIS router has an OSI NET address of 8 to 20 octets in length. The NET address consists of two parts: an Area ID and a System ID. The Area ID has a number of different formats defined in OSI specifications. The System ID may be from 1 to 8 octets in length. The default System ID length defaults to 6 octets and must be the same length for every router in the domain. The System ID is unique within its ISIS **area** for Level 1, or unique within the ISIS routing **domain** for Level 2 or Level 1/2. Two types of network connections are supported: broadcast and point-to-point. In a broadcast network, each interface on an ISIS dual-mode router must have an IP address and mask.

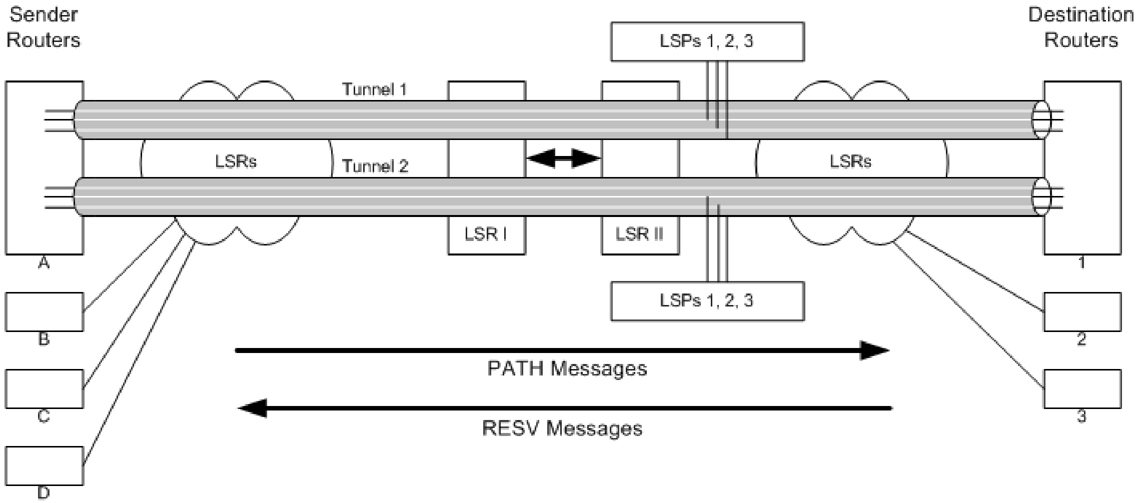
RSVP-TE

The Ixia protocol server implements a part of the Resource Reservation Protocol (RSVP) used for Traffic Engineering (TE). This subset of the RSVP protocol, referred to as RSVP-TE, is used in the process of constructing a path through a sequence of MPLS-enabled label switched routers (LSRs), while reserving necessary bandwidth resources. The use of an internal gateway routing protocol (IGP), such as OSPF, is also required to automatically determine the 'next hop' router.

Multi-Protocol Label Switching (MPLS) allows rapid forwarding of packets across a sequence of routers, without time-consuming examination of the packet contents at each hop. Label switching has been used extensively for ATM traffic, where overhead bytes for each 'cell,' or packet, of data constitute a large percentage of the overall data transmitted. The addition of a 'label' value to the header information in each cell or packet supplies the only forwarding information required to transit the MPLS domain. Based on information in its forwarding table, each LSR replaces (swaps) the incoming label with a new one which directs the packet to the next hop.

The most important output from an RSVP-TE setup session is the set of *MPLS labels*, which are used by the MPLS-enabled routers along the path to efficiently forward network traffic. The operation of RSVP-TE is shown in Figure 3-10 on page 3-20.

Figure 3-10. RSVP-TE Overview



Through the use of RSVP-TE message exchanges, the router at the entry to the MPLS domain, also known as an Ingress LSR, initiates the creation of a dynamic 'tunneled' pathway to the Egress LSR, the router at the exit side of the MPLS domain. Packets which pass through this 'tunnel' are essentially 'protected' from the extensive packet processing normally imposed by each router it traverses. Once this special pathway or Label Switched Path (LSP) is established, the router can **forward,** rather than route, packets across the domain, saving considerable processing time at each intermediate LSR (Transit LSR). The resulting tunneled pathway is known as an LSP Tunnel. The traffic flows through an LSP Tunnel are unidirectional. To establish bidirectional traffic through the MPLS domain, a second LSP Tunnel must be created in the opposite direction.

An LSP Tunnel is defined by a Destination Address (the IP address of the Egress LSR), and a Tunnel ID. At a finer level of granularity are LSP IDs. Essentially, these LSP IDs can serve to provide a set of aliases for alternate hop-by-hop paths between a single pair of Ingress and Egress LSRs, and therefore exist within the same LSP Tunnel.

Note: Ingress LSRs and Egress LSRs are also known as Label Edge Routers (LERs).

Two principal RSVP-TE message types are used to establish LSP Tunnels:

- PATH message. A PATH message is generated by the ingress router and sent toward the egress router. This is termed the *downstream* direction. This PATH message is a request by the sending LSR for the establishment of an LSP to the egress router. Each LSR in the path to the destination router digests the PATH message and does one of three things:
    - If the LSR cannot accommodate the request, it rejects the request by sending a PATH_ERR message back to the source indicating the nature of the rejection.
    - If the LSR is not the egress router, it sends a PATH message to the next LSR toward the destination router.
    - If the LSR is the egress router, it should respond with a RESV message back to its most recent neighbor.
- RESV message. A RESV message is generated by the egress router and sent over the reverse path that the PATH messages took. This is termed the *upstream* direction.

An additional *HELLO* message is used between neighbor LSRs to ensure that LSRs are alive. This allows for quick tunnel replacement in the case of link or router failure.

A set of labels is passed in the RESV messages sent upstream from the egress to the ingress router. A label is sent from one LSR to its upstream neighbor telling the upstream router which label to use when later sending downstream traffic.

Three scenarios are currently supported to test MPLS/RSVP-TE on a DUT using Ixia equipment:

1. The DUT acts as the Ingress LSR, and the Egress LSR is simulated by an Ixia port.
2. The DUT acts as the Egress LSR, and the Ingress LSR is simulated by an Ixia port.
3. The DUT acts as a Transit/Intermediate LSR, and the Ingress and Egress LSRs are simulated by Ixia ports.

PATH Messages

PATH messages contain a number of objects which define the tunnel to be established. These are shown in Table 3-4 on page 3-22.

Table 3-4. RSVP-TE PATH Message Objects

Object	Contents	Usage
SESSION		Describes the destination router and associates a tunnel ID with the session.
	tunnel endpoint	The destination router’s IP address.
	tunnel ID	A unique LSP tunnel ID.
SENDER_TEMPLATE		The description of the sender.
	tunnel sender address	The sender router’s IP address.
	LSP ID	A unique LSP ID.
LABEL_REQUEST		Asks all the LSRs to send back label values through RESV messages.
SENDER_TSPEC and ADSPEC		Both of these objects deal with bandwidth and other QoS requirements for the path.
TIME_VALUES		Timing values related to the refresh of tunnel information.
	refresh interval	The interval between messages.
EXPLICIT_ROUTE		Allows the sender to request that the LSP tunnel follow a specific path from ingress to egress router. See *Explicit_Route* on page 3-22 for more details.
SESSION_ATTRIBUTE		Other attributes associated with the session: tunnel establishment priorities, session name, and optionally resource affinity.
RSVP_HOP		Describes the immediate upstream router’s address to the downstream router.

Explicit_Route

An explicit route is a particular path in the network topology. Typically, the explicit route is determined by a node with the intent of directing traffic along that path. An explicit route is described as a list of groups of nodes along the explicit route. In addition to the ability to identify specific nodes along the path, an explicit route can identify a group of nodes that must be traversed along the path. Each group of nodes is called an *abstract node*. Thus, an explicit route is a specification of a set of abstract nodes to be traversed.

There are three types of objects in an explicit route:

- IPv4 prefix
- IPv6 prefix
- Autonomous system number

Each node has a *loose* bit associated with it. If the bit is not set, the node is considered *strict*. The path between a strict node and its preceding node may only include network nodes from the strict node and its preceding abstract node. The path between a loose node and its preceding node may include other network nodes that are not part of the strict node or its preceding abstract node.

RESV Message

The RESV message contains object that indicate the success of the PATH request and the details of the assigned tunnel. These are shown in Table 3-5 on page 3-23.

Table 3-5. RSVP-TE RESV Message Objects

Object	Usage
SESSION	Indicates which session is being responded to.
TIME_VALUES	As in the PATH message but from the downstream LSR to the upstream LSR.
STYLE	The type of reservation assigned by the egress router. This relates to whether individual tunnels are requested for each sender-destination connection or whether some connections may use the same tunnel.
FILTER_SPEC	The sender router's IP address and the LSP ID.
LABEL	The label value assigned by the downstream router for use by the upstream router.
RECORD_ROUTE	If requested, the complete route from the destination back to the source. The contents of this object include the IP addresses in either v4 or v6 format of all the LSRs encountered in the formation of the LSP, and optionally the labels used at each step. Each LSR on the upstream path perpends its own address information.
RESV_CONF	If present, it indicates that the ingress router should send a RESV_CONF message in response to the destination to indicate that the tunnel has been completely established.

Other Messages

Several additional messages are used in RSVP-TE, as explained in Table 3-6 on page 3-24.

Table 3-6. Additional RSVP-TE Messages

Message	Usage
PATH_ERR	Any LSR may determine that it cannot accommodate the tunnel requested in a PATH message. In this case it sends a PATH_ERR message back to the sender.
PATH_TEAR	When a sender router determines that it wants to tear down a tunnel, it sends a PATH_TEAR message to the destination router.
RESV_ERR	If a router cannot handle a reservation, it sends a RESV_ERR back to the destination router.
RESV_TEAR	When a destination router determines that it wants to tear down a tunnel, it sends a RESV_TEAR message upstream to the source router.
RESV_CONF	When requested, a sender router responds to the destination router with a RESV_CONF message to indicate that a complete tunnel has been successfully established.

RSVP-TE Fast Reroute

RSVP-TE Fast Reroute allows to configure backup LSP tunnels to provide local repair/protection ONLY for **explicitly-routed** LSPs/LSP tunnels, termed *protected LSPs* as described in IETF DRAFT draft-ietf-mpls-rsvp-lsp-fastreroute-03.

An example diagram of a backup scenario for rerouting around a downed link, using an LSP Detour, is shown in Figure 3-11 on page 3-25. Figure 3-12 on page 3-25 shows an example diagram for a backup scenario to reroute around a downed node, using an LSP Detour.

Figure 3-11. RSVP-TE Fast Reroute Backup Link (Detour) Example

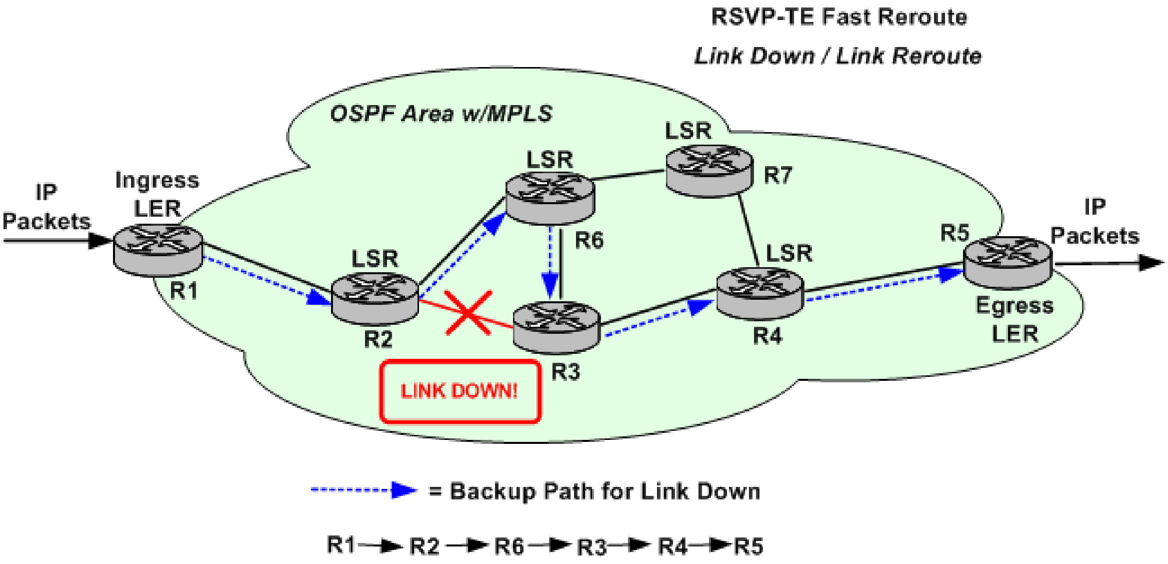
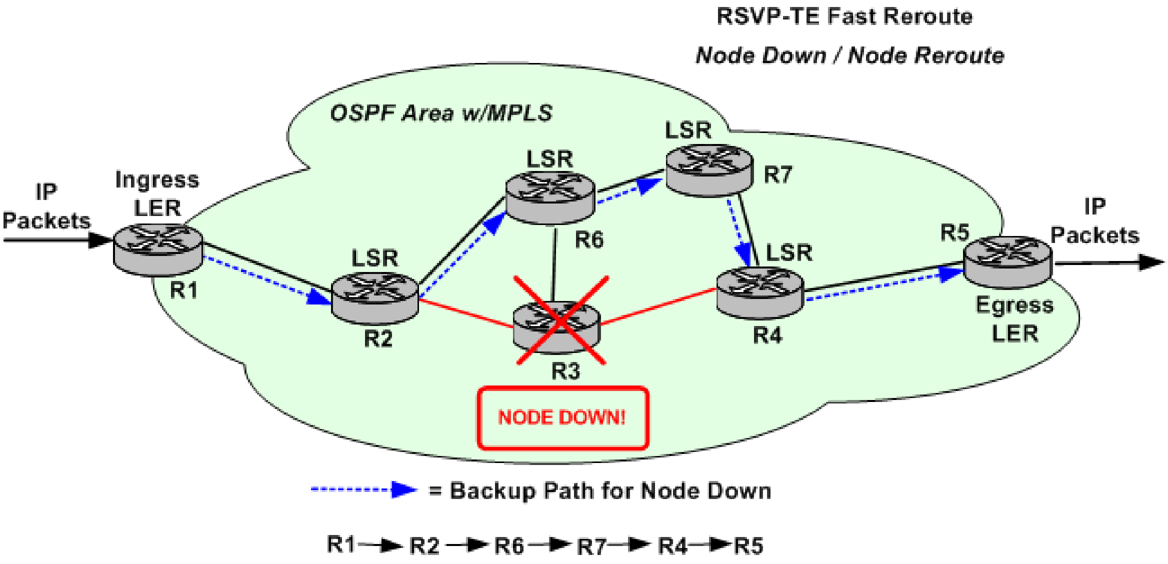


Figure 3-12. RSVP-TE Fast Reroute Backup Node (Detour) Example



The one-to-one backup method is based on including a DETOUR object in the Path message. The head-end router, the Point of Local Repair (PLR), sets up a separate detour LSP for each LSP it protects. For the Facility backup method, the PLR sets up a tunnel to protect multiple LSPs simultaneously, by using the MPLS label stack.

Ixia Test Model

The Ixia test process is designed so as to fully exercise RSVP functionality in MPLS routers. An Ixia port can simulate any number of LSR routers at the same time. Each router operates in an ingress or egress mode. In the following discussion, LSRs I and II refer to Figure 3-10 on page 3-20:

- Ingress mode: LSRs I and II are termed a neighbor pair, where LSR I is the upstream router being simulated and LSR II is its immediate downstream neighbor. The Ixia port generates the PATH and HELLO messages that LSR I would send. LSR II is the Device Under Test (DUT) and may be an egress router or be connected to other LSRs, as shown in the figure.
- Egress mode: the Ixia port simulates LSR II while LSR I is the DUT. The Ixia port interprets PATH messages that it receives to determine if they are directed for any of the defined destination routers. If that is the case, it responds with appropriate RESV messages.

If requested, HELLO messages are generated and responded to in either mode.

When the Ixia port operates in Ingress mode, it attempts to set up LSP tunnels for each combination of sender router and destination router, using any number of LSP tunnels and any number of LSP IDs for each LSP tunnel. Thus the number of PATH messages that the Ixia port attempts to generate for each refresh interval is:

# of sender routers
x # of destination routers
x # of LSPs
x # of LSP tunnels

The protocol server records all labels and other information that it receives on behalf of its simulated routers and displays those in a convenient format.

LDP

The Label Distribution Protocol (LDP) version 1, defined in RFC 3036, works in conjunction with Multi-Protocol Label Switching (MPLS), to efficiently ‘tunnel’ IP traffic across backbone topologies between Label Switching Routers (LSRs).

MPLS forwards packets based on added labels, so IP routing table lookups are not required along the length of the tunnel. RFC 3031 defines Forwarding Equivalence Classes (FECs) for use with MPLS, for purposes such as Quality of Service (QoS). LDP utilizes this option, assigning an FEC to every Label Switched Path (LSP) it sets up.

The LDP protocol creates peer sessions through a bidirectional exchange of messages, which include label requests and labels. While the initial Hello messages are based on UDP and sent to well-known port ‘646,’ all other messages are based on TCP.

The following global timers can be configured: Hello Hold timer, Hello Interval timer, KeepAlive Hold timer, and KeepAlive Interval timer. The values for these timers can be entered by you, but the final values are negotiated during the Discovery and Session setup processes. When the LDP remote peer has a timeout value which is lower than the one configured for the local LDP router, the lower value is used by both peers.

Virtual Circuit (VC) Ranges of MAC Addresses can be created to simulate Virtual Private LAN Services (VPLS), where L2 PDUs can be carried over VC LSPs, which, in turn, are carried over MPLS. This creates a ‘bridged,’ Ethernet Layer Two Virtual Private Network (Ethernet L2VPN). Refer to IETF DRAFT draft-lasserre-vkompella-ppvpn-vpls-03, which defines the VC Type - Ethernet VPLS, and also discusses the use of MPLS transport tunnels by pseudowires (PWs).

A pseudowire is a logical link through the tunnel, made up of two parallel VC LSPs using the same VC Identifier (VCID), as shown in Figure 3-13 on page 3-27, and in more detail in Figure 3-14 on page 3-28.

Figure 3-13. LDP VPLS Example

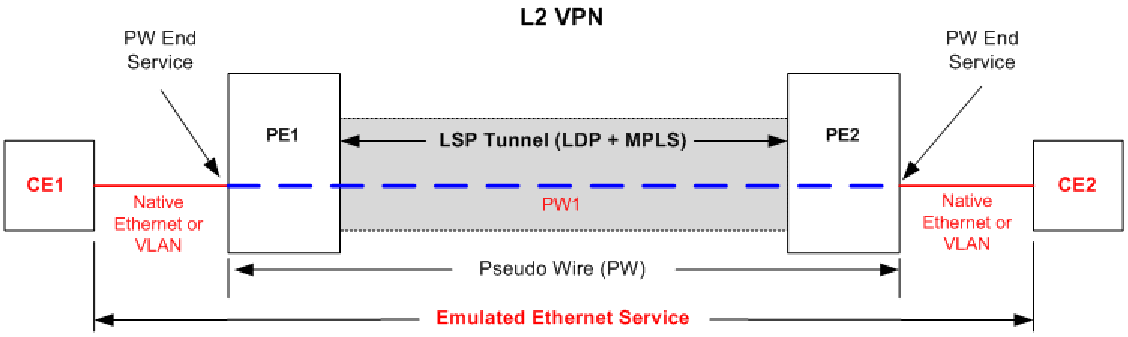
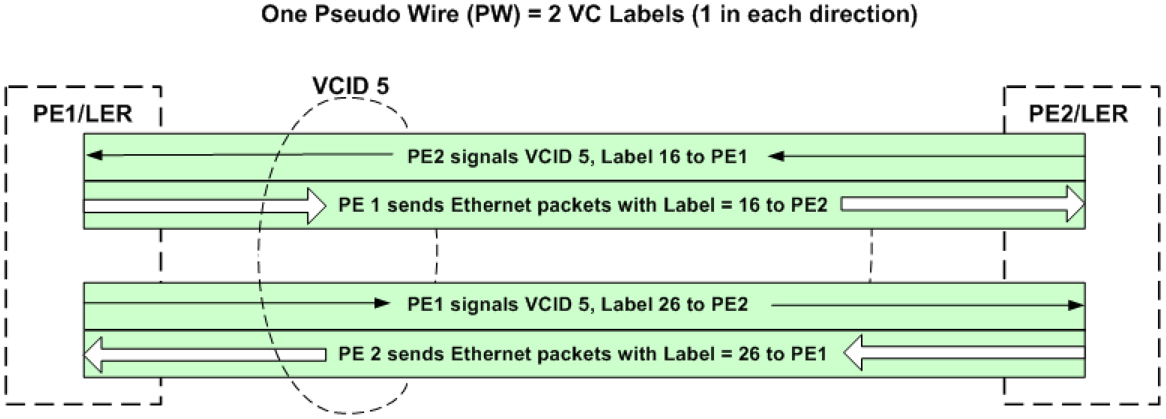


Figure 3-14. LDP VPLS Pseudowire Diagram


One Pseudo Wire (PW) = 2 VC Labels (1 in each direction)
VCID 5
PE1/LER
PE2/LER
PE2 signals VCID 5, Label 16 to PE1
PE 1 sends Ethernet packets with Label = 16 to PE2
PE1 signals VCID 5, Label 26 to PE2
PE 2 sends Ethernet packets with Label = 26 to PE1

MLD

The Multicast Listener Discovery (MLD) protocol is integral to the operation of Internet Protocol Version 6 (IPv6). MLDv1 is defined by RFC 2710, while MLDv2 is defined by RFC 3810. The MLD operations are based on operations similar to the Internet Group Management Protocol (IGMP) that supports IPv4. MLDv2 corresponds to IGMPv3. Both versions are supported by the protocol server.

An IPv6 router uses MLD to: (1) discover multicast listeners (nodes) on the directly attached links, and (2) find out which multicast addresses those nodes have interest in. In MLDv2, nodes can indicate interest in listening to packets that are sent to a specific multicast address from a filtered group of source IP addresses. This filtering can be based on ‘all but’ (Excluding) or ‘only’ (Including) certain source addresses. Host nodes can only be multicast ‘listeners,’ while the multicast routers can act as routers or listeners.

PIM-SM/SSM-v4/v6

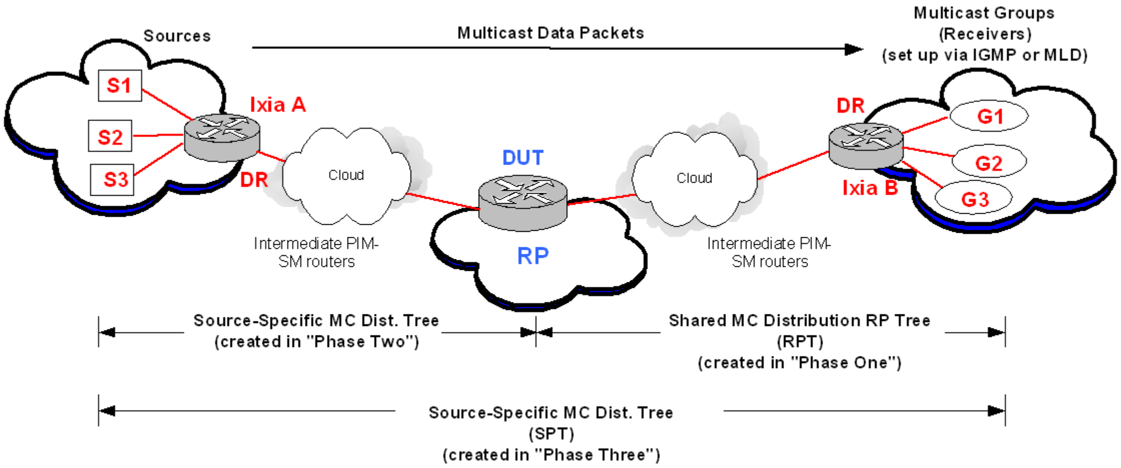
Protocol Independent Multicast - Sparse Mode (PIM-SM) Version 2 protocol is designed for multicast routing, and is defined in RFC 2362. IETF DRAFT draft-ietf-pim-sm-v2-new-06.txt is being designed to obsolete RFC 2362.

There is one Rendezvous Point (RP) per multicast group, and this router serves as the root of a unidirectional *shared* distribution tree whose 'leaves' consist of multicast receivers. In addition, PIM-SM can create an optional shortest-path tree for an *individual* source (where the source is the root). The term *upstream* is used to indicate the direction toward the root of the tree; *downstream* indicates the direction away from the root of the tree. The address of the RP can be configured statically by an administrator, or configured through a Bootstrap router (BSR) mechanism.

PIM-SM can use two sources of topology information to populate its routing table, the Multicast Routing Information Base (MRIB): unicast or multicast-capable. In a LAN where there are multiple PIM-SM routers and directly-connected hosts, one of the routers is elected as Designated Router (DR) to act on the behalf of the hosts.

The diagram in Figure 3-15 on page 3-29 shows a simplified PIM-SM test setup using Ixia ports.

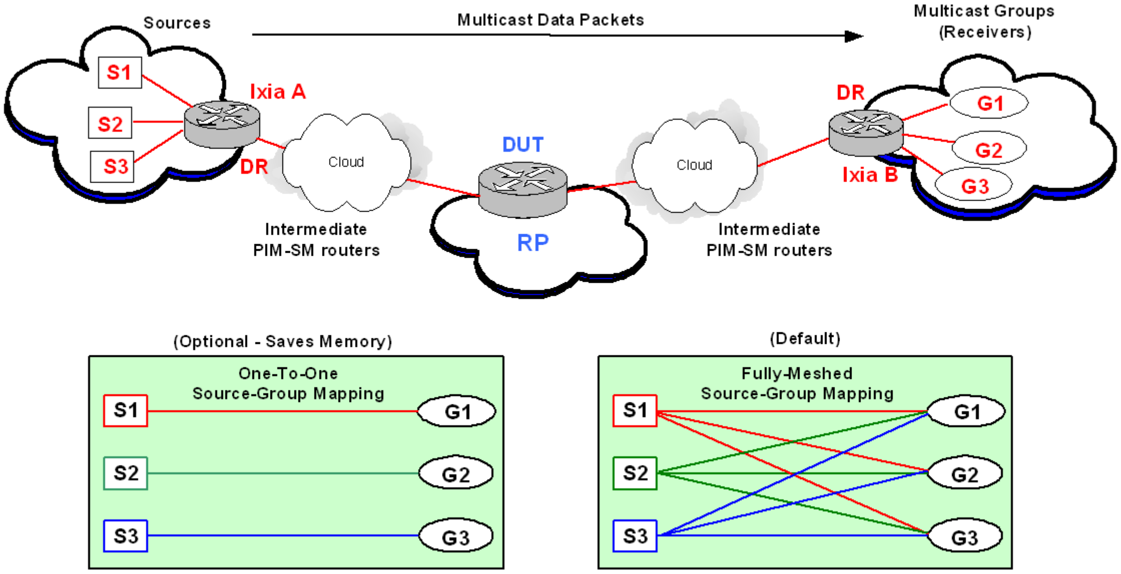
Figure 3-15. PIM-SM Diagram



PIM-SM Source-Group Mapping

PIM-SM Source-Group mapping involves the pairing of Sources and Groups. The default method is a fully-meshed mapping of sources to groups, where every source is paired with every group. For a situation where there are ‘X’ number of sources and ‘Y’ number of groups, there will be ‘X x Y’ number of mappings, resulting in a great deal of memory usage for processing. When full-mesh mapping is not desired, the optional ‘One-To-One’ Source-Group Mapping can be used to save memory. In comparison, if a one-to-one type mapping behavior was preferred and only a full-mesh setup was available, you would have to create ‘N’ fully-meshed source-group mapping ranges of size ‘1’ to emulate the one-to-one behavior. An example showing the differences between the two types of mapping is shown in Figure 3-16 on page 3-30.

Figure 3-16. PIM-SM Source-Group Mapping Example

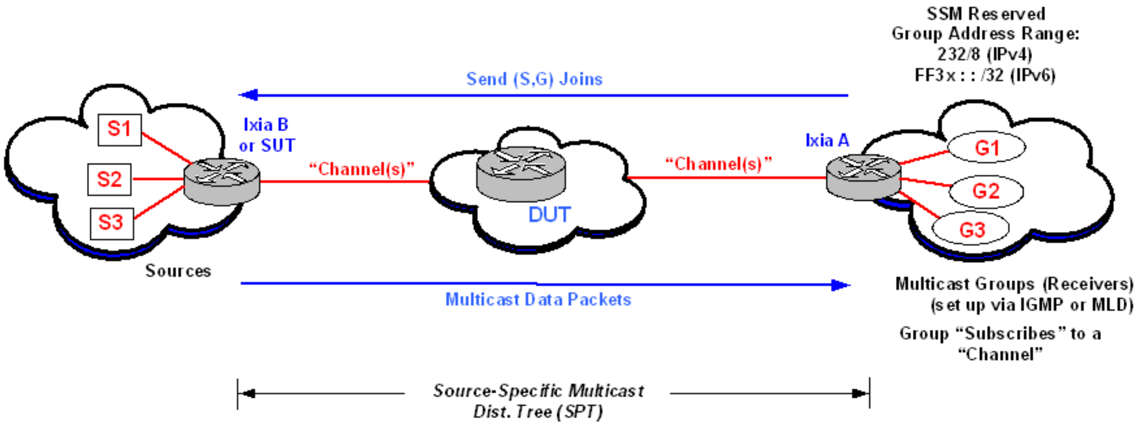


PIM-SSM

Protocol Independent Multicast - Source-Specific Multicast (PIM-SSM) uses a subset of the PIM-SM protocol, described in draft-ietf-ssm-arch-06, Source-Specific Multicast for IP, and in Section 4.8 of draft-ietf-pim-sm-v2-new-11, Protocol Independent Multicast - Sparse Mode (PIM-SM): Protocol Specification (Revised).

PIM-SSM is useful for broadcast-type applications, where one source sends packets to many host groups. There is no shared distribution tree topology, but there is a shortest-path tree (SPT) established, where the source is the root of the tree. In the case of SSM, the usual PIM-SM multicast terminology is modified, and the term *Channel* is used instead of *Group* and *Subscription* replaces *Join*. A multicast group (G) router that wants to receive packets from a specific Source (S) for its hosts/listeners, will 'Subscribe' to 'Channel (S,G).' An example of an PIM-SSM topology is shown in Figure 3-17 on page 3-31.

Figure 3-17. PIM-SSM Topology Example



An existing PIM-SM network can be modified to run SSM by enabling PIM-SSM on the source and destination/group routers. The typical PIM-SM signaling is not used for PIM-SSM, since the role of Rendezvous Point (RP) router is eliminated. The Subscribe (Join) message travels directly from the Destination router to the Source router, and data packets are transmitted in the opposite direction.

PIM-SSM Addressing

The PIM-SSM protocol uses a restricted addressing scheme, with reserved values for IPv4 SSM addresses defined by the IANA as 232.0.0.0 through 232.255.255.255 (232/8). IPv6 SSM addresses are defined in IETF DRAFT draft-ietf-ssm-arch- 06 and draft-ietf-pim-sm-v2-new-11 as FF3x: : /32. The range of FF3x: : /96 is proposed by RFC 3307, 'Allocation Guidelines for IPv6 Multicast Addresses.'

Differences Between PIM-SM and PIM-SSM

Some of the principal differences between PIM-SM and PIM-SSM routers, per draft-ietf-ssm-arch-06, are mentioned in the following list:

- PIM-SSM-only routers must not send (*,G) Join/Prune messages.
- PIM-SSM-only routers must not send (S,G,rpt) Join/Prune messages.
- PIM-SSM-only routers must not send Register messages for packets with SSM destination addresses.
- PIM-SSM-only routers must act in accordance with (*,G) or (S,G,rpt) state by forwarding packets with SSM destination addresses.
- PIM-SSM-only routers acting as RPs must not forward Register messages for packets with SSM destination addresses.

Protocol Elements for PIM-SSM

Protocol elements ***required*** for PIM-SSM-only routers are mentioned in the following list:

- (S,G) Downstream and Upstream state machines.
- Hello messages, neighbor discovery, and DR election.
- Packet forwarding rules.
- [(S,G) Assert state machine]

Some of the Protocol elements ***not required*** for PIM-SSM-only routers are mentioned in the following list:

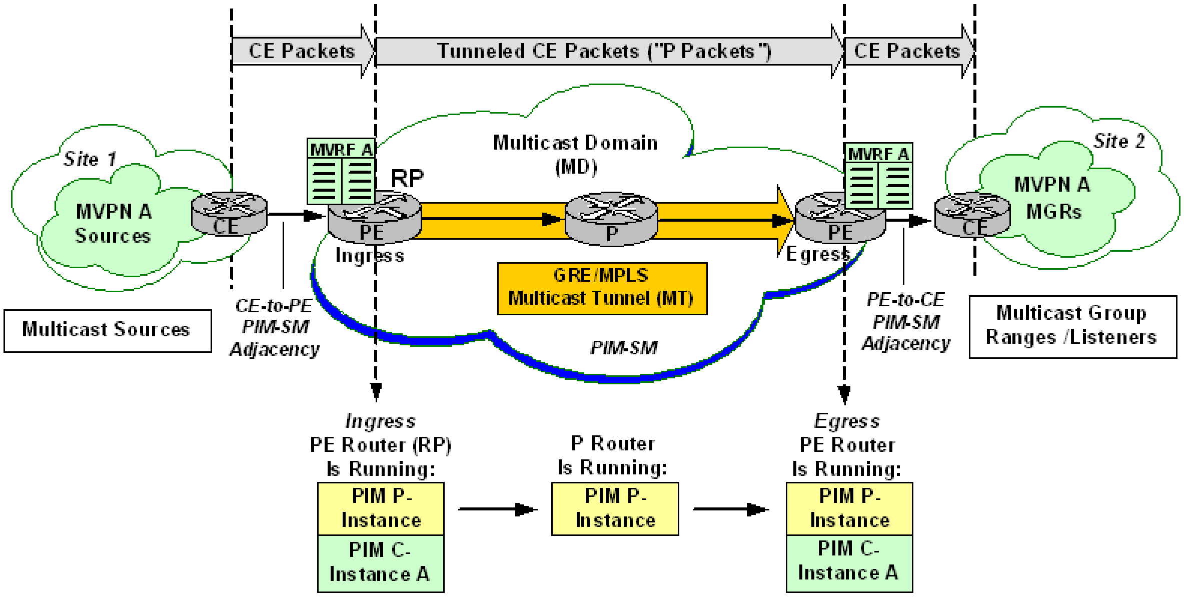
- Register state machine
- (*,G), (S,G,rpt), and (*,*,RP) Downstream and Upstream state machines.
- Keepalive Timer (treated as always running)
- SptBit (treated as always set for an SSM address)

Multicast VPNs

Multicast VPNs (MVPNs) can be created through the use of MP-BGP combined with PIM-SM. Multicast VPNs can be set up by a Service Provider to support scalable, IPv4 multicast traffic solutions, based on IETF draft-rosen-vpn-mcast-07, ‘Multicast in MPLS/BGP IP VPNs.’

Multicast VRFs (MVRFs) on each PE router contain multicast routing tables. Within a Service Provider’s domain, each MVRF is assigned to a Multicast Domain (MD), which is a set of MVRFs that can send multicast traffic to one another. Multicast packets from CE routers are sent over a (GRE) multicast tunnel to other PE routers in the multicast domain. A simplified example of a Multicast VPN topology, with one MVPN, is shown in Figure 3-18 on page 3-33.

Figure 3-18. Multicast VPN Topology Example



Each CE and its connected PE set up a PIM-SM adjacency. However, CEs do not set up PIM-SM adjacencies with each other. Separate CE-associated instances of PIM are run by each PE router, and these are called ‘PIM C-instances.’ Each C-instance is MVRF-specific. As each PE can be affiliated with many MVPNs/ MVRFs, the router can run many PIM C-instances simultaneously, up to a maximum of one C-instance per MVRF.

PIM Provider-wide instances (‘PIM P-instances’) are run by each PE router, creating a global PIM adjacency with all of its IGP PIM-SM-enabled neighbors (P routers). P routers cannot set up PIM-SM C-instances.

At startup for the multicast domain’s Provider Edge (PE) routers, the default Multicast Distribution Tree (MDT) is set up automatically. Each Multicast Domain is identified by a globally unique Service Provider (P) Group address and a Route Distinguisher. The MD group address is created by using BGP (L3 Site window). It is a valid 4-byte IPv4 multicast address prefix (for example, 239.1.1.1/32). The 12-byte Route Distinguisher is also created through BGP (L3 Site window). This Ixia implementation uses an RD value = 2. One C-Multicast Group Range (MGR) can be configured for each MVRF.

MPLS

Multi-Protocol Label Switching (MPLS) is based on the concept of label switching: and independent and unique 'label' is added to each data packet and this label is used to switch and route the packet through the network. The label is simple, essentially a shorthand version of the packet's header information, so network equipment can be optimized around processing the label and forwarding traffic. This concept has been around the data communications industry for years. X.25, Frame Relay, and ATM are examples of label switching technologies.

It is important to understand the differences in the way MPLS and IP routing forward data across a network. Traditional IP packet forwarding uses the IP destination address in the packet's header to make an independent forwarding decision at each router in the network. These hop-by-hop decisions are based on network layer routing protocols, such as Open Shortest Path First (OPSF) or Border Gateway Protocol (BGP). These routing protocols are designed to find the shortest path through the network, and do not consider other factors, such as latency or traffic congestion.

MPLS creates a connection-based model overlaid onto the traditionally connectionless framework of IP routed networks. This connection-oriented architecture opens the door to a wealth of new possibilities for managing traffic on an IP network. MPLS builds on IP, combining the intelligence of routing, which is fundamental to the operation of the Internet and today's IP networks, with the high performance of switching. Beyond its applicability to IP networking, MPLS is being expanded for more general applications in the form of Generalized MPLS (GMPLS), with applications in optical and Time-Division Multiplexing (TDM) networks.

One of the primary original goals of MPLS, boosting the performance of software-based IP routers, has been superseded as advances in silicon technology have enabled line-rate routing performance implemented in router hardware. In the meantime, additional benefits of MPLS have been realized, notably Virtual Private Network (VPN) services and traffic engineering (TE).

Advantages of MPLS

Some of the advantages of using MPLS are:

- MPLS enables traffic engineering. Explicit traffic routing and engineering help squeeze more data into available bandwidth.
- MPLS supports the delivery of services with Quality of Service (QoS) guarantees. Packets can be marked for high quality, enabling providers to maintain a specified low end-to-end latency for voice and video.
- MPLS reduces router processing requirements, since routers simply forward packets based on fixed labels.

How Does MPLS Work?

MPLS is a technology used for optimizing forwarding through a network. Though MPLS can be applied in many different network environments, this discussion focuses primarily on MPLS in IP packet networks, by far the most common application of MPLS today.

MPLS assigns labels to packets for transport across a network. The labels are contained in an MPLS header inserted into the data packet.

These short, fixed-length labels carry the information that tells each switching node (router) how to process and forward the packets, from source to destination. They have significance only on a local node-to-node connection. As each node forwards the packet, it swaps the current label for the appropriate label to route the packet to the next node. This mechanism enables very-high-speed switching of the packets through the core MPLS network.

MPLS combines the best of both Layer 3 IP routing and Layer 2 switching. In fact, it is sometimes called a ‘Layer 2-1/2’ protocol. While routers require network-level intelligence to determine where to send traffic, switches only send data to the next hop, and so are inherently simpler, faster, and less costly. MPLS relies on traditional IP routing protocols to advertise and establish the network topology. MPLS is then overlaid on top of this topology. MPLS predetermines the path data takes across a network and encodes that information into a label that the network’s routers can understand. This is the connection-oriented approach previously discussed. Since route planning occurs ahead of time and at the edge of the network (where the customer and service provider network meet), MPLS-labeled data requires less router horsepower to traverse the core of the service provider’s network.

MPLS Routing

MPLS networks establish Label-Switched Paths (LSPs) for data crossing the network. An LSP is defined by a sequence of labels assigned to nodes on the packet’s path from source to destination. LSPs direct packets in one of two ways: hop-by-hop routing or explicit routing.

Hop-by-Hop Routing

In hop-by-hop routing, each MPLS router independently selects the next hop for a given Forwarding Equivalency Class (FEC). A FEC describes a group of packets of the same type; all packets assigned to a FEC receive the same routing treatment. FECs can be based on an IP address route or the service requirements for a packet, such as low latency.

In the case of hop-by-hop routing, MPLS uses the network topology information distributed by traditional Interior Gateway Protocols (IGPs) routing protocols such as OPSF or IS-IS. This process is similar to traditional routing in IP networks, and the LSPs follow the routes the IGPs dictate.

Explicit Routing

In explicit routing, the entire list of nodes traversed by the LSP is specified in advance. The path specified could be optimal or not, but is based on the overall view of the network topology and, potentially, on additional constraints. This is called Constraint-Based Routing. Along the path, resources may be reserved to ensure QoS. This permits traffic engineering to be deployed in the network to optimize use of bandwidth.

Label Information Base

As the network is established and signaled, each MPLS router builds a Label Information Base (LIB), a table that specifies how to forward a packet. This table associates each label with its corresponding FEC and the outbound port to forward the packet to. This LIB is typically established in addition to the routing table and Forwarding Information Base (FIC) that traditional routers maintain.

Connections are signaled and labels are distributed among nodes in an MPLS network using one of several signaling protocols, including Label Distribution Protocol (LDP) and Resource reSerVation Protocol with Tunneling Extensions (RSVPTE). Alternatively, label assignment can be piggybacked onto existing IP routing protocols such as BGP.

The most commonly used MPLS signaling protocol is LDP. LDP defines a set of procedures used by MPLS routers to exchange label and stream mapping information. It is used to establish LSPs, mapping routing information directly to Layer 2 switched paths. It is also commonly used to signal at the edge of the MPLS network, the critical point where non-MPLS traffic enters. Such signaling is required when establishing MPLS VPNs.

RSVP-TE is also used for label distribution, most commonly in the core of networks that require traffic engineering and QoS. A set of extensions to the original RSVP protocol, RSVP-TE provides additional functionality beyond label distribution, such as explicit LSP routing, dynamic rerouting around network failures, preemption of LSPs, and loop detection. RSVP-TE can distribute traffic engineering parameters such as bandwidth reservations and QoS requirements.

Multi-protocol extensions have been defined for BGP, enabling the protocol to also be used to distribute MPLS labels. MPLS labels are piggybacked onto the same BGP messages used to distribute the associated routes. MPLS allows multiple labels (called a label stack) to be carried on a packet. Label stacking enables MPLS nodes to differentiate between types of data flows, and to set up and distribute LSPs accordingly. A common use of label stacking is for establishing tunnels through MPLS networks for VPN applications.

BFD

Bidirectional Forwarding Detection (BFD) is a network protocol used to detect faults between two forwarding engines. It provides low-overhead detection of faults even on physical media that don't support failure detection of any kind, such as ethernet, virtual circuits, tunnels and MPLS LSPs.

BFD establishes a session between two endpoints over a particular link. If more than one link exists between two systems, multiple BFD sessions may be established to monitor each one of them. The session is established with a three-way handshake, and is torn down the same way. Authentication may be enabled on the session. A choice of simple password, MD5 or SHA1 authentication is available.

BFD does not have a discovery mechanism; sessions must be explicitly configured between endpoints. BFD may be used on many different underlying transport mechanisms and layers, and operates independently of all of these. Therefore, it needs to be encapsulated by whatever transport it uses. For example, monitoring MPLS LSPs involves piggybacking session establishment on LSP-Ping packets. Protocols that support some form of adjacency setup, such as OSPF or IS-IS, may also be used to bootstrap a BFD session. These protocols may then use BFD to receive faster notification of failing links than would normally be possible using the protocol's own keepalive mechanism.

A session may operate in one of two modes: asynchronous mode and demand mode. In asynchronous mode, both endpoints periodically send Hello packets to each other. If a number of those packets are not received, the session is considered down.

In demand mode, no Hello packets are exchanged after the session is established; it is assumed that the endpoints have another way to verify connectivity to each other, perhaps on the underlying physical layer. However, either host may still send Hello packets if needed.

Regardless of which mode is in use, either endpoint may also initiate an Echo function. When this function is active, a stream of Echo packets is sent, and the other endpoint then sends these back to the sender through its forwarding plane. This is used to test the forwarding path on the remote system.

CFM

Ethernet CFM is an end-to-end per-service-instance Ethernet layer OAM protocol that includes proactive connectivity monitoring, fault verification, and fault isolation. End to end can be PE to PE or customer edge (CE) to CE. Per service instance means per VLAN.

Being an end-to-end technology is the distinction between CFM and other metro-Ethernet OAM protocols. For example, MPLS, ATM, and SONET OAM help in debugging Ethernet wires but are not always end-to-end. 802.3ah OAM is a single-hop and per-physical-wire protocol. It is not end to end or service aware. Ethernet Local Management Interface (E-LMI) is confined between the uPE and CE and relies on CFM for reporting status of the metro-Ethernet network to the CE.

Troubleshooting carrier networks offering Ethernet Layer 2 services can be difficult. Customers contract with service providers for end-to-end Ethernet service and service providers may subcontract with operators to provide equipment and networks. Compared to enterprise networks, where Ethernet traditionally has been implemented, these constituent networks belong to distinct organizations or departments, are substantially larger and more complex, and have a wider user base. Ethernet CFM provides a competitive advantage to service providers for which the operational management of link uptime and timeliness in isolating and responding to failures is crucial to daily operations.

FCoE and NPIV

IxExplorer provides GUI access to all Ixia platform functionality with full support for stateless FCoE functional and scalability testing. The FCoE and Priority Flow Control (PFC) and FCoE Initialization Protocol (FIP) features allow testing of FCoE switches running both FCoE traffic and traditional Ethernet traffic.

Supported Load Modules

The following Ixia load modules have the Fibre Channel over Ethernet (FCoE) capability:

- LSM10GXM8-01, GXMR8-01, and GXM8XP-01, including 10GBASE-T versions LSM10GXM(R)8GBT-01
- LSM10GXM4-01, GXMR4-01, and GXM4XP-01, including 10GBASE-T versions LSM10GXM(R)4GBT-01
- LSM10GXM2XP-01 and GXMR2-01, including 10GBASE-T versions LSM10GXM(R)2GBT-01
- LSM1000XMVDCx-01 load modules. 4-port, 8-port, 12-port, and 16-port
- LSM1000XMVDC4-NG load modules. 4-port

Data Center Mode

FCoE support requires a new port mode, Data Center Mode. You need to switch port mode between Normal Mode and Data Center Mode to use the desired features in each mode.

- Mode switching (to or from Data Center Mode) triggers an FPGA re-download.
- There is no Packet Stream Mode support in Data Center Mode; only Advanced Scheduler Mode is supported.
- Supports 4-Priority traffic mapping for frame size up to 9216-byte. The different frame size support is determined by a sub mode in Data Center Mode. This limitation applies to all frames in Data Center Mode, whether FCoE frame or not.
- Data Center Mode only supports auto instrumentation mode for both TX and RX.
- When the port is in Data Center Mode, both existing Ethernet frames and FCoE frames are generated.

Priority Traffic Generation

The scheduling function is based on the existing Advanced Scheduler. A new parameter called ‘Priority Group’ has been added to each stream. You can map Priority Group to the priority field in the frame. The priority field in the same stream should not change (for example, if the priority is a VLAN priority field, then you cannot configure a UDF to control this field within a stream).

Priority-based Flow Control (PFC)

The Ixia port responds to either IEEE 802.3x pause frame or to IEEE 802.1Qbb Priority-based Flow Control (PFC) frame. The flow control type is determined by the selection made on the Flow Control tab of the Port Properties dialog, in IxExplorer.

IxExplorer Reference

See IxExplorer User Guide, Chapter 6 topic *Frame Data for FCoE Support*, subtopic *Priority-based Flow Control*.

Fibre Channel over Ethernet

When the port is in Data Center Mode, both existing Ethernet frames and FCoE frames are generated.

The Fibre Channel CRC is generated on the fly. This CRC is inserted at offset of Ethernet frame size minus 12 bytes. For example:

Ethernet Frame Size (bytes)	2000	2001	2002	2003
FC-CRC Offset in FCoE Frame (bytes)	1988	1989	1990	1991

The FC-CRC can be set to No Error or to Bad CRC.

Packet View Support

For Fibre Channel frame, there is no Extended Header and Optional Header support. It decodes only FC-2 Frame Header field.

FCoE Initialization Protocol (FIP)

FIP (FCoE Initialization Protocol) has been implemented (in addition to FCoE). It is used to discover and initialize FCoE capable entities connected to an Ethernet cloud.

IxExplorer Reference

See the IxExplorer User Guide, Chapter 6 topic *Frame Data for FCoE Support*.

NPIV Protocol Interface

NPIV stands for N_Port ID Virtualization. These can be used to virtually share a single physical N_Port. This allows multiple Fibre Channel initiators to occupy a single physical port, easing hardware requirements in SAN design. Up to 256 N_Port_IDs can be assigned to a single N_Port. NPIV interfaces can be configured using the Protocol Interface Wizard.

See the IxExplorer User Guide, Chapter 10, topic *NPIV Protocol Interface*.

Precision Time Protocol (PTP) IEEE 1588v2

Precision Time Protocol (IEEE 1588v2) allows precise synchronization of clocks in measurement and controls systems implemented with technologies such as network communications, local computing and distributed objects. The protocol supports system wide synchronization accuracy in sub-microseconds range with minimal network and local clock computing resources. The protocol operates in master/subordinate configuration. IEEE1588 deploys Multicast over an Ethernet network, and devices such as routers and switches can sync to the provided timing source.

Supported Load Modules

The following Ixia load modules have the PTP capability:

- LSM1000XMV(R)16, XMV(R)12, XMV(R)8, XMV(R)4
- ASM1000XMV12X
- Xcellon-Ultra XP, NP, and NG

Supported Messages

The following messages are supported between clocks participating in the PTP protocol.

- Event messages
  - Sync
  - Delay Request
- General Messages
  - Announce
  - Follow_up
  - Delay_Response

Supported Features

The following PTP features are supported.

- Only one two-step clock is supported on an Ixia port, at this time. One-step clock is not supported.
- Ixia ports can run other (non-PTP) traffic along with PTP traffic. Ixia ports have the ability to throttle transmit based on flow control packets being received.
- IEEE 1588 version 2.2 in IPV4 (multicast) is supported.
- Ports are manually configured in Master or Subordinate mode.
- A histogram reporting Subordinate clock OFFSET from master is provided in the form of plot along with PTP messages transmitted and received.
- Aggregate statistics are displayed in Statistics View in IxExplorer.
- Session/Per Interface stats is displayed in IxNw/Tcl/csv file.

- Per Interface configuration is done in protocol interfaces.
- Ability to compose or decode PTP messages from the IxExplorer user interface.
- Negative testing is supported.
    - Programmable follow-up messages as a percentage of sync messages. See how dropping 10-90% of follow-up messages (while sending 100% of sync messages) affects the DUT.
    - Send follow-up messages with a bad packet.
    - Purposely send data with timestamps that include jitter (to try forcing a sync to fail).
    - Negative testing is done with packet streams/linux.

Local Clock synchronization through PTP to another PTP clock

The local clock of the is synchronized to the ’s master clock by minimizing the Offset_from_master value of the current data set. The time and the rate characteristics of the local clock are modified upon receipt of either a sync message or follow-up message. Figure 3-19 illustrates the PTP communication path.

Figure 3-19. PTP Communication Path

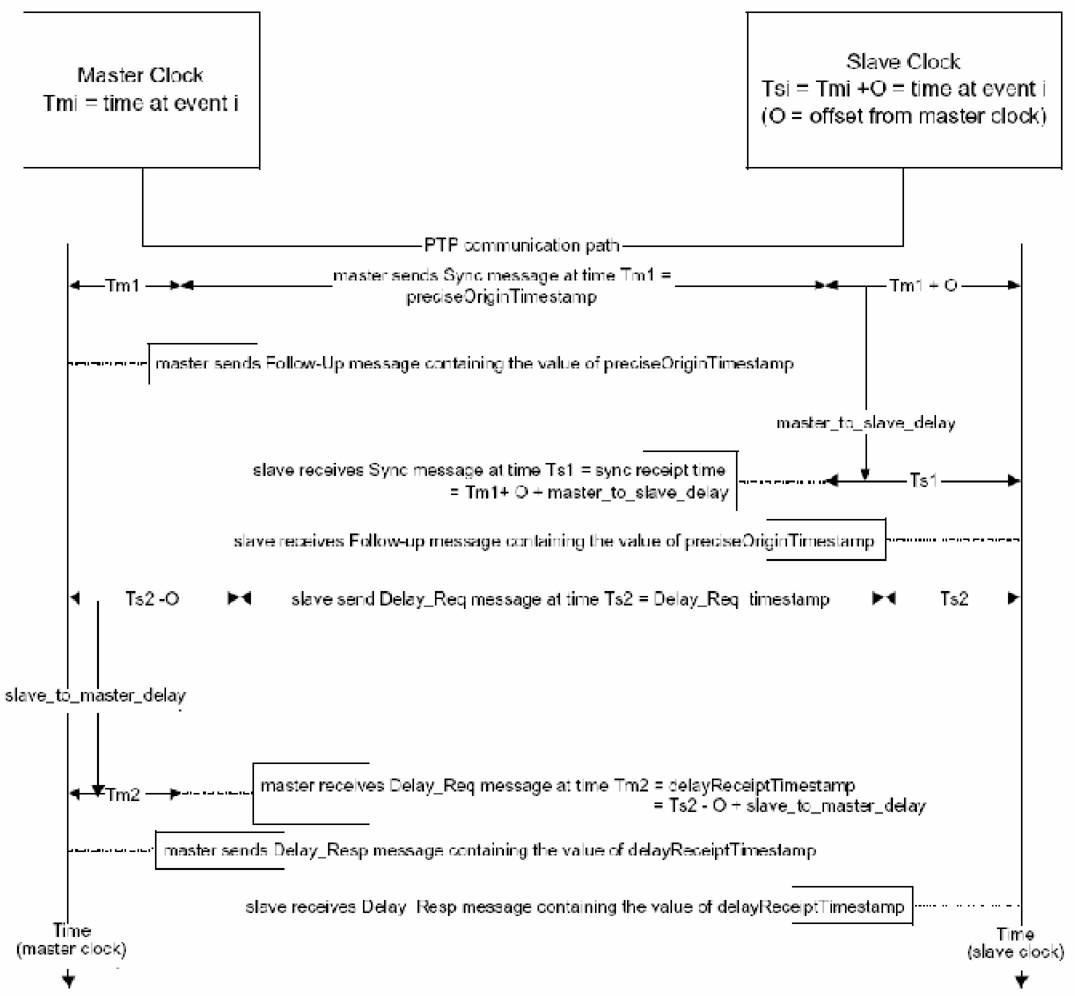

Master Clock
Tmi = time at event i
Slave Clock
Tsi = Tmi +O = time at event i
(O = offset from master clock)
PTP communication path
Tm1
master sends Sync message at time Tm1 = preciseOriginTimestamp
Tm1 + O
master sends Follow-Up message containing the value of preciseOriginTimestamp
master_to_slave_delay
slave receives Sync message at time Ts1 = sync receipt time = Tm1+ O + master_to_slave_delay
Ts1
slave receives Follow-up message containing the value of preciseOriginTimestamp
Ts2 -O
slave send Delay_Req message at time Ts2 = Delay_Req timestamp
Ts2
slave_to_master_delay
Tm2
master receives Delay_Req message at time Tm2 = delayReceiptTimestamp = Ts2 - O + slave_to_master_delay
master sends Delay_Resp message containing the value of delayReceiptTimestamp
slave receives Delay Resp message containing the value of delayReceiptTimestamp
Time (master clock)
Time (slave clock)

Table 3-7. PTP Communication Path

Term	Value
sync_receipt_time = Ts1	Ts1 =Tm1+O+master_to__delay
preciseOriginTimestamp = Tm1	Tm1
master_to__delay (computed)	Ts1–Tm1
delay_req_sending_time = Ts2	Ts2
delayReceiptTimestamp = Tm2	Tm2 = Ts2 – O + _to_master_delay
_to_master_delay (computed)	Tm2 – Ts2

Table 3-7. PTP Communication Path

Term	Value
one_way_delay	{(master_to__delay as computed) + (_to_master_delay as computed)}/2 {(Ts1-Tm1) + (Tm2-Ts2)}/2 {(O + master_to__delay) + –O +_to_master_delay)}/2 {(master_to__delay) + (_to_master_delay)}/2 master_to__delay if path is symmetrical

Notes:

1. Offset shall be computed as O= Ts1-Tm1 - one_way_Delay. Offset and One way delay shall be stored.
2. Offset correction shall be applied to the local clock.

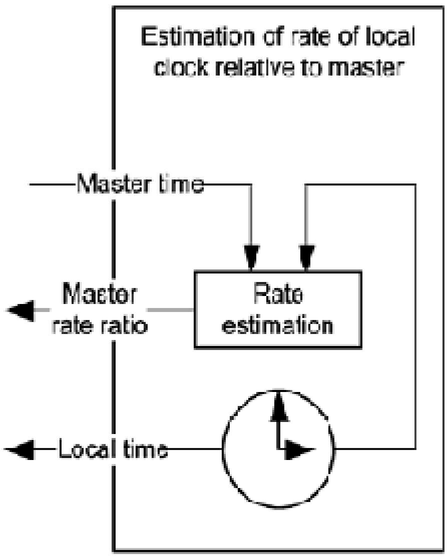
Local clock frequency transfer

In Slave mode of operation, the Ixia port implements a local clock in software (Linux). The frequency of the oscillator is not adjusted but allowed to free-run. The local clock shall be implemented based on time information synchronized from sync/follow_up messages and hardware timestamps associated with these messages. The local clock is associated with a constant and a slope. The rate of a local clock relative to a master clock is illustrated in Figure 3-20.

IxExplorer References

See the IxExplorer User Guide, Chapter 10, Protocol Interfaces, especially topics *Protocol Interfaces Tab*, *PTP Discovered Information*, and *PTP Clock Configuration*.

Figure 3-20. Rate of Local Clock Relative to Master



- Clock = K+ Slope*(TS-TS1),

- Slope = (T2-T1)/(TS2-TS1).
- K = T1

Where T1 is the time synchronized from the master and TS1 is the hardware timestamp associated with sync message 1. T2 and TS2 are corresponding parameters associated with sync message 2. T is the time at any point of time.

With a sync message, the parameters K and the slope are updated. The Clock Offset from master is calculated as discussed above and applied to K for correction.

Timestamps are cleared once when PTP is enabled.

In master mode of operation, server provides timestamp to the ports at the instant timestamps are cleared and slope is 1. OFFSET from master is 0.

If a GPS source is interfaced to the chassis, ports emulating the master are configured as Grand Master.

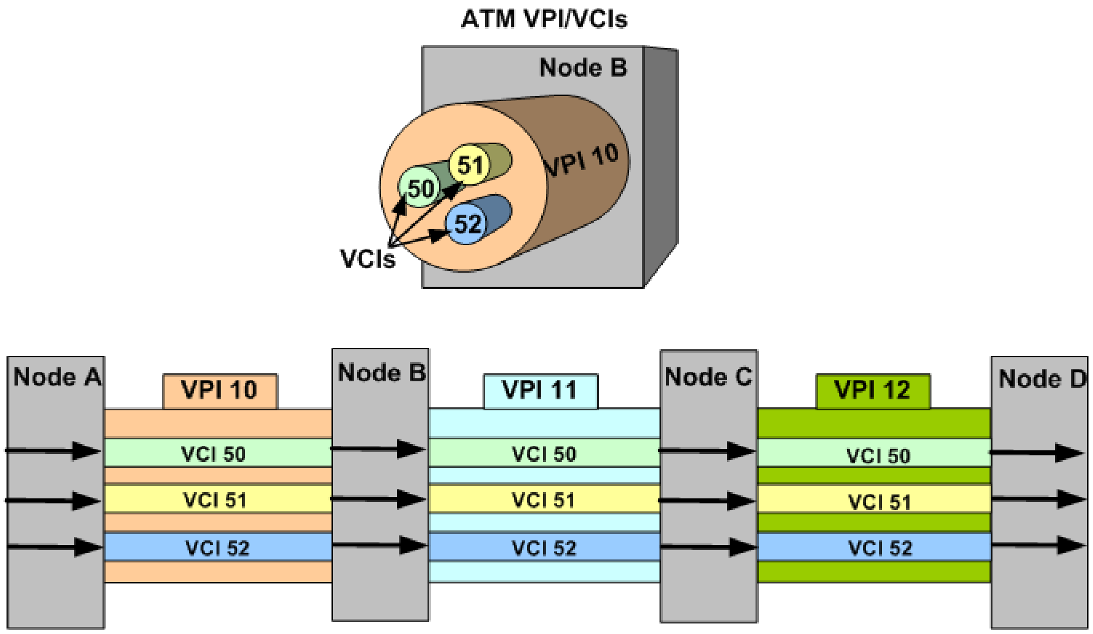
Local clock time format is seconds (32 bits) and nanoseconds (32 bits). The Ixia port supports a 2-step clock.

ATM Interfaces

On Asynchronous Transport Mode (ATM) is a Layer 2, connection-oriented, switching protocol, based on L2 Virtual Circuits (VCs). For operation in a connectionless IP routing or bridging environment, the IP PDUs must be encapsulated within the **payload field** of an ATM AAL5 CPCS-PDU (ATM Adaptation Layer 5—Common Part Convergence Sublayer—Protocol Data Unit). The ATM CPCS-PDUs are divided into 48-byte segments which receive 5-byte headers to form 53-byte ATM cells.

The ATM cells are then switched across the ATM network, based on the Virtual Port Identifiers (VPIs) and the Virtual Connection Identifiers (VCIs). The relationship between VPIs (identifying one hop between adjacent nodes) and VCIs (identifying the end-to-end virtual connection) is illustrated in Figure 3-21 on page 3-46.

Figure 3-21. ATM VPI/VCI Pairs (PVCs)



‘Bridged ATM’ Versus ‘Routed ATM’

The ATM AAL5 frames allow for the overlay of the connectionless IP bridging or routing environment over the network of ATM nodes (that have frame handling capability). Each ATM node examines the payload of the AAL5 frame, and forwards the frame to the next node, based on the payload’s MAC destination address (for IP bridging) or IP destination address (for IP routing). In effect, the ATM environment functions as a simulated Ethernet or IP network, respectively.

In the case of Label Distribution Protocol (LDP) routing over ATM, the process becomes more complex since MPLS tunnels are created over ATM core networks. For more information on the signaling, session setup, and label distribution for LDP routing over ATM, see the *IxNetwork Users Guide: Network Protocols - LDP chapter*.

ATM Encapsulation Types

There are two main types of ATM Multiplexing encapsulations defined by RFC 2684, 'Multiprotocol Encapsulation over ATM Adaptation Layer 5.' The ATM AAL5 Frame is described in *ATM Frame Formats* on page 3-50. The various encapsulation types and references to diagrams of the encapsulated frame payloads are listed as follows:

- **VC Multiplexing (VC Mux)**: used when only one protocol is to be carried on a single ATM VC. Separate VCs are used if multiple protocols are being transported.
  - VC Mux IPv4 Routed: see Figure 3-27 on page 3-52
  - VC Mux IPv6 Routed: see Figure 3-28 on page 3-52
  - VC Mux Bridged Ethernet/802.3 (FCS): see Figure 3-23 on page 3-51
  - VC Mux Bridged Ethernet/802.3 (no FCS): see Figure 3-24 on page 3-51
- **Logical Link Control (LLC)**: used for multiplexing multiple protocols over a single ATM virtual connection (VC).
  - LLC Routed AAL 5 Snap: see Figure 3-29 on page 3-53
  - LLC Bridged Ethernet (FCS): see Figure 3-25 on page 3-51
  - LLC Bridged Ethernet (no FCS): see Figure 3-26 on page 3-52

> **Note**: The Protocol Configuration Wizards for BGP, OSPFv2, and ISIS allow configuration on ATM ports, but ***ONLY*** for the VC Mux Bridged Ethernet/802.3 (FCS) encapsulation type.

Encapsulation Types by Protocol

The types of RFC 2684 ATM encapsulations available for each Ixia routing protocol emulation are listed in Table 3-8 on page 3-47.

Table 3-8. ATM Encapsulations for Protocols

Routing Protocol	ATM Encapsulation Type
BGP	***'Bridged ATM':*** • VC Mux Bridged Ethernet/802.3 (FCS) - (the default) • VC Mux Bridged Ethernet/802.3 (no FCS) • LLC Bridged Ethernet (FCS) • LLC Bridged Ethernet (no FCS) ***'Routed ATM':*** • VC Mux IPv4 Routed • VC Mux IPv6 Routed • LLC Routed AAL5 Snap

Table 3-8. ATM Encapsulations for Protocols

Routing Protocol	ATM Encapsulation Type
OSPF (v2 only) **Note**: Supported for both Point-to-Point and Point-to-MultiPoint links.	***'Bridged ATM':*** • VC Mux Bridged Ethernet/802.3 (FCS) - (the default) • VC Mux Bridged Ethernet/802.3 (no FCS) • LLC Bridged Ethernet (FCS) • LLC Bridged Ethernet (no FCS) ***'Routed ATM':*** • VC Mux IPv4 Routed • LLC Routed AAL5 Snap
LDP **Note**: Discovery Mode must be set to Basic, and Advertising Mode must be set to Downstream on Demand (DoD).	***'Bridged ATM':*** • VC Mux Bridged Ethernet/802.3 (FCS) - (the default) • VC Mux Bridged Ethernet/802.3 (no FCS) • LLC Bridged Ethernet (FCS) • LLC Bridged Ethernet (no FCS) ***'Routed ATM':*** • VC Mux IPv4 Routed • LLC Routed AAL5 Snap
RSVP-TE	***'Bridged ATM':*** • VC Mux Bridged Ethernet/802.3 (FCS) - (the default) • VC Mux Bridged Ethernet/802.3 (no FCS) • LLC Bridged Ethernet (FCS) • LLC Bridged Ethernet (no FCS)
ISIS	***'Bridged ATM':*** • VC Mux Bridged Ethernet/802.3 (FCS) - (the default) • VC Mux Bridged Ethernet/802.3 (no FCS) • LLC Bridged Ethernet (FCS) • LLC Bridged Ethernet (no FCS)
RIP	***'Bridged ATM':*** • VC Mux Bridged Ethernet/802.3 (FCS) - (the default) • VC Mux Bridged Ethernet/802.3 (no FCS) • LLC Bridged Ethernet (FCS) • LLC Bridged Ethernet (no FCS)
RIPng	***'Bridged ATM':*** • VC Mux Bridged Ethernet/802.3 (FCS) - (the default) • VC Mux Bridged Ethernet/802.3 (no FCS) • LLC Bridged Ethernet (FCS) • LLC Bridged Ethernet (no FCS)
IGMP	***'Bridged ATM':*** • VC Mux Bridged Ethernet/802.3 (FCS) - (the default) • VC Mux Bridged Ethernet/802.3 (no FCS) • LLC Bridged Ethernet (FCS) • LLC Bridged Ethernet (no FCS)

Table 3-8. ATM Encapsulations for Protocols

Routing Protocol	ATM Encapsulation Type
MLD	***'Bridged ATM':*** • VC Mux Bridged Ethernet/802.3 (FCS)-(the default) • VC Mux Bridged Ethernet/802.3 (no FCS) • LLC Bridged Ethernet (FCS) • LLC Bridged Ethernet (no FCS)
PIM-SM	***'Bridged ATM':*** • VC Mux Bridged Ethernet/802.3 (FCS)-(the default) • VC Mux Bridged Ethernet/802.3 (no FCS) • LLC Bridged Ethernet (FCS) • LLC Bridged Ethernet (no FCS)

ATM Frame Formats

The format of the ATM AAL5 CPCS-PDU (ATM AAL5 Frame) is shown in Figure 3-22 on page 3-50. The formats of the various types of AAL5 CPCS-PDU payloads for these frames are shown in the following diagrams:

- **BRIDGED:**
  - VC Mux Bridged Ethernet/802.3 (FCS): see Figure 3-23 on page 3-51
  - VC Mux Bridged Ethernet/802.3 (no FCS): see Figure 3-24 on page 3-51
  - LLC Mux Bridged Ethernet (FCS): see Figure 3-25 on page 3-51
  - LLC Mux Bridged Ethernet (no FCS): see Figure 3-26 on page 3-52
- **ROUTED:**
  - VC Mux IPv4 Routed: see Figure 3-27 on page 3-52
  - VC Mux IPv6 Routed: see Figure 3-28 on page 3-52
  - LLC Routed AAL5 Snap: see Figure 3-29 on page 3-53

Figure 3-22. ATM AAL5 CPCS-PDU (ATM AAL5 Frame)

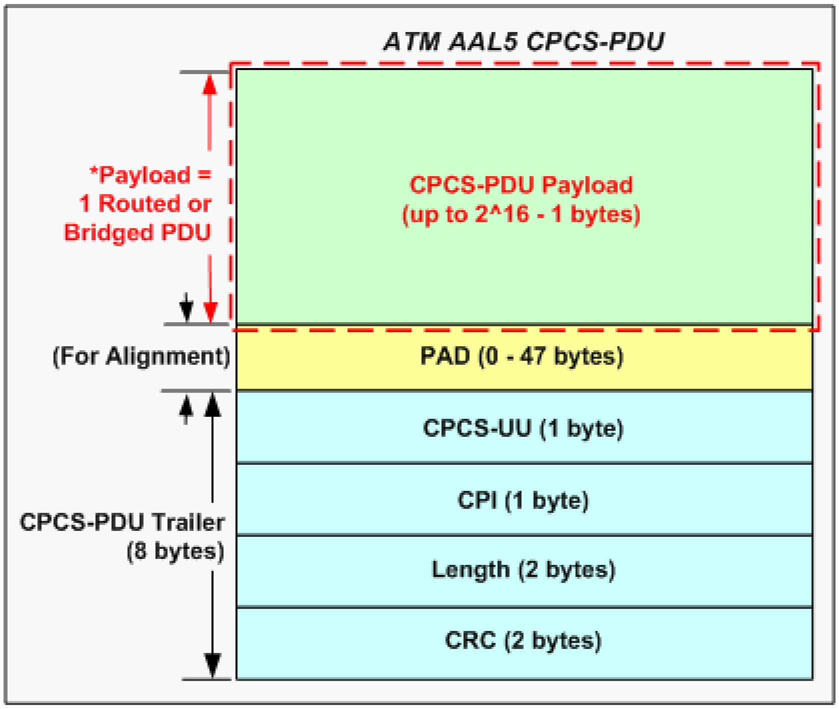


Figure 3-23. VC Mux Bridged Ethernet/802.3 (FCS)

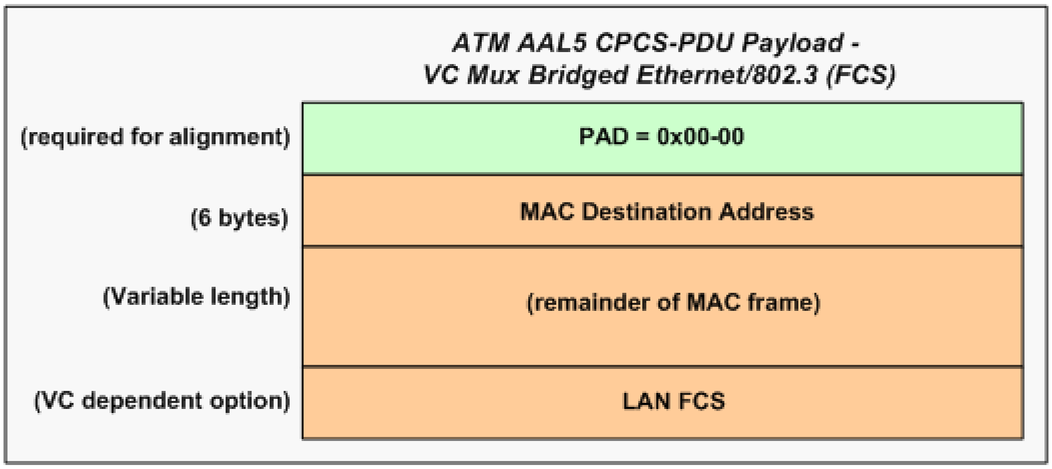


Figure 3-24. VC Mux Bridged Ethernet/802.3 (no FCS)

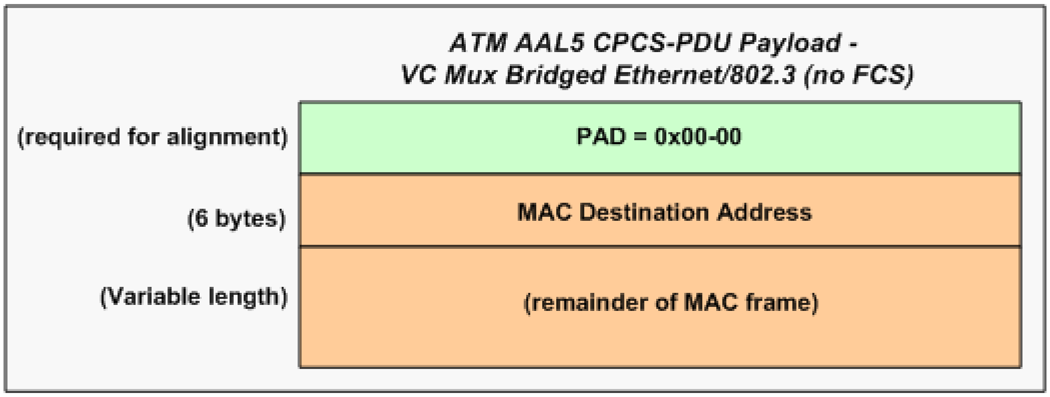


Figure 3-25. LLC Bridged Ethernet (FCS)

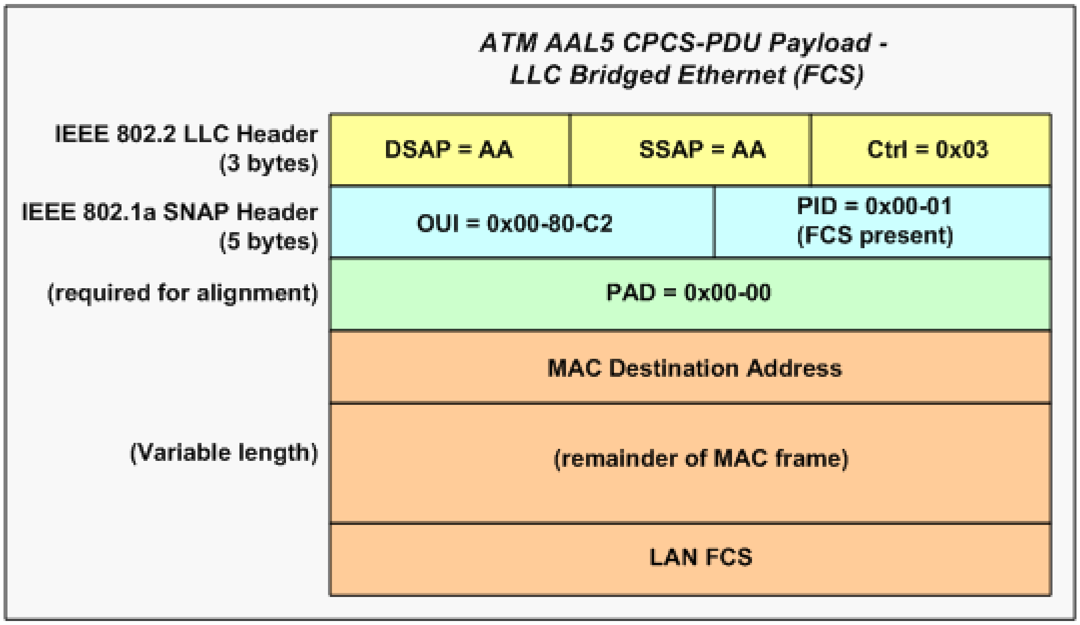


Figure 3-26. LLC Bridged Ethernet (no FCS)

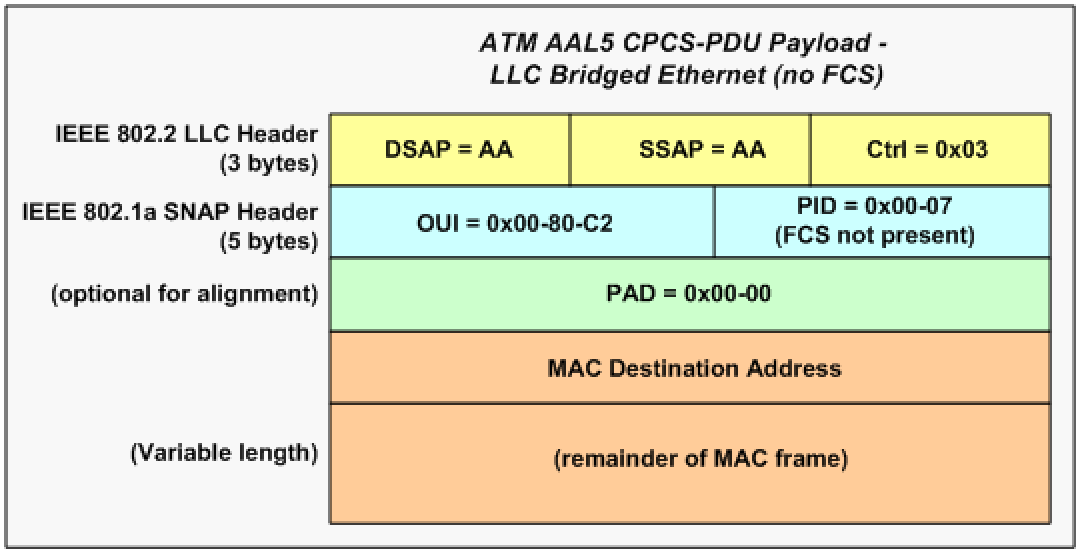


Figure 3-27. VC Mux IPv4 Routed

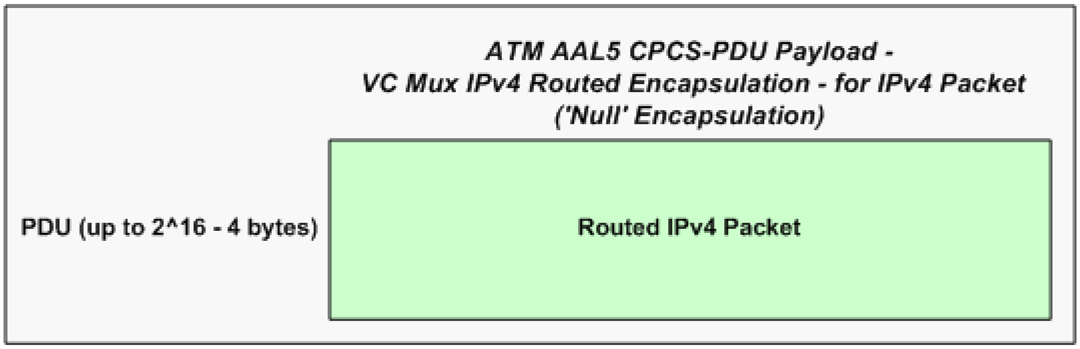

ATM AAL5 CPCS-PDU Payload -
VC Mux IPv4 Routed Encapsulation - for IPv4 Packet
('Null' Encapsulation)
PDU (up to 2^16 - 4 bytes)
Routed IPv4 Packet

Figure 3-28. VC Mux IPv6 Routed

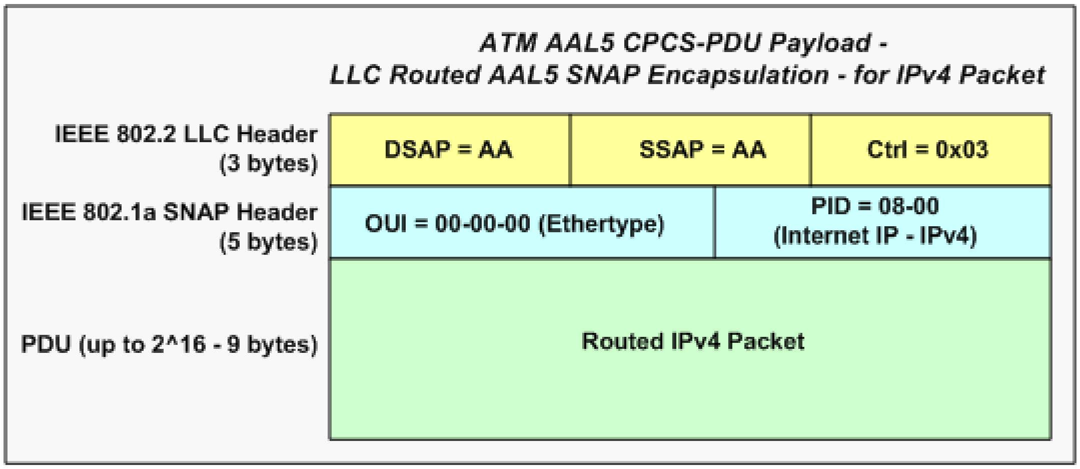
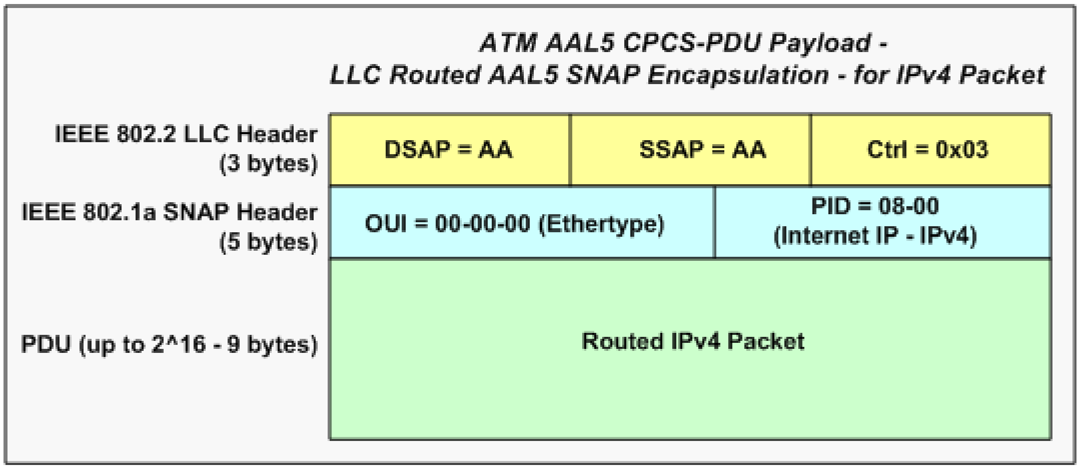


Figure 3-29. LLC Routed AAL5 Snap


ATM AAL5 CPCS-PDU Payload -
LLC Routed AAL5 SNAP Encapsulation - for IPv4 Packet
IEEE 802.2 LLC Header (3 bytes)
DSAP = AA
SSAP = AA
Ctrl = 0x03
IEEE 802.1a SNAP Header (5 bytes)
OUI = 00-00-00 (Ethertype)
PID = 08-00 (Internet IP - IPv4)
PDU (up to 2^16 - 9 bytes)
Routed IPv4 Packet

Generic Routing Encapsulation (GRE)

RFC 2784, ‘Generic Routing Encapsulation’ (GRE), provides a mechanism for encapsulating a payload packet to send that packet over a network of a different type. First, a GRE header is prepended to the payload packet, and the Ethertype for the protocol used in that packet is included in the GRE header. Then, a Delivery header is prepended to the GRE header, which adds a Layer 2 Data Link Layer address plus a Layer 3 Network address (for a network protocol in this implementation, either IPv4 or IPv6). After a GRE-encapsulated payload packet has reached the last router of the GRE ‘tunnel,’ this router removes the GRE header and forwards the payload as a ‘normal’ packet for the native protocol in the network.

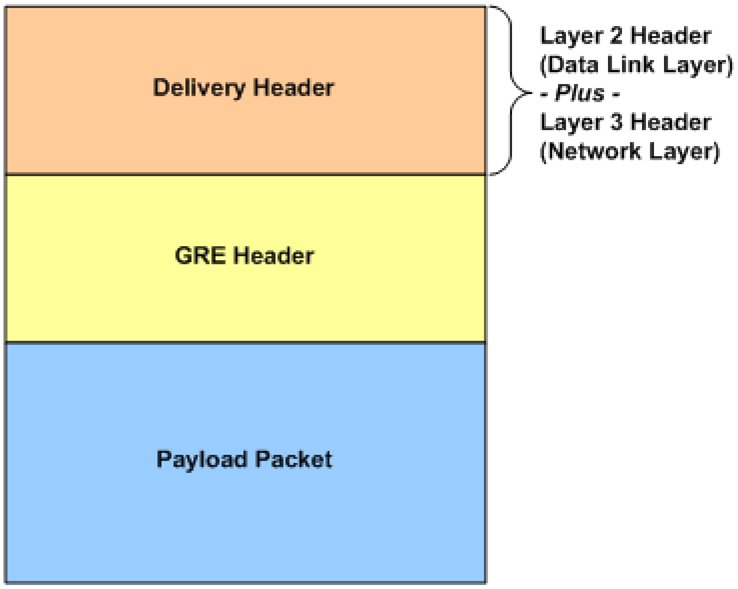
This is a relatively simple type of encapsulation and can be used to transparently carry packets for many different protocols, since it is based on Ethertypes. The original specifications for this encapsulation were RFC 1701, ‘Generic Routing Encapsulation (GRE),’ published in 1994, and RFC 1702, ‘Generic Routing Encapsulation over IPv4 Networks,’ also published in 1994.

RFC 2890, ‘Key and Sequence Number Extensions to GRE,’ provides optional fields for identifying individual traffic flows within a GRE tunnel through an authentication key value, and for monitoring the sequence of packets within each GRE tunnel.

GRE Packet Format

Both control and data packets can be GRE-encapsulated. The overall format of a GRE-encapsulated packet is shown in Figure 3-30 on page 3-54.

Figure 3-30. GRE-Encapsulated Packet



GRE Packet Headers

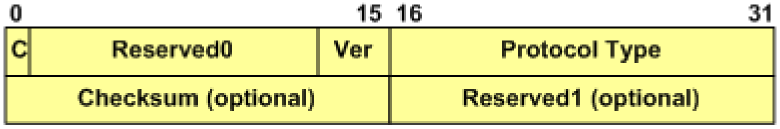
There are two formats for the GRE Packet Headers:

- *GRE Header per RFC 2784* on page 3-54
- *GRE Header per RFC 2890* on page 3-55

GRE Header per RFC 2784

The format of a GRE packet header per RFC 2784 is shown in Figure 3-31 on page 3-54.

Figure 3-31. GRE Packet Header (per RFC 2784)



The fields in the GRE header, per RFC 2784, are described in Table 3-9 on page 3-54.

Table 3-9. GRE Header Fields (per RFC 2784)

Field	Description
C	The Checksum Present flag bit. If set (= 1), the Checksum and Reserved1 fields are present, and the information in the Checksum field is valid.

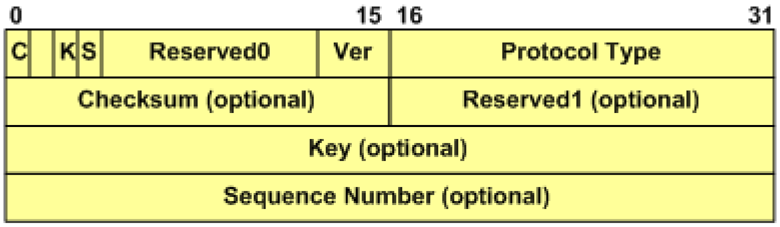
Table 3-9. GRE Header Fields (per RFC 2784)

Field	Description
Reserved0	(Bits 1 - 12) • Bits 1 - 5 unless the receiver is implementing RFC 1701, the receiver must discard the packet if any of these bits are non-zero. • Bits 6 - 12 reserved for future use.
Ver	The Version Number field. The value must be zero.
Protocol Type	Protocol Type field. The protocol type of the payload packet. These values are defined in RFC 1700, ‘Assigned Numbers’ and by the IANA ‘ETHER TYPES’ document. When the payload is an IPv4 packet, the protocol type must be set to 0x800 (Ethertype for IPv4).
Checksum	(Optional) The IP (one’s complement) checksum of all of the 16-bit words in the GRE header and the payload packet. The value of the checksum field = zero for the purpose of computing the checksum. The checksum field is present only if Checksum Present bit is set (= 1).
Reserved1	(Optional) These bits are reserved for future use. This field is present only if the Checksum field is present (that is, the Checksum Present bit = 0). If present, this field must be transmitted as zero.

GRE Header per RFC 2890

The format of a GRE header, with added information per RFC 2890, is shown in Figure 3-32 on page 3-55.

Figure 3-32. GRE Header (per RFC 2890)



The fields in the GRE header, per RFC 2890, are described in Table 3-10 on page 3-56.

Table 3-10. GRE Header Fields (per RFC 2890)

Field	Description
C	The Checksum Present flag bit. If set (= 1), the Checksum field and the Reserved1 is present, and the information in the Checksum field is valid.
Reserved0	Bits 1 - 12. For bits 1 - 5, unless the receiver is implementing RFC 1701 the receiver must discard the packet if any of these bits are non-zero. For bits 6 - 12, these bits are reserved for future use.
K	The Key Present flag bit. If set (= 1), the Key field is present. If not set (= 0), this field is not present. (Compatible with RFC 1701)
S	The Sequence Number Present flag bit. If set (= 1), the Sequence Number Present field is present. If not set (= 0), this field is not present. (Compatible with RFC 1701)
Ver	The Version Number field. The value must be zero.
Protocol Type	Protocol Type field. The protocol type of the payload packet. These values are defined in RFC 1700, 'Assigned Numbers' and by the IANA 'ETHER TYPES' document (located at www.iana.org/assignments/ethernet-numbers). When the payload is an IPv4 packet, the protocol type must be set to 0x800 (Ethertype for IPv4).
Checksum	(Optional) The IP (one's complement) checksum of all of the 16-bit words in the GRE header and the payload packet. The value of the checksum field = zero for the purpose of computing the checksum. The checksum field is present only if Checksum Present bit is set (= 1).
Reserved1	(Optional) These bits are reserved for future use. This field is only present if the Checksum field is present (that is, the Checksum Present bit = 0). If present, this field must be transmitted as zero.

Table 3-10. GRE Header Fields (per RFC 2890)

Field	Description
Key Present	(Optional) This field is present only if the Key Present bit is set (= 1). A 4-octet number that can be used to identify an individual, logical traffic flow within the GRE tunnel. The encapsulator/sender uses the same key value for all packets within a single flow, for identification by the decapsulator/receiver.
Sequence Number Present	(Optional) This field is present only if the Sequence Number Present bit is set (= 1). A 4-octet number that can be used to identify the order of transmission of the packets, with the goal of providing unreliable, but in-order delivery of packets. The decapsulator/receiver uses the sequence number to monitor the order of the packets as they are received. Out-of-sequence packets should be silently discarded. The sequence number of the first packet = 0. The value range is from 0 to (2 ** 32) -1.

DHCP Protocol

Dynamic Host Configuration Protocol (DHCP) is defined in RFC 2131, and it is based on earlier work with the protocol for BOOTP relay agents, which was specified in RFC 951. A DHCP Server provides permanent storage and dynamic allocation of IPv4 network addresses and other network configuration information. A DHCP Server is a host, and a DHCP Client is also a host. This protocol is designed for allocating IPv4 addresses to hosts, but not to routers.

A Client Identifier (Client ID) is required so that the DHCP Server can match a DHCP client with its ‘lease.’ If the Client does not supply a Client Identifier option, the Client Hardware MAC Address (chaddr) is used by the Server to identify the Client. A lease is the period of time that a DHCP Client may use an IPv4 address that has been allocated by the DHCP Server. This lease period may be extended, and may even be set to ‘infinity’ (0xffffffff hex), to indicate a ‘permanent’ IPv4 address allocation.

DHCP messages are exchanged between client and server using UDP as the transport protocol. The DHCP Server port is UDP Port 67, and the DHCP Client port is UDP Port 68.

DHCPDISCOVER messages are broadcast by the Client on the local subnet, to reach the DHCP Server. Suggested values for a network address and lease period may be included in the Discover message. The Server(s) may respond with a DHCPOFFER message. The Offer message includes available IPv4 network address, plus configuration parameters contained in the DHCP options (TLVs/objects).

Note: You will not be able to select DHCP-enabled protocol interfaces for use with Ixia protocol emulations, with the exception of IGMP.

DHCPv6 Protocol

The Dynamic Host Control Protocol for Version 6 (DHCPv6) is defined in RFC 3315. DHCPv6 uses UDP packets to exchange messages between servers and clients. The servers provide IPv6 addresses and additional configuration information to clients. A DHCPv6 server listens on a reserved, link-scope multicast address. A client identifies itself to the server by a link-local source address.

The groups of IPv6 addresses managed by the servers and clients are called Identity Associations (IAs), where each IA has a unique identifier. IA_NAs are identity associations of non-temporary (permanent) IPv6 addresses. IA_TAs are identity associations of temporary addresses.

RFC 3633, ‘IPv6 Prefix Options for Dynamic Host Configuration Protocol (DHCP) Version 6,’ adds capability for *automated* allocation of IPv6 prefixes from a delegating router to a requesting router. IA_PDs are identity associations used for delegated IPv6 address prefixes.

The setup for DHCPv6 involves a four-message exchange ‘handshake.’ Maintaining the DHCPv6 client-server relationship, and managing the return or deletion of IPv6 addresses involves three additional messages. These messages are described in the following list:

Message exchange (handshake):

- SOLICIT: Client sends a DHCPv6 SOLICIT message to the all DHCPv6 Agents multicast address to locate suitable servers.
- ADVERTISE: Multiple servers respond to the client’s SOLICIT message by sending an ADVERTISE message to the client. The Client receives and stores ADVERTISE messages until the first retransmit timeout for SOLICIT messages, then accepts the message with the highest preference value. Or, the client immediately accepts an ADVERTISE message that has the preference value set to 255.
- REQUEST: Client sends a REQUEST message to the DHCPv6 server that has the highest preference value.
- REPLY: Server responds to the client’s REQUEST message with a REPLY message containing the IPv6 address and configuration parameters required by the client.

Additional messages for Maintenance/Return/Deletion of Addresses:

- RENEW: Client sends a RENEW message to the assigned server after the Renew time specified for the IA. The server may respond with a REPLY message.
- REBIND: If the client does not receive a response (REPLY) from the primary (assigned) server, it multicasts a REBIND packet according to the Rebind time specified for the IA. The server(s) may each respond with a REPLY message.
- RELEASE: Client sends a RELEASE message to return one or more IPv6 addresses to the server when it has completed using the IPv6 address(es).
- **Note**: If the client does not receive any REPLY messages from the server in response to its RENEW or REBIND messages, the client deletes the assigned addresses according to the valid lifetimes of the addresses.

Ethernet OAM

The IEEE Std 802.3ah Operations, Administration, and Maintenance (OAM) sublayer provides mechanisms useful for monitoring link operation such as remote fault indication and remote loopback control. In general, OAM provides network operators the ability to monitor the health of the network and quickly determine the location of failing links or fault conditions.

OAM information is conveyed in Slow Protocol frames called OAM Protocol Data Units (OAM PDUs). OAM PDUs contain the appropriate control and status information used to monitor, test and troubleshoot OAM-enabled links.

The addition of Ethernet OAM support in IxOS involves the following :

- support in stream configuration dialogs to send OAM packets.
- support for a PCPU based state machine that is configured to act as a *passive* mode endpoint and reply to OAM packets.

A list of load modules and the Ethernet OAM statistics they can generate are provided in Table B-31 on page B-160. Ethernet OAM statistics counters are defined in Table B-6 on page B-9.