

Software Overview

3

This chapter describes all the software components of the ERX system. It includes a general overview of the software architecture and information on software traffic flow, access protocols, routing protocols, and the QoS mechanisms.

Topic	Page
Architecture Overview	3-2
Software Summary	3-3
ERX Software Traffic Flow	3-5
IP	3-9
IPSec	3-22
DS3/DS1/E3/E1	3-24
SONET	3-26
Switched Multimegabit Data Service (SMDS)	3-28
Routing	3-29
QoS	3-41
Policy Management	3-48
Dynamic Interfaces	3-51
B-RAS Support	3-52
L2TP	3-57
L2F	3-57
MPLS Support	3-58
Encapsulating Layer 2 Services	3-62
DHCP Local Server	3-63
VLAN Support	3-65

Architecture Overview

As network transport becomes more and more critical to daily business operations, service providers need to deploy products that can guarantee maximum uptime. In the ERX system, both the hardware *and* the software are designed to be extremely robust. The system combines a redundant carrier-class hardware design with a software architecture that is modular, object oriented, and component based. This software modularity increases overall system reliability, eases software upgrades, and reduces development time to speed new feature release by making each software component highly autonomous.

Each major software subsystem within the system (such as BGP-4, IP, SNMP, Frame Relay, SONET) is independent and has a set of dedicated resources, such as memory, buffers, and processing cycles. This allows each subsystem to minimize unwanted interactions with other subsystems and thus to minimize negative overall system impact. Since each subsystem has dedicated system resources, no one subsystem can consume an unfair share of memory, buffers, or processing cycles.

In addition, software subsystems can be independently loaded, shut down, and restarted. All actions are dynamic; the loading of one module does not affect the performance or operation of another.

The modularity of the system architecture is in direct contrast to older, monolithic, code-based designs. The modular approach improves stability and reliability by ensuring that the behavior of one program module does not adversely affect others. The system modularity provides several advantages:

- Seamless recovery – Individual software subsystems can be restarted without affecting the rest of the system's operation.
- Increased system uptime – Each software module interacts with the other software modules through a defined set of interfaces, protecting subsystem interactions.
- Predictable software behavior – System resources are protected so that even if a software subsystem has a problem, overall system operation will be minimally affected.
- Nondisruptive software upgrades – Individual modules can be upgraded without affecting overall system operation.

The ERX system also implements a distributed software architecture, where each line module is capable of making routing, QoS, and forwarding decisions. This distribution of software functions makes the

system highly scalable. As more line modules are added to the system, more packet processing capability is added.

The SRP module is responsible for downloading software images to each of the line modules on system boot. In addition, the SRP is also responsible for sending down routing table updates (see *Routing*, later in this chapter). The distributed software processing allows for maximum system performance even when the chassis is fully loaded.

Software Summary

IP has emerged as the dominant traffic type for wide area transport. With this in mind, the ERX system has an IP-optimized software set designed to meet the demanding needs of the service provider edge. The ERX software set combines robust IP routing protocol support with next-generation IP features such as QoS, VPNs, multiple virtual routers, and detailed accounting. This software set is backed by the wire-speed performance of the ERX hardware. The comprehensive software set allows service providers to deploy the system to fulfill immediate subscriber demand for high-speed Internet access, while also enabling them to develop new, high-value IP service plans.

Software Features

The system supports the following software features, each of which is described in more detail in the following sections:

- Robust IP protocol
- Routing protocols: BGP-4, IS-IS, OSPF, RIPv2, VRRP
- PPP (including packet over SONET), Multilink PPP, Frame Relay, Multilink Frame Relay, and ATM, including zero-touch provisioning
- SONET, DS3, DS1, fractional DS1, E3, E1, fractional E1
- MPLS (including RSVP, CR-LDP, Martini and RFC 2547 BGP MPLS)
- Multicasting protocols: DVMRP, IGMP, MBGP, PIM
- VPNs
- Virtual routers
- IP QoS
- Dynamic policy management
- Dynamic interfaces

- B-RAS features: PPP, PAP, CHAP, RADIUS, DHCP
- Comprehensive network management (covered in *Chapter 4*)
- Complete statistics gathering per interface (covered in *Chapter 5*)
- Strong diagnostics (covered in *Chapter 5*)

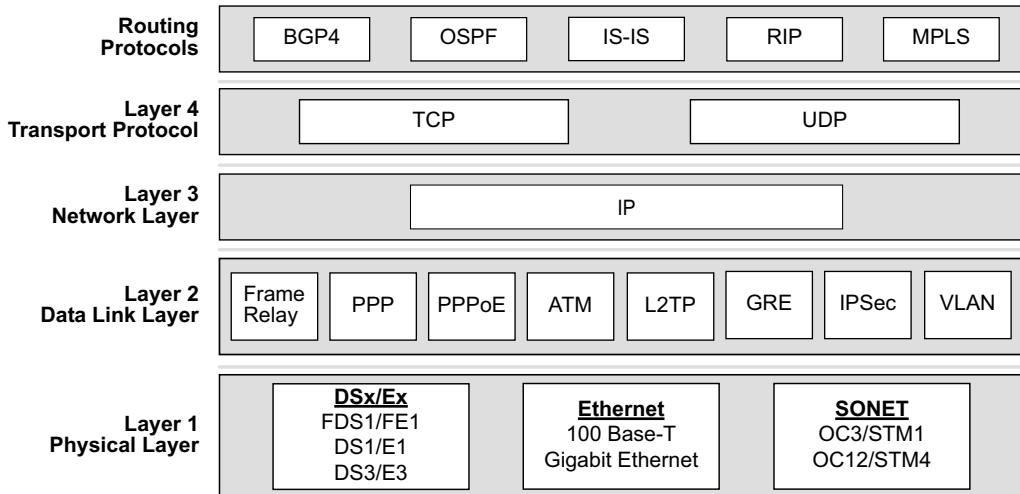


Figure 3-1 ERX system software overview

Figure 3-1 shows the breadth of the ERX system software support. Software functions are layered on top of physical (copper or optical) interfaces. Multiple access protocols (PPP, Frame Relay, ATM) are supported to allow service providers to offer a number of access methods and line speeds to their subscribers. The system is optimized to handle IP connections no matter what the access protocol.

The system also supports a number of protocols that are specific to the B-RAS application. These are shown in Figure 3-26, and include:

- IP/PPP/ATM
- IP/PPP/Ethernet/ATM
- IP/PPP/Ethernet
- IP/PPP/VLAN/Ethernet

For routing, the system supports all major IP-based routing protocols, which give the service provider flexibility in deploying routing protocols. This allows service providers to use protocols such as IS-IS or OSPF for internal network connections and BGP-4 for external connections. The ERX system supports both interior routing protocols (such as RIP, OSPF,

IS-IS) as well as exterior ones (such as BGP-4). Both types of routing protocols are required for large-scale service provider network deployment.

The ERX system also supports the complete suite of multicast routing protocols. IP multicasting improves network efficiency by allowing a host to transmit a datagram to a targeted group of receivers.

In addition to all the protocols listed above, the system supports a leading-edge set of next-generation IP features to allow service providers to implement new IP services. These features include support for IP QoS, IP VPN services, detailed accounting, virtual routing, and IP service plan creation and enforcement.

ERX Software Traffic Flow

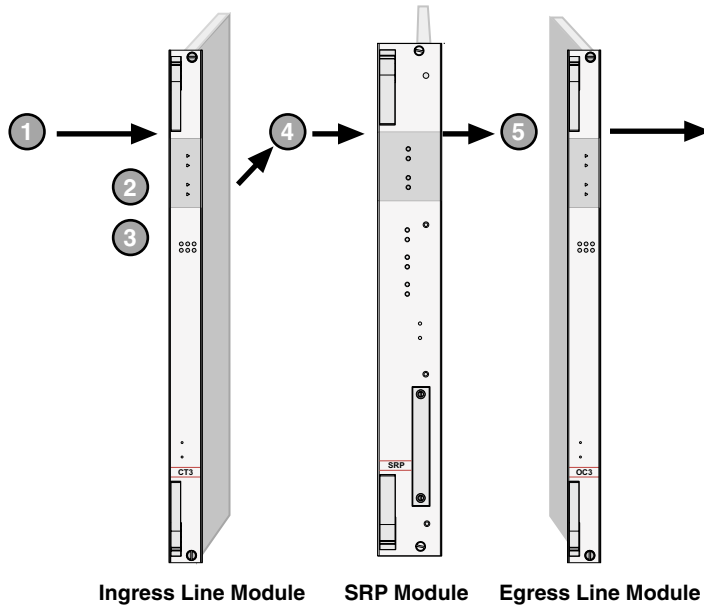
This section describes the software events that occur as traffic is processed by the system.

At a high level, packets are received by an ERX line module interface. Any number of interfaces can be receiving packets simultaneously. The packets are then processed by each line module. Various QoS or routing policies can be applied to make the packet conform to the subscriber's service plan. The packets are then routed out of an ERX interface.

Packet policies can be created and enforced for packets initiated by the subscriber as well as packets destined for the subscriber. This feature is important because it ensures that traffic entering into the network, as well as traffic returning from the network, is in compliance with the subscriber's service plan. It also allows for different policies for CPE outgoing traffic versus CPE incoming traffic.

The ability to set up different policies per traffic direction also prevents misuse of system resources. For example, if traffic from the network is blasting to a specific subscriber destination in excess of its service plan, ERX system resources could be improperly consumed. The system has the intelligence to ensure that the traffic remains balanced for both ingress and egress in accordance with the subscriber's service plan.

Figure 3-2 illustrates the flow of traffic in the system.



- ① Packets are received by a line module.
- ② Packets are classified based on ingress interface, MPLS label, or any field in the packet, and a lookup is performed. Classification could include Gold, Silver, and Bronze traffic queues.
- ③ Actions are performed on the packet as a result of classification.
- ④ Packet is forwarded to the egress destination via IP forwarding function. Packets could be forwarded into a specific traffic queue (such as Gold, Silver, or Bronze).
- ⑤ Line module handles packet scheduling, Layer 2 encapsulation, and transmission of the packet on the egress network. Weighted-based scheduling could provide relative priority between queues (such as Gold, Silver, or Bronze).

Figure 3-2 Packet processing in the ERX system

Packet Processing

The following steps describe how the ERX system processes packets:

- 1 Packets are received by a line module.

Any line module in the ERX chassis can act as the access line module. For a list of available line modules, see *Chapter 2, Hardware Overview*. As the packet is received on the line module, the layer 2 header is removed from the IP packet.
- 2 The packet is classified and a lookup is performed. The result of the classification and lookup is an internal egress destination address.

The packet can be classified in a number of ways:

- Layer 2 – classifies the packet based on the ingress interface. This allows the service provider to define subscriber-specific policies, such as map all traffic from port X or virtual circuit Y to the “Gold” service.
- Layer 2+ – classifies the packet based on the MPLS label. This allows the service provider to use MPLS labels to assign traffic paths.
- Layer 3 – classifies the packet based on any field in the packet, including source IP address, destination IP address, DiffServ (DS) field, and application type and port number. This gives the service provider more fine-grained control over service definition. For example, “map all traffic that is VoIP for low-latency handling,” or “map all traffic received from this subscriber source address to a best-effort service.”

Once the packet is classified, the lookup function finds the internal egress destination address for the packet and tags the packet with the internal route address. The packet is placed in a service queue based on its classification.

- 3 A number of actions can be performed on the packet as a result of its classification, including:
 - Permit or deny forwarding based on the subscriber’s service profile. For example, if a subscriber is restricted from targeting a specific destination address, drop all packets.
 - Target the internal route path to assign the packet to a certain level of treatment. For example, forward the packet into a specific service queue (such as Gold, Silver, or Bronze).
 - Give a service queue a scheduling weight or a fixed amount of bandwidth.
 - Police the packets to a specified level. Up to two thresholds can be set. This allows packets to be set as committed (green), conforming (yellow), or excess (red).
 - Mark the packet with a label to indicate treatment by the rest of the network. Marks could include MPLS labels, DS byte, or discard-eligible (DE) bits. For example, either identify all VoIP traffic as high priority by appending an MPLS label that designates a priority handling; or, if a subscriber’s traffic is bursting out of his or her traffic profile, mark the packet as DE so that these packets can be dropped if congestion occurs.

- Apply a suite of VPN services. Depending on the service provider demand, these services could include ATM PVC mapping, routing policy assignment, virtual router creation, and VPN inclusion. For example, map traffic from an incoming subscriber to an ATM PVC that serves the VPN destination.
- Gather statistics or accounting information about the packet or flow.
- Direct traffic to an IPSec or Tunnel Service module for further specialized processing.
- Any combination of the above actions can be used together. For example, a subscriber could be given Gold service with a guaranteed bandwidth of 384 Kbps.

Figure 3-3 shows how traffic might be segmented into three traffic queues based on classification, and it shows the resulting actions.

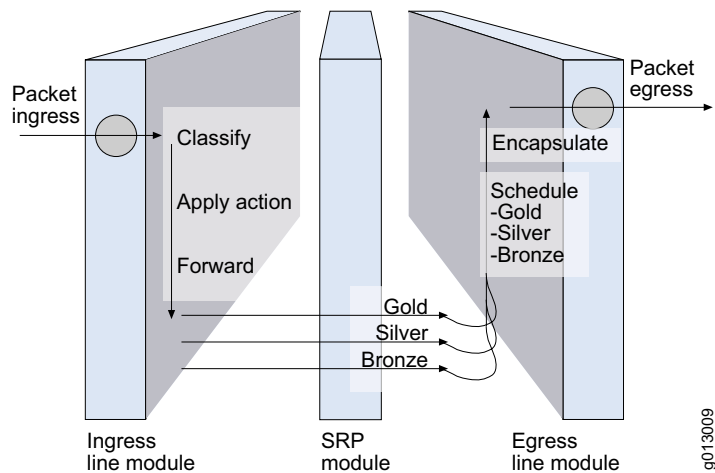


Figure 3-3 Segmenting and processing traffic

- 4 The packet is forwarded internally to the egress destination via the IP forwarding function.

The switch fabric located on the SRP module handles the packet forwarding to the internal egress port destination.

- 5 The line module handles packet scheduling, layer 2 encapsulation, and transmission of the packet onto the egress network.

Any line module in the ERX chassis can route packets to the egress network. For a list of available line modules, see *Chapter 2, Hardware Overview*. The line module handles packet scheduling. The packet scheduler draws from the packet queues that are targeted at that interface.

The queues can be emptied according to one of the following criteria: weight-based priority, rate-based priority, or strict priority. These criteria can be used independently or combined together and are configured by the service provider to match the QoS policies. Weight-based scheduling provides relative priority between queues and is commonly used for Gold/Silver/Bronze type services. Rate-based scheduling provides a fixed rate guarantee that might be useful for some applications. Strict policy-based scheduling provides guarantees for low-latency applications, like VoIP.

If there is congestion, the packets that have been marked out of profile can be dropped by the line module.

Once the packet is scheduled, the router on the line module removes the internal routing tag, encapsulates the packet with a layer 2 header (ATM, POS, Frame Relay, PPP, VLAN), and transmits the packet to the network. All traffic flows are handled at wire speed. For more detailed information on classification, actions, and scheduling, see *QoS* later in this chapter.

IP

Internet access is now a business necessity for most companies, and IP is the protocol of the Internet. With this in mind, Juniper Networks developed a highly robust software set that is optimized to process IP packets at wire speed.

IP Features

The following IP features are supported:

- RFC 2236 – Internet Group Management Protocol, Version 2 (November 1997)
- RFC 1812 – Requirements for IP Version 4 Routers (June 1995)
- RFC 1122 – Requirements for Internet Hosts—Communication Layers (October 1989)
- RFC 1112 – Host Extensions for IP Multicasting (August 1989)
- RFC 950 – Internet Standard Subnetting Procedure (August 1985)

- RFC 922 – Broadcasting Internet Datagrams in the Presence of Subnets (October 1984)
- RFC 919 – Broadcasting Internet Datagrams (October 1984)
- RFC 854 – Telnet Protocol Specification (May 1983)
- RFC 793 – Transmission Control Protocol (September 1981)
- RFC 792 – Internet Control Message Protocol (September 1981), including ICMP error messages (Destination Unreachable, Redirect, Time Exceeded, and Parameter Problem), as well as ICMP query messages (Echo, Timestamp, Address Mask, Router Discovery Protocol) on all connected networks supporting either IP multicast or IP broadcast addressing
- RFC 791 – Internet Protocol DARPA Internet Program Protocol Specification (September 1981)
- RFC 768 – User Datagram Protocol (August 1980)
- Classless Interdomain routing (CIDR) – 32-bit address (network and host ID) with support for prefix length indicating the number of bits in the network prefix
- Maximum transmission unit (MTU)
- Reachability commands – **traceroute** and **ping**
- Support for simultaneous multiple logical IP stacks on the same ERX system. Each IP stack operates independently with its own data structures and is assigned to a set of IP interfaces in the chassis. This feature enables full support for multiple logically separate virtual routers.
- Flexible IP address assignment to support any portion of a physical interface (for example, a channel or circuit), exactly one physical interface, or multilink PPP interfaces
- IP fragmentation
- Loose source routing to specify the IP route
- Strict source routing to specify the IP route for each hop
- Record route to track the route taken
- Internet timestamp
- Broadcast addressing, both limited broadcast and directed broadcast
- Bridged IP
- Cisco HDLC

- Response Time Reporter (RTR)
- Support for 48,000 discrete, simultaneous IP interfaces per system to support thousands of logical connections
- Capability of all line modules to detect and report changes in the state (up or down) of any IP interfaces resulting from changes to the state of the underlying physical interface.
- Shared IP interfaces over the same layer 2 logical interface, which enable more than one IP interface to share the same logical resources
- Subscriber interfaces over an Ethernet interface; this feature allows demultiplexing of traffic associated with different subscribers

The ERX system implements an IP stack architecture that supports dynamic configuration to ease network changes with minimal service disruption. All IP stack configuration information is stored in nonvolatile storage, and all changes to the stack's configuration are dynamic, without requiring the system to be restarted.

An interface table stores information about each interface associated with an IP stack, including index, IP address, prefix length (subnet mask), association with lower-layer interfaces, MTU, state (up or down), and interface state timer. This interface-dense support allows service providers to deploy the system for IP session termination applications, such as aggregating the output from DSLAMs for B-RAS applications.

Profiles

You can configure an IP interface dynamically by creating a profile. This capability allows you to manage a large number of IP interfaces efficiently by creating a profile with a specific set of characteristics. In addition, you can create a profile to assign an IP interface to a virtual router.

A profile can contain one or more of the following characteristics:

- Access routes
- IP address
- Directed broadcast forwarding
- Maximum transmission unit
- Transmission of ICMP redirect messages
- Unnumbered IP address
- Virtual router

RTR

The Response Time Reporter (RTR) feature allows you to monitor your network's performance and its resources. It does this by measuring response times and the availability of your network devices.

IP/PPP

The ERX system supports IP/PPP on any module that is channelized to fractional or full T1/E1 or T3/E3, including modules that channelize down to T1/E1 or T3/E3 (for example, CT3, cOCx/STMx modules) and IP/PPP/SONET on any module that supports SONET/STM interfaces (for example, the OCx/STMx modules). IP/PPP support allows service providers to accept traffic from subscribers who have CPE equipment, such as routers with PPP interfaces, and to output traffic in PPP format to other network devices.

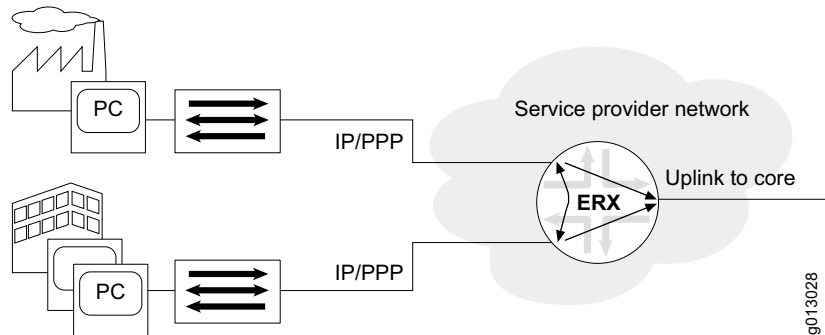


Figure 3-4 The ERX system supports IP/PPP connections from the CPE

Figure 3-4 shows that the ERX system supports the incoming IP/PPP traffic from the CPE. This traffic can then be routed to the uplink(s) attached to the system or to other CPEs that are attached to the system.

PPP Features

The following PPP features are supported:

- RFC 2615 – PPP over SONET/SDH (June 1999)
- RFC 1661 – The Point-to-Point Protocol (PPP) (July 1994)
- RFC 1662 – PPP in HDLC-like Framing (July 1994)
- RFC 1332 – The PPP Internet Protocol Control Protocol (IPCP) (May 1992)
- OSI Network Layer Control Protocols (OSINLCP)

- Bit-synchronous HDLC support on the E1/T3/E3 modules; byte-synchronous support on the OC3/STM1/OC12/STM4 modules
- Link Control Protocol (LCP) support – maximum receive unit (MRU) (RFC 1661), magic number (RFC 1661): Detection of looped-back links recommended for SONET
- Reporting of link events (up or down) to the IP layer

PPP Statistics Tracking

The PPP layer is capable of tracking a number of statistics:

- Total number of frames discarded on input that are not LCP, IPCP, IP, OSINLCP, or OSI frames
- Number of frames received with an incorrect address/control field
- Number of frames transmitted that exceeded the MTU
- Number of frames received with a checksum error
- Number of frames and bytes successfully received and transmitted
- Statistics collected via the Generic Interface MIB, including:
 - > Octets received and transmitted
 - > Receive errors, discards, and unknown protocols
 - > Transmit errors and discards
 - > Received and transmitted unicast packets
- Statistics collected via the PPP MIB, including bad addresses, controls, frame check sequence (FCS), and packets too long

PPP Structure

As shown in Figure 3-5, PPP can exist directly on top of the HDLC layer or on top of a layer 2 ATM interface. In either case, IP rides on top of PPP, providing support for IP/PPP/ATM and IP/PPP/HDLC. Both SONET and DSx/Ex interfaces are supported at the physical layer.

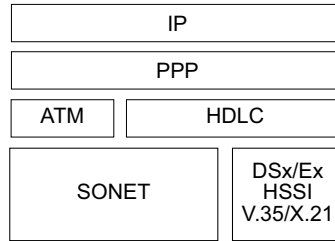


Figure 3-5 Structure of PPP

Multilink PPP

The ERX system supports MLPPP on any modules that are channelized to fractional or full T1/E1. MLPPP aggregates multiple physical links into a single logical bundle. More specifically, MLPPP bundles multiple link-layer channels into a single network-layer channel. Peers negotiate MLPPP during the initial phase of LCP option negotiation. Each system indicates that it is multilink capable by sending the multilink option as part of its initial LCP configuration request.

The systems joined by the multilink each assign the same unique name to the bundle. A bundle can consist of multiple physical links of the same type—such as multiple asynchronous lines—or can consist of physical links of different types—such as leased synchronous lines and dial-up asynchronous lines.

The system treats MLPPP like another PPP Network Control Protocol (NCP). Packets received with an MLPPP header are subject to fragmentation, reassembly, and sequencing. Packets received without the MLPPP header cannot be sequenced and can be delivered only on a first-come, first-served basis

IP/Frame Relay

The ERX system supports IP over Frame Relay PVCs on any module that supports fractional or full T1/E1 or T3/E3. The system supports Frame Relay over POS interfaces on the OCx/STMx POS modules.

With this interface, the service provider can:

- Take in traffic from subscribers that have CPE equipment such as routers with Frame Relay interfaces
- Take in traffic from other network devices that send output in Frame Relay such as Frame Relay switches
- Use Frame Relay as an uplink technology on an unchannelized T3 or E3 link

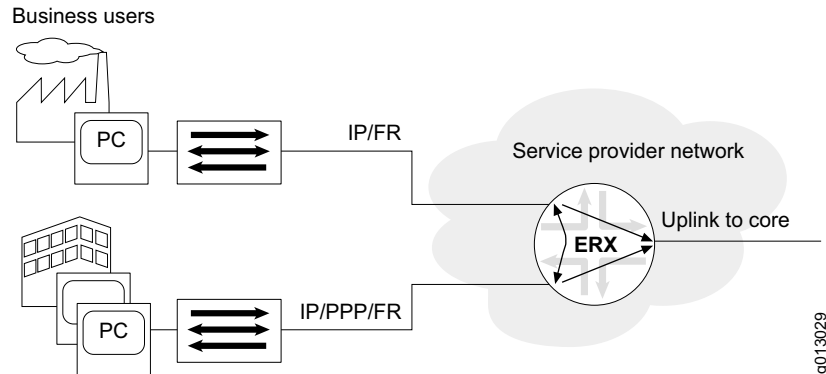


Figure 3-6 ERX system IP/PPP and IP/Frame Relay access connections

Figure 3-6 illustrates that the ERX system can accept a number of access technologies and support return paths to the CPE or paths to the core. Typical Frame Relay interfaces come from a CPE-based router or a public network-based Frame Relay switch.

Frame Relay Features

The ERX system supports the following Frame Relay features:

- RFC 2427 – Multiprotocol Interconnect over Frame Relay (September 1998) – standards compliance for IP and IS-IS encapsulation (obsoletes RFC 1490)
- Configuration of either DTE or DCE Frame Relay User-to-Network Interfaces (UNIs) or network-to-network Interfaces (NNIs)
- Packet format: 8188-byte information field size (8192 minus 2 bytes for the address and a 16-bit CRC) or 8186-byte information field size (8192 minus 2 bytes for the address and a 32-bit CRC)
- Multiple Frame Relay subinterfaces per major interface
- Local management interface (LMI) support in adherence with the ANSI (T1.617 Annex D), CCITT (ITU-T Q.933 Annex A), and original Group of Four (DEC, Nortel, Stratacom, and Cisco) specifications
- End-to-End Fragmentation and Reassembly per Frame Relay Forum – Frame Relay Fragmentation Implementation Agreement, FRF.12 (December 1997)
- Support for 1,000 Frame Relay VCs on a single line module

Figure 3-7 shows the structure of the ERX system Frame Relay interface. Each Frame Relay major interface sits on top of an HDLC interface. The Frame Relay implementation is divided into two levels: a “major” interface and one or more “subinterfaces.” This line allows a single physical interface to support multiple logical interfaces. Multiple IP interfaces can also be assigned to each Frame Relay major interface via the subinterfaces.

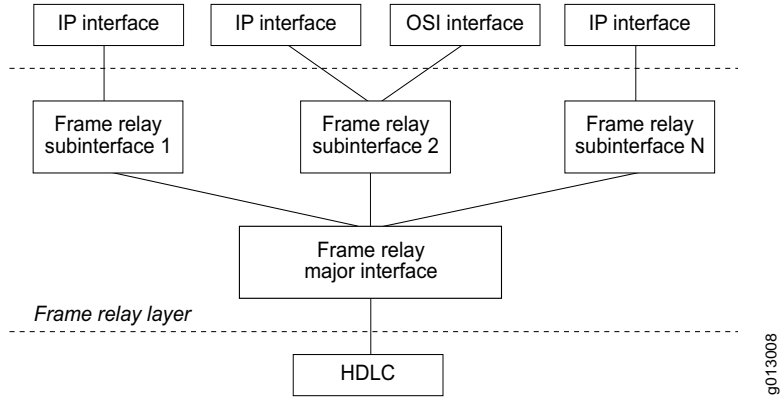


Figure 3-7 ERX system Frame Relay interface design

Figure 3-8 shows the structure of the Frame Relay protocols with the physical layer as the foundation. For Frame Relay, this can be E1, E3, T1, T3, or fractional amounts, as supported by the different line module ports. The HDLC layer is on top of the physical layer and can support flexible assignment of physical resources. For example, an HDLC channel can support one DS0, fractional T1s, or an entire T1. The major Frame Relay interface sits on top of the HDLC resource, and the subinterfaces sit on top of the major interface. The Frame Relay subinterfaces connect to the IP interface layer.

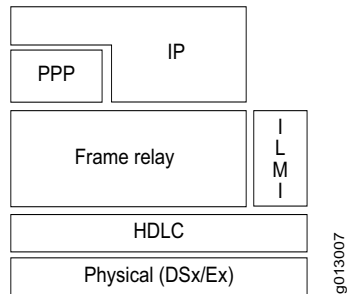


Figure 3-8 Structure of Frame Relay protocols

The ERX system supports Frame Relay LMI (local management interface) to provide the operator with configuration and status information relating to the Frame Relay VCs in operation. LMI specifies a polling mechanism to receive incremental and full status updates from the network. The system can represent either side of the UNI interface and supports unidirectional LMI. Bidirectional support for NNI is also provided.

Juniper Networks supports only end-to-end fragmentation per the FRF.12 Implementation Agreement standard. Unlike UNI and NNI fragmentation, end-to-end supports fragmentation only at the endpoints. End-to-end fragmentation and reassembly are supported only on nonmultilink Frame Relay interfaces on cOC12/STM4 and CT3 12 FO modules.

Configurable LMI Parameters

The following LMI parameters can be configured on the system:

- The DTE *keepalive* polling interval to determine how often the DTE polls the network for a status request (T391: range 5–30 seconds)
- The DCE network response interval for status verification (T392: range 5–30 seconds)
- The DTE number of normal status requests before a full status is requested (N391: range 1–255)
- The number of error counts such as invalid sequence number, CRC error, or nonreceipt of a status response (N392: range 1–10) of the last “n” packets received (N393: range 1–10) before the interface is to be marked inactive

Multilink Frame Relay

The ERX system supports Multilink Frame Relay (MLFR) on any modules that are channelized to fractional or full T1/E1. The system also supports MLFR over POS interfaces on the OCx/STMx POS modules.

MLFR aggregates multiple physical links into a single logical bundle. More specifically, MLFR bundles multiple link-layer channels into a single network-layer channel.

The systems joined by the multilink each assign the same unique name to the bundle. A bundle can consist of multiple physical links of the same type—such as multiple asynchronous lines—or can consist of physical links of different types—such as leased synchronous lines and dial-up asynchronous lines.

The system treats MLFR like nonmultilink Frame Relay. Packets received with an MLFR header are subject to sequencing. Packets received without the MLFR header cannot be sequenced and can be delivered only on a first-come, first-served basis.

IP/ATM

The ERX system supports IP over ATM PVCs and SVCs on modules that support T3/E3 or SONET/STM interfaces (for example, T3/E3, OC3c/STM1, and OC12c/STM4). This feature allows service providers to take in traffic from subscribers that have CPE equipment such as routers with ATM interfaces, to take in traffic from other network devices that send output in ATM such as DSLAMs, and to connect to service providers with ATM backbone structures. See Figure 3-9.

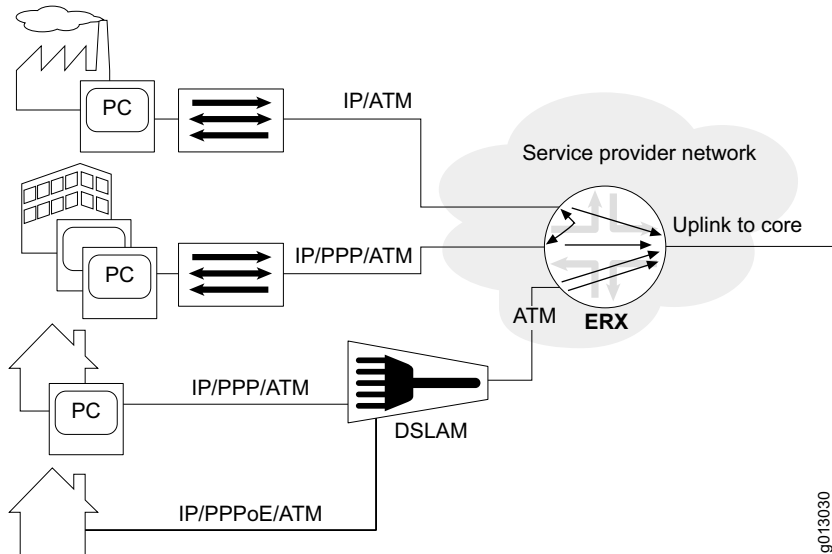


Figure 3-9 ERX system IP/ATM access connection

ATM SVCs

In addition to PVCs, ATM supports switched virtual circuits (SVCs). PVCs provide a static connection that is usually set up manually. SVCs are established on demand via signaling. Table 3-1 shows the differences between PVCs and SVCs.

Table 3-1 Differences between PVCs and SVCs

PVCs	SVCs
Connected on a permanent basis. Users are charged a flat rate.	Dynamically connected as needed. Users are charged only for time and resources used.
Manually configured, permanent connections. Each PVC must be configured at both end systems and on all ATM switches in the network.	Dynamically signaled. SVCs are configured only on the end systems; they do not require configuration on all ATM switches in the network.
Provisioned when the connection is set up. Bandwidth and services allocated to the PVC are not available to other applications even when not in use.	Can request bandwidth and ATM service quality information needed for a particular connection. Once the connection is released, network resources are made available to other users or applications.
Cannot take alternate routes in the event of a failure in the network.	Can take alternate routes in the event of a failure in the network.

ATM Features

The ERX system supports the following ATM features:

- IP transport over ATM – RFC 2684 – Multiprotocol Encapsulation over ATM Adaptation Layer 5 (September 1999), which replaces RFC 1483. Both methods of RFC 2684 encapsulation are supported: LLC encapsulation, where the LLC header is used to multiplex multiple protocols on a single circuit; and VC multiplexing, where no headers are used and only a single protocol may use that circuit.
- ATM PVCs and SVCs
- ATM Forum User-Network Interface (UNI) versions: 3.0, 3.1, 4.0 selectable per port
- ATM adaptation layer 5 (AAL5) frame format
- ATM monitoring per interface with support for ILMI. The ILMI version is selectable and optionally enabled on a per-port basis. ATM ILMI MIB – MIB-II System Group, Physical Port Group, ATM Layer Group, Virtual Path Group, and Virtual Channel Group are supported.
- ATM monitoring per circuit or path with support for operations, administration, and maintenance (OAM). Specifically, the system supports F4 and F5 cell flows, including ATM ping. Two types of tests are supported to allow for test isolation: from the ERX system to the end host and from the ERX system to the ATM switch.
- Traffic shaping for UBR, VBR-RT, VBR-NRT, and CBR service on an individual VC basis.

- Connection admission control (CAC), which ATM networks use to determine whether to accept a connection request, based on whether allocating the connection's requested bandwidth would cause the network to violate the traffic contracts of existing connections.
- Point-to-multipoint connection using nonbroadcast multiaccess (NBMA) interfaces, allowing you to connect an ERX system to multiple stations.

ATM Protocol Structures

Figure 3-10 shows the structure of the ATM protocols with the physical layer as the foundation. For ATM, this can be SONET, DS3, or E3 as supported by the different line modules. The major ATM interface sits on top of the SONET/DS3/E3 resource, and the subinterfaces sit on top of the major interface. The ATM subinterfaces connect to the IP/OSI interface layer.

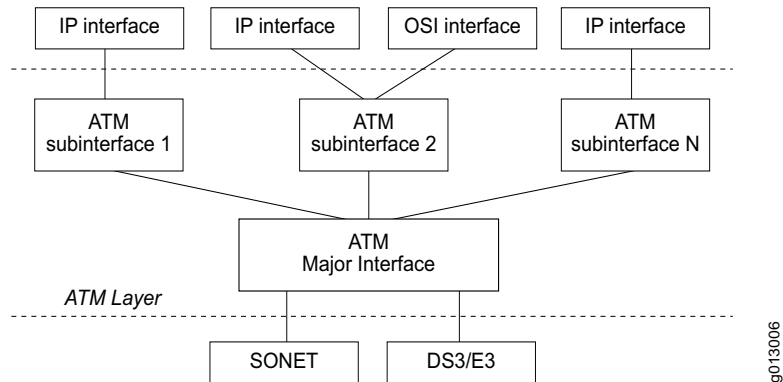


Figure 3-10 Structure of the ATM interface design

Figure 3-11 shows the structure of the ATM protocols. The physical layer (SONET and DS_x/E_x) is the foundation and provider of layer 1 framing service. The ATM layer is on top and provides cell, circuit, and OAM services. The AAL5 layer provides a frame-oriented interface to the ATM layer. ILMI provides local management across the UNI.

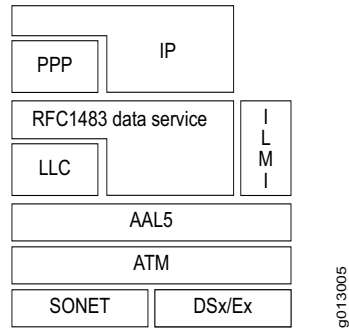


Figure 3-11 Structure of ATM protocol

The AAL5 layer also supports RFC 2684 encapsulation. The RFC 2684 data service layer defines the ERX system subinterfaces that are used by higher-level ATM services. The ATM subinterfaces connect to the IP interface layer.

IP/Cisco HDLC

The ERX system supports IP over Cisco HDLC on any module that supports HSSI, T1/E1, T3/E3, or SONET/STM interfaces.

Cisco HDLC monitors line status on a serial interface by exchanging keepalive request messages with peer network devices. It also allows routers to discover IP addresses of neighbors by exchanging SLARP address request and address response messages with peer network devices.

The ERX system Cisco HDLC is compatible with Cisco Systems HDLC protocol, the default protocol for all Cisco serial interfaces.

Cisco HDLC Features

The ERX system supports the following framing features:

- HDLC for data-link framing
- 18,000-byte information field size

The system also supports the exchange of keepalive messages that identify inactive or failed connections.

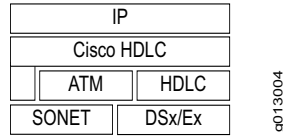


Figure 3-12 Structure of Cisco HDLC protocol

Figure 3-12 shows that the Cisco HDLC protocol can exist directly on top of the HDLC layer or ATM or SONET interface. Both SONET and DSx/Ex interfaces are supported at the physical layer.

Bridged IP

Your ERX system supports bridged IP (RFC 1483) on the T3-ATM, E3-ATM, and OC3 modules. You can configure a bridged IP interface on your ERX system to manage IP packets that are encapsulated inside an Ethernet frame running over a permanent virtual circuit (PVC).

When you configure a bridged IP interface, you can also configure the system as a relay agent that forwards Dynamic Host Configuration Protocol (DHCP) broadcasts.

Bridged Ethernet

Bridged Ethernet allows multiple upper-layer interface types (IP, PPPoE, and CBF) to be simultaneously multiplexed over the same interface. You can set up the system to either terminate interfaces and route data or to pass data transparently through the system to another terminating device. This capability supports multiple client devices that use both IP and PPPoE encapsulation over an Ethernet LAN.

IPSec

The system supports IP security (IPSec), which consists of a set of standards used to provide privacy and authentication services at the IP layer. There are two aspects of IPSec: packet level transport and key management.

Packet Level Transport

Packet level transport involves the encapsulation, encryption, and authentication of the data itself. It also provides antireplay protection.

Encapsulation

IPSec defines two encapsulation protocols. The system supports both of these protocols:

- Authentication Header (AH) – ensures the integrity of the data and authenticates the origin of the data
- Encapsulating Security Payload (ESP) – protects the confidentiality and integrity of the data and authenticates the origin of the data

Encryption

Encryption provides confidentiality of data by making it unreadable to everyone but those to whom the data is intended. Encryption uses the concept of an encryption key to encrypt data. A key can be thought of as a mathematical routine that is applied to a string of data to scramble the contents of the packet. Once encrypted, the recipient must use the same key (or a key that applies the mathematical routine in reverse order) to unscramble the contents of the packet.

IPSec implementations must support two encryption algorithms. Those are:

- DES – Data Encryption Standard
- 3DES – Triple Data Encryption Standard

Authentication

Authentication guarantees that data was not altered during transmission. It is accomplished by applying a hash algorithm across the data and generating a 96-bit value or fingerprint that is appended to the data. Since the pair of communicating parties (for example, two ERX systems) shares the same hash algorithm, the hash result should be the same for both sides. If it is not the same, the data has been compromised.

The system supports two hash algorithms:

- MD5 – Message Digest 5
- SHA-1 – Secure Hash Algorithm

Antireplay

Antireplay prevents a third party from eavesdropping on an IPSec conversation, stealing packets, and injecting those packets into the session at a later time. To provide antireplay protection, IPSec uses sequential counters to guarantee that packets are received and processed in order. Packets that are received out of sequence are dropped.

Key Management

Key management allows the system to create IPsec tunnels without requiring administrators to manually program keys. Key management, or the ability to dynamically negotiate keys between two parties, relies on the Internet Key Exchange (IKE) protocol. IKE is used to negotiate a key exchange between two IPsec nodes so that both devices can share data across an IPsec tunnel. There are two important aspects of key management: key generation and authentication.

Key Generation

IKE uses an algorithm called Diffie-Hellman to generate secret session keys without ever having to send the actual keys themselves across the communication line. Diffie-Hellman uses numbers exchanged by two communicating devices and applies exponential math to derive a shared secret key that allows two parties to encrypt and decrypt data across an IPsec tunnel. Using Diffie-Hellman, two devices (such as two ERX systems) can create secure tunnels to one another without requiring the manual programming of keys on each system. This automatic generation of keys provides a much more scalable model than manual tunnel creation as well as a more secure model, since new keys are generated with each new session.

Authentication

One thing that Diffie-Hellman cannot do is determine if the two routers establishing a key management session are, in actuality, who they claim to be. For example, ERX-1 might be negotiating a key session using Diffie-Hellman with ERX-2. A third party using the IP address of ERX-2 could intercept the exchange so that all data exchanged between ERX-1 to ERX-2 passes through the intruder. This is called a man in the middle attack. To overcome this vulnerability, the devices attempting to establish an IPsec tunnel must be authenticated.

Not to be confused with the authentication of data as described in the *Packet Level Transport* section earlier in this chapter, this authentication is relative to the identity of the two end stations attempting to set up an IPsec tunnel. To provide this authentication, the ERX system supports preshared keys (secrets).

Again, the problem to resolve is that to generate a secret session key to be used to create an encrypted IPsec tunnel, there must be a way to make sure that the party with which we are attempting to communicate is not an imposter. To do this, a shared secret is manually installed on each pair of machines, which is relevant only to the IPsec communication between

these two devices. If the two machines share the same secret, they can establish the authenticity of each other.

DS3/DS1/E3/E1

The ERX system supports a number of line rates; these are listed per line module below:

- CT3 line module supports channelized DS3 (channelized to DS1, fractional DS1, or the DS0 level).
- T3 line module supports unchannelized DS3 and fractional T3.
- CT1 line module supports T1 and fractional T1.
- E3 line module supports unchannelized E3.
- CE1 line module supports E1 and fractional E1.
- cOCx/STMx line modules support T3, E3, T1, E1, and DS0 rates.

For specifications on line modules, see *Chapter 2, Hardware Overview*.

A variety of protocols are supported over these interfaces, including IP/Frame Relay, IP/ATM, IP/PPP, as well as the protocols to enable B-RAS services (see *B-RAS Support* in this chapter for more information). The ERX system DSx and E1/E3 implementations support termination, statistics gathering, alarm surveillance, and performance monitoring. These links can be used for either network ingress or network egress.

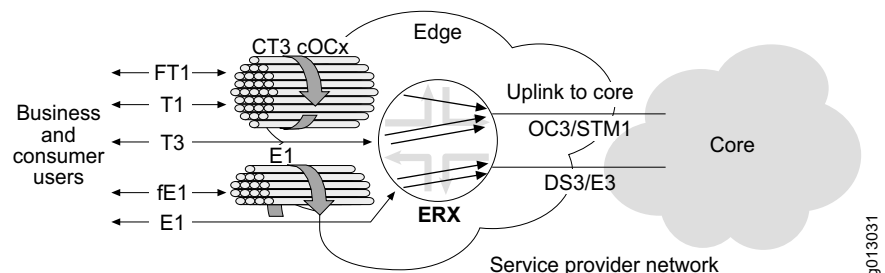


Figure 3-13 The ERX system supports fractional T1/E1 through T3/E3 interfaces

As shown in Figure 3-13, the system can support fractional, full, and channelized interfaces.

Line Module Support

The line modules listed above support:

- Three different clocking options: internal timing, loop timing, and chassis timing
- DS3 and DSI bit error rate tests (BERTs)
- DS3 framing type – both M23 framing and C-bit parity
- DS1 framing type – both D4 framing mode and ESF framing mode
- DS3 loopback – for line, payload, diagnostic, and DS1 (see *Diagnostic Support in Chapter 5* for more information)
- DS1 loopback – for line, payload, and diagnostic (see *Diagnostic Support in Chapter 5* for more information)
- SONET/SDH loopback
- DS3/DS1 line status/alarm monitoring
- DS1 line coding type – both AMI line encoding and B8ZS line encoding
- Unique IP interface support – for each PPP or Frame PVC interface

Configurable HDLC Parameters

The following HDLC parameters are configurable:

- Mapping of DS0 timeslots for T1/FT1 DS0 mapping
- Setting the speed of the DS0 to Nx56 or Nx64
- HDLC CRC checking (enable/disable)
- HDLC CRC algorithm (CRC16 or CRC32)
- Channel Data Inversion (enable/disable)
- MRU
- MTU

Statistics are also gathered per line interface.

SONET

The ERX system supports:

- IP/PPP and IP/Frame Relay over SONET on the cOCx/STMx line modules
- IP/ATM and IP/PPP over SONET on the OCx/STMx line modules

This support allows service providers to accept incoming optical connections or connect the system to the backbone network via optical connections. The ERX system SONET implementation supports termination, statistic gathering, alarm surveillance, and performance monitoring at the section, line, and path layers of a SONET interface. See Figure 3-14.

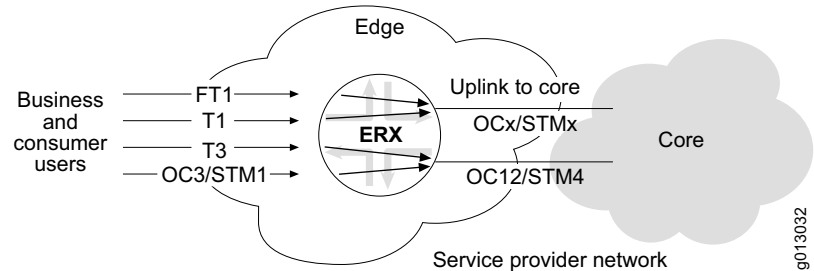


Figure 3-14 The ERX system supports SONET interfaces

SONET Features

The ERX system supports the following SONET features:

- Telcordia Technologies GR-253-CORE common generic criteria feature set
- Configurable SONET or SDH operational support
- SONET statistics collection
- Generation of remote defect indication (RDI) signals at the line and path layers for loss of signal (LOS), loss of frame (LOF), and loss of pointer (LOP); and alarm indication signals at the line and path (AIS-L and AIS-P) layers
- Signaling of remote error indication (REI) at the line and path layers
- Transmit clock configuration options for recovered receive clock (loop timing) or internal generation. The clock source can be pulled from the internal clock on the SRP module or any externally connected clock source. A common systemwide clock is supported, as well as configuration options for a different clock source per line module.
- Support for two loopback modes: *Line Loopback* to connect the received network signal directly to the transmit network signal line and *Internal Loopback* to connect the local transmitted signal to the local received signal

SONET Statistics

Statistics are gathered on the SONET section, line, and path interfaces in 15-minute intervals:

- Errored seconds – section: SEF defect, LOS defect, or coding violation (CV); line: seconds with AIS-L or CV; path: seconds with AIS-P, LOP defect, or CV
- Severely errored seconds – section: errored seconds with x CV, SEF defect, or LOS defect; line: errored seconds with x CV or AIS-L; path: errored seconds with x CV, AIS-P, or LOP defect
- Severely errored framing seconds – section: seconds with SEF defect
- Coding violations – section: BIP-8 errors (B1 byte); line: BIP-8*N errors, (N=STS-N; B2 bytes); path: BIP-8 errors, (B3 byte)
- Unavailable seconds – line: 10 consecutive line SESs; path: 10 consecutive path SESs

Switched Multimegabit Data Service (SMDS)

SMDS is a wide area networking service designed for LAN interconnection. SMDS is a connectionless service. An SMDS network is composed of:

- A series of SMDS switches inside a service provider's network
- A series of CSUs/DSUs that connect subscribers to the network
- Routers and gateways to connect to each CSU/DSU

SMDS Components

The following components support the SMDS application:

- HSSI line module – provides the physical interface to connect to the SMDS switch.
- GRE tunnels – transport SMDS traffic among ERX systems.
- SMDS trunk interface – runs over the HSSI line module and GRE tunnels.
- Connection-based forwarding (CBF) interfaces and connections – forward SMDS traffic among SMDS trunk interfaces.

Figure 3-15 shows how these components make up the interface columns for the SMDS application.

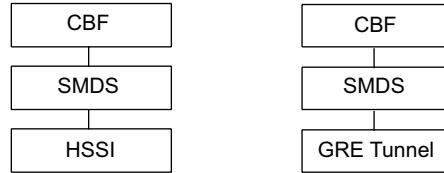


Figure 3-15 Interface columns for the SMDS application

Routing

The ERX system is a carrier-class router that fully supports both the interior and exterior IP-based routing protocols used by large service providers, including BGP-4, IS-IS, OSPF, RIP, VRRP, MPLS, and IP multicasting. This routing support allows service providers to deploy the system into existing network architectures and delivers full interoperability with older routing products.

Full routing is supported between all line interfaces in the system, including U-turn routing on a single port. All routing support integrates fully into the other features supported by the system, including layer 2 support and QoS capabilities, allowing a service provider to build leading-edge routing services for their subscribers.

Routing Features

There are three key features of the ERX system routing implementation:

- The system implements a fully distributed routing scheme that supports wire-speed packet forwarding for small packet sizes across all interfaces simultaneously.
- All Tier-1 ISP routing protocols are supported, including BGP-4, OSPF, IS-IS, RIP, and MPLS, and are highly scalable to address the needs of large networks.
- The system allows multiple virtual routers to be created, with separate route protocols and route tables supported on the different router instances.

The following are additional features of the routing implementation:

- Scalable IP routing with standards support for routing requirements for IPv4 – RFC 1812
- Support for emerging IPv6 requirements
- Virtual router support

- IP multicast support for DVMRP, IGMP, MBGP, PIM
- Access list, route map, and distribution list support for creating and applying routing filters
- Wire-speed performance forwarding with minimum packet sizes
- Load balancing to compute equal cost paths based on routing protocol information

ERX Routing Architecture

The ERX routing architecture is designed specifically to address the needs of large service provider routed networks. It has a highly scalable design; as more line modules are added to a chassis, more packet processing power is added. The system performs at wire speed, even with 40-byte packets and all IP features (QoS, VPNs, statistics) enabled. The routing architecture is distributed—each individual line module executes route lookups and forwarding decisions based on a common routing table. These two functions are described in detail below.

The ERX SRP module builds and stores a common IP route database for up to 1,500,000 routes. The information maintained for each route includes at least: destination IP address, prefix length, equal cost path calculations, route redistribution information, the protocol that added the route (static, RIP, OSPF, BGP-4), and attributes specific to that protocol (such as layer 1 vs. layer 2 and interior vs. exterior).

Routes in the routing table can be shared between routing protocols, and any routing protocol can add routes to the table. The level of interaction between the different routing protocols is configurable. The common routing table is stored centrally on the SRP module, but it is compressed and multicast downloaded to each individual line module.



Note: *If virtual routers are configured, each virtual router has its own logically distinct routing database.*

Figure 3-16 shows the route table creation and distribution process. The SRP module gathers the routing information and stores it in a centralized table. It then sends updates to each line module. Each line module maintains a complete routing table. This feature allows incoming packets to be processed quickly and efficiently in contrast to approaches that must rely on IP address or route caching.

The ERX route lookup and forwarding execution is completely distributed. When a packet is received on the line module, the route table lookup is performed entirely on that module with the routing table that is already located on the module. If the route is not in the routing table, the

destination is considered unreachable and an ICMP unreachable packet is sent. This distributed processing architecture eliminates the performance restrictions found in centralized approaches.

A route tag is prepended to the packet, which indicates the ERX egress port associated with the route, and the packet is sent to the switch fabric for transport to the egress port.

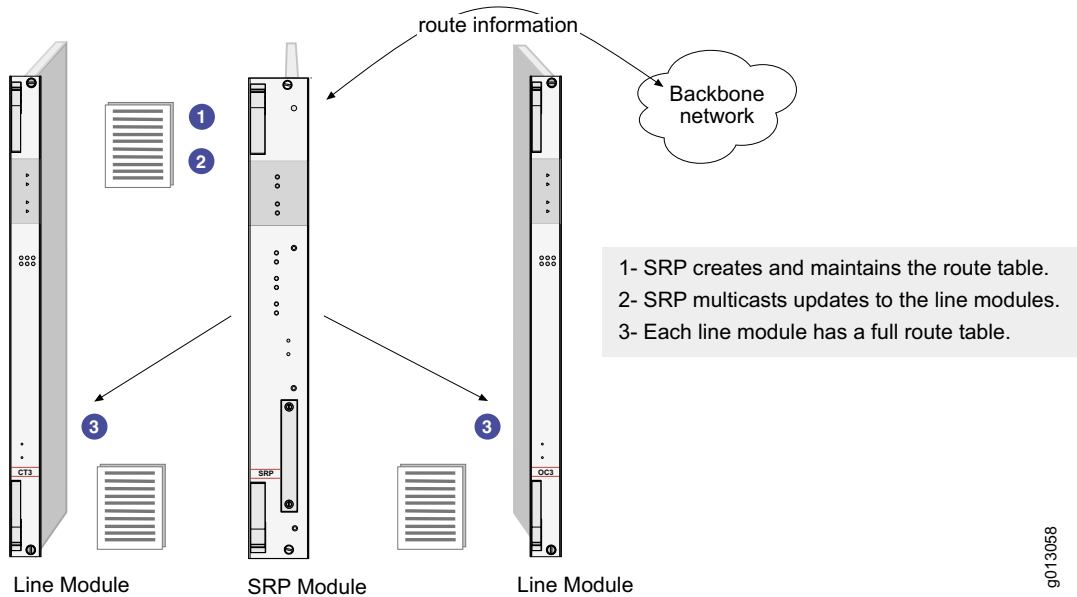


Figure 3-16 ERX system routing table distribution overview

Route Control

The ERX system supports a number of features that allow the service provider to control the exchange of routing information between virtual routers in the system, between routers in the network, and between protocols within a router:

- Access lists – Provide filters that can be applied to route maps or distribution lists. They allow policies to be created, such as those to prevent forwarding of specified routes between the BGP-4 and IS-IS routing tables.
- Route maps – Modify the characteristics of a route (generally to set its metric or to specify additional attributes) as it is transmitted or accepted by a router. Route maps may use access lists to identify the set of routes to modify.

- Distribution lists – Control the routing information that is accepted or transmitted to peer routers. Distribution lists always use access lists to identify routes for distribution. For example, distribution lists could use access lists to specify routes to advertise.
- Redistribute routes – Redistribution allows routes to be shared between routing protocols and routing domains. For example, a subset of BGP-4 routes could be leaked into the IS-IS routing tables.

Virtual Routers

The ERX system supports multiple distinct routers within a single system. This support allows service providers to configure multiple separate secure routers within a single chassis. Applications for this function include the creation of individual routers dedicated to wholesale customers and corporate VPN users, or routers dedicated to a specific traffic type. The system can support up to 1,000 virtual routers per module or per chassis.

A network device attaching to an ERX system sees a router interface and has no notion of the virtual router behind the interface. This makes the routing completely secure and protected as compared with other shared table models. For example, a physical Frame Relay link may have circuits that are connected to different virtual routers. The physical and data-link layers are not aware that there are multiple router instances. Each customer represented by a Frame Relay VC gets its own route table. In this manner, the system can seamlessly handle even overlapping addresses.

The ERX system implements the virtual routers by maintaining a separate instance of each data structure for each virtual router and allowing each protocol (TCP/UDP, RIP, OSPF, BGP-4, IS-IS) to be enabled on a case by case basis. There is a table of router interfaces to associate user connections (for example, PPP or Frame Relay) with one or more IP interfaces within a virtual router.

As mentioned above, the ERX routing protocol suite provides full support for BGP-4, IS-IS, OSPF, and RIP. These protocols can be enabled or disabled for each instance of a router in the ERX system. Each of these protocols is covered in detail in the following subsections. In addition, each virtual router can be independently managed with secure access between virtual routers.

BGP-4 Support

The ERX system supports a highly scalable and robust version of the Border Gateway Protocol (BGP-4), designed specifically for large-scale

networks. Large service providers typically use BGP-4 as the protocol of choice for interaction with external networks, such as peer and wholesale networks.

BGP Features

BGP-4 features include:

- BGP-4 – RFC 1771 – A Border Gateway Protocol 4 (BGP-4) (March 1995)
- RFC 1745 – BGP4/IDRP for IP—OSPF Interaction (December 1994)
- Communities – RFC 1997 – BGP Communities Attribute (August 1996)
- Community attributes – RFC 1998 – An Application of the BGP Community Attribute in Multi-home Routing (August 1996)
- Extended communities – draft RFC
- Confederations – RFC 1965 – Autonomous System Confederations for BGP (June 1996)
- Route reflectors – RFC 1966 – BGP Route Reflection An alternative to full mesh IBGP (June 1996)
- Route flap dampening – RFC 2439 – BGP Route Flap Damping (November 1998)
- Tier 1 ISP class route filtering, route mapping and attribute manipulation capabilities, as well as route redistribution
- Highly scalable BGP-4 architecture to support hundreds of BGP-4 peers and hundreds of thousands of routes. Scaling will increase with future software releases.
- BGP-4 neighbor setup
- Exterior BGP (EBGP) multihop to accept and attempt BGP-4 connections to external peers residing on networks that are not directly connected
- Next-hop self to disable the next-hop BGP-4 update route processing
- Soft-reconfiguration inbound to enable storing of received updates without regard to the inbound policy
- Update source to allow internal BGP (IBGP) sessions to use any interface for TCP connections
- Advertisement interval, which sets the interval between sending BGP-4 routing updates
- Shutdown to stop BGP-4
- Maximum prefix to restrict the number of prefixes that can be received from a neighbor

- MD5 authentication to support MD5 authentication on a TCP connection between two BGP-4 peers
- **network** command to specify BGP-4 route origin as IGP
- **no synchronization** command support to support network route advertisement without waiting for the IGP
- BGP-4 peer groups to add BGP-4 neighbors as members of a peer group
- BGP-4 aggregation – with/without AS sets, attribute modification during aggregation
- BGP-4 access list – AS-path list, community list, distance/BGP distance, distribute-list in/out, regular expressions (for AS-path list comparisons)
- BGP multiprotocol extensions – RFC 2858 – Multiprotocol Extensions for BGP-4 (June 2000)
- BGP/MPLS VPNs – virtual private networks using MPLS to carry data and BGP to carry routes and labels – RFC 2547 – BGP/MPLS VPNs (March 1999)
- Route map support
- Overlapping VPNs
- Automatic route-target filtering
- Support for automatically ending BGP sessions if the link to any adjacent external peer fails
- Table maps to apply routing policy to routes about to be added to the IP routing table
- Routes intended to be sent to peer group members stored in a single Adj-RIBS-Out table for the peer group rather than in tables for each peer group member
- Configurable redistribution of IBGP routes in addition to EBGP routes into IGP configured for BGP route redistribution

IS-IS Support

The ERX system supports a highly scalable version of the Intermediate System-to-Intermediate System (IS-IS) protocol, designed specifically for large-scale networks. IS-IS is a link state protocol. It is increasingly being used by large service providers as the interior routing protocol to share routing table information between routers in the same service provider network.

IS-IS Features

IS-IS features include:

- Integrated IS-IS – ISO 10589 standard and RFC 1195, use of OSI IS-IS for routing in TCP/IP and dual environments, and draft extensions to RFC 1195
- Level 1 and level 2 routing
- Internal optimization of route leaking from level 1 to level 2
- Highly scalable IS-IS architecture. The number of adjacencies and routes will be scaled higher with future software releases.
- Subnetwork-dependent functions to enable IS-IS transport over all media types supported by the interfaces
- Configuration control over the maximum number of addresses per area
- IS-IS computation for equal cost paths
- Adjacency and link state protocol (LSP) overrun to prevent other routers from using ERX paths in overrun states during route calculation
- Overload bit to force the system into overrun state
- ID field length configurable from 1-8 octets to support large-scale networks
- Router ID included in IS-IS hello PDUs for unnumbered point-to-point links
- Mesh groups to reduce the IS-IS protocol overhead within fully meshed networks. When configured in a mesh group, LSPs seen on one interface in a mesh group will not be flooded to another interface in the same mesh group. Alternatively, you can select a specific router in the mesh to be a server for all other routers within the mesh group.
- Periodic complete sequence number packet (CSNP) on point-to-point links to help mesh groups keep their databases synchronized to recover quickly from failed mesh matrices
- No IS-IS database purge on checksum error/ignore LSP error
- Checksum errored IS-IS LSPs allowed to be ignored rather than purging the LSPs
- Ability to configure up to three areas in one router
- Authentication within the IS-IS packet with PDU or MDU

- Selective route redistribution to control which routes should pass between the different databases (level 1 and level 2)
- Commands – Timers: IS-IS hello-interval per level, hello-multiplier per interface, LSP-interval per interface, retransmit-interval per interface, retransmit-throttle-interval per interface, SPF interval, LSP maximum life, summary address level 1, level 1-2, level 2, IS-IS metric, CLNS host, log-adjacency-changes, IS-IS mesh-group blocked
- Support for MPLS traffic engineering

OSPF Support

The ERX system fully supports the Open Shortest Path First (OSPF) routing protocol. This protocol is used by service providers to keep routing table information updated between the routers in a service provider network, and as a means to communicate with other POP-based network equipment. OSPF is an interior link state-based routing protocol.

OSPF Features

The ERX system implementation supports the following key features:

- OSPF Version 2 – RFC 2328 – OSPF Version 2 (April 1998)
- NSSA and Opaque LSA – RFC 1587 – The OSPF NSSA Option (March 1994) and RFC 2370 – The OSPF Opaque LSA Option (July 1998)
- Intra-area and interarea routes, with each area running as a separate network in order to reduce the size of the link state database in large networks
- Highly scalable OSPF architecture to support hundreds of OSPF adjacencies and tens of thousands of routes. The adjacencies and routes will be scaled higher with future software releases.
- Virtual link to create a logical link between two backbone area routers that are not physically adjacent, where the logical link tunnels through a nonbackbone area
- Nonbroadcast network, where a separate IP interface is defined for each neighbor, creating a series of separate point-to-point links
- Three types of authentication for protocol exchanges: null authentication, simple password authentication, and cryptographic authentication
- Opaque LSAs that will be accepted from other routers and be flooded accordingly

- Capability of new routes to be leaked into the routing table either by another routing protocol or through a static route
- Equal-cost multipath to insert all equal instances of available routes
- MD5 authentication
- Type 1 and Type 2 external routes
- Load sharing to balance loads across equal cost paths
- Control of the generation of type 7 default LSAs by NSSA border routers
- Adjacencies through unidirectional interfaces to remote neighbors that can be more than one hop away through intermediate routes that are not running OSPF
- Connection of OSPF domains via BGP/MPLS VPNs without creating OSPF adjacencies between the domains

RIP Support

The ERX system supports RIPv1 and RIPv2 routing protocols. Service providers use RIP to communicate with other POP-based networking equipment or, on subscriber access links, to communicate with CPE devices that are running RIP.

RIP Features

The system supports the following features:

- Compliance with RFC 2453 – RIP Version 2 (November, 1998)
- Backward compatibility with RIPv1
- Unicast messages to communicate with a RIP neighbor
- Adjacencies through unidirectional interfaces to remote neighbors that can be more than one hop away through intermediate routes that are not running RIP

VRRP

The ERX system supports the Virtual Router Redundancy Protocol (VRRP), which can prevent loss of network connectivity to end hosts if the static default IP gateway fails. When VRRP is implemented, a number of routers can be designated as “backup” routers in the event that the default “master” router fails.

In case of a failure, VRRP dynamically shifts the packet-forwarding responsibility to a backup router. A redundancy scheme is created by VRRP, which allows hosts to keep a single IP address for the default gateway, but maps the IP address to a well-known virtual MAC address. VRRP provides this redundancy without user intervention or additional configuration at the end hosts.

VRRP Features

The system supports the following features:

- Compliance with RFC 2338 – Virtual Router Redundancy Protocol (April 1998) and RFC 2787 – Definitions of Managed Objects for the Virtual Router Redundancy Protocol (March 2000)
- Supported on the FE-2 line module and the GE/FE line module
- Currently supports 200 virtual router IDs (VRID) on the FE-2 line module, and supports 800 VRIDs on the GE/FE line module
- Fully supports VLANs

IP Multicasting

The ERX system supports IP multicasting. IP multicasting improves network efficiency by allowing a host to transmit a datagram to a targeted group of receivers. For example, a host may want to send a large video clip to a group of selected recipients. It would be time-consuming for the host to unicast the datagram to each recipient individually. If the host broadcasts the video clip throughout the network, network resources are not available for other tasks. The host uses only the resources it needs by multicasting the datagram. A multicast datagram is effectively a broadcast to a limited group of network devices.

Routers use multicast routing algorithms to determine the best route and transmit multicast datagrams throughout the network. Table 3-2 lists the protocols the system supports and the function that each protocol provides.

Table 3-2 Function of multicast protocols on a router

Protocol	Function
Internet Group Membership Protocol (IGMP)	Discovers hosts that belong to multicast group.
Protocol Independent Multicast Protocol (PIM)	Discovers other multicast routers that should receive multicast packets.

Table 3-2 Function of multicast protocols on a router (continued)

Protocol	Function
Distance Vector Multicast Routing Protocol (DVMRP)	Routes multicast datagrams within autonomous systems.
BGP Multicasting Protocol	Routes multicast datagrams between autonomous systems.

The ERX system supports up to 16,384 multicast forwarding entries (multicast routes) at any time.

IGMP

IP hosts use IGMP to report their multicast group memberships to neighboring routers. Similarly, multicast routers, such as the ERX system, use IGMP to discover which of their hosts belong to multicast groups.

The IPv4 address scheme assigns Class D addresses for IP multicasting. IGMP is the protocol that uses these addresses, which can be in the range 224.0.0.0 to 239.255.255.255. The following addresses have specific functions or are unavailable:

- 224.0.0.0 is reserved and you cannot assign it to a group.
- 224.0.0.1 is the all-hosts address – a packet sent to this address reaches all hosts on a subnet.
- 224.0.0.2 is the all-routers address – a packet sent to this address reaches all routers on a subnet.

This implementation of IGMP complies with RFC 2236 – Internet Group Management Protocol, Version 2 (November 1997). IGMPv2 routers support both IGMPv1 and IGMPv2 hosts.

PIM

PIM is the protocol that allows multicast routers to identify other multicast routers that should receive packets. This implementation of PIM supports PIM dense mode (DM), PIM sparse mode (SM), and PIM sparse-dense mode (S-DM).

Figure 3-17 represents how PIM builds a source-group entry in a source based tree. The designated router (DR) receives data from the source on interface 1/0 and multicasts the data to its downstream neighbors on interfaces 1/1, 2/0, and 2/1. In the DR routing table, the entry for this operation lists the source as the IP address of the source and the group as the IP address of the multicast group.

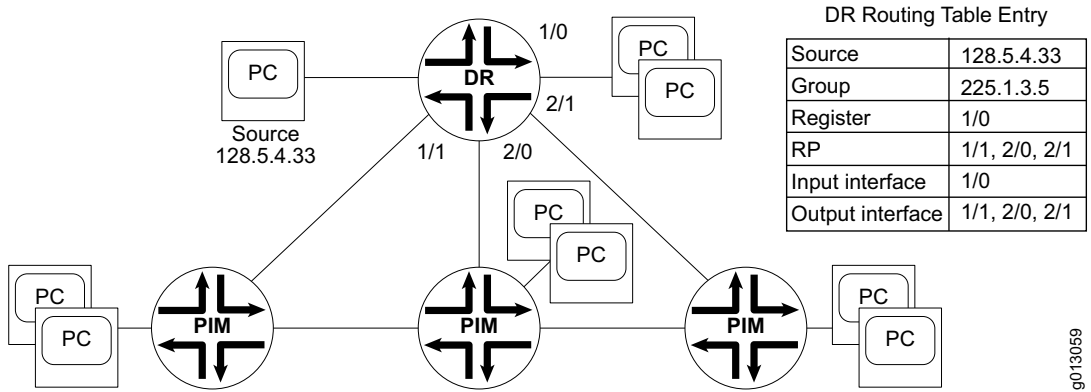


Figure 3-17 Source-routed tree

DVMRP

The ERX system supports Distance Vector Multicast Routing Protocol (DVMRP) on virtual routers to forward multicast datagrams through a network. DVMRP is an interior gateway protocol that supports operations within an autonomous system, but not between autonomous systems. The multicast backbone of the Internet, MBONE, uses DVMRP to forward multicast datagrams.

DVMRP is a dense mode multicasting protocol and therefore uses a broadcast and prune mechanism. The protocol builds a source-routed tree in a similar way to PIM-DM (see Figure 3-17). DVMRP routers flood datagrams to all interfaces except the one that provides the shortest unicast route to the source. To prevent unnecessary sending of multicast messages through the source-routed tree, DVMRP uses pruning. A DVMRP router sends prune messages to its neighbors if it discovers that:

- The network to which a host is attached has no active members of the multicast group.
- All neighbors, except the next-hop neighbor connected to the source, have pruned the source and the group.

When a neighbor receives a prune message from a DVMRP router, it removes that neighbor from its source group table, which provides information to the multicast forwarding table.

If a host on a previously pruned branch wants to join a multicast group, it sends an IGMP message to its first-hop router. The first-hop router then sends a graft message upstream.

BGP Multicasting

BGP multicasting is an extension of the BGP unicast routing protocol. Many of the functions available for BGP unicasting are also available for BGP multicasting.

The BGP multiprotocol extensions specify that BGP can exchange information within different types of *address families*. The address families available are unicast, multicast, and unicast vpn. When you enable BGP, the ERX system employs unicast IPv4 addresses by default. To enable BGP multicasting, you must configure commands for the multicast address family.

QoS

Quality of service is becoming more and more important for several reasons, including:

- Bandwidth oversubscription of the edge and the need to prioritize traffic; for example, business subscribers can be prioritized over residential users and mission-critical applications over best-effort data.
- Converged network infrastructure; that is, voice traffic is now carried over data networks.
- Service providers wanting to offer differentiated traffic classes to gain new revenue streams.

Bandwidth Oversubscription

In most service provider networks, the IP edge is oversubscribed in terms of bandwidth. Oversubscription at the edge calls for mechanisms to provide contractually agreed upon bandwidth guarantees, as well as to make sure that subscribers get at least a fair amount of the bandwidth that is available.

Therefore, the following features must be supported in the edge router to ensure traffic prioritization and true QoS at the edge:

- S-CBQ (subscriber class-based queuing), which allows a set of different traffic classes per IP interface and subscriber
- Per subscriber and per IP interface queuing; that is, a queue per traffic class and subscriber to support fairness
- Varied queuing algorithms such as (weighted round-robin) WRR and SPQ (strict-priority queuing) that can be attached to a physical port, subport level (for example, ATM PVC), and IP interface (subscriber)

- Wire-speed forwarding on all interfaces and all packet sizes; that is, the ability to perform route lookup and packet forwarding at line rate in all circumstances without dropping a single packet, even at packet sizes of 40 bytes
- Wire-speed classification of all packets on all interfaces—on the ingress *and* the egress interface. Packet classification is used to determine the traffic class to which an IP packet belongs and the service it gets.
- Wire-speed performance while applying policy decisions on forwarding, imposing rate limits, and reclassifying packets, such as marking and queuing packets

A high-aggregation edge router, such as the ERX system, must potentially support tens of thousands of queues for each line module, because each subscriber can have multiple queues. Figure 3-18 shows a single ingress queue being demultiplexed to many egress queues.

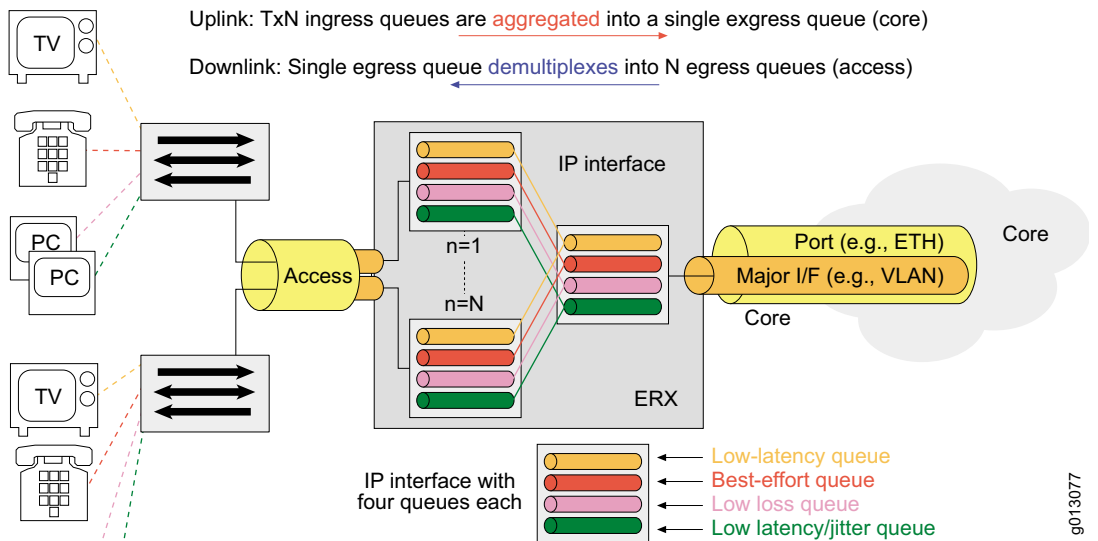
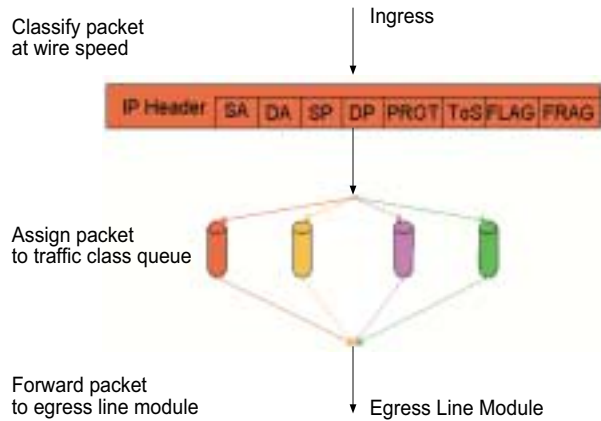


Figure 3-18 Interface queuing at the edge router

Implementing QoS

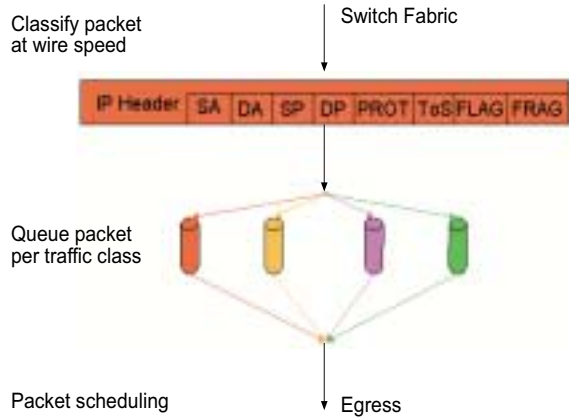
Figure 3-19 shows the common flow of an IP packet through the ERX system. The FPGAs and ASICs provide hardware assistance on the line modules for QoS. For example, hardware assists with color-based thresholding and a single queue per subscriber and IP interface.



g013011

Figure 3-19 Packet flow in the ERX system at ingress

Packets can be classified on the ingress and/or the egress line module. Policies such as rate-limit profiles can be applied to the result of the classification. Figure 3-20 shows packet flow in the ERX system at egress.



g013012

Figure 3-20 Packet flow in the ERX system at egress

Traffic Classes

In today's network environments, the edge router must ensure that real-time traffic is treated differently and is prioritized over all other traffic, because real-time applications, such as voice, are sensitive to delay and jitter.

Traffic classes that are emerging in service provider networks are:

- Low latency and low jitter for voice applications

- Low latency for streaming applications, such as video
- Low loss for mission-critical applications as well as signaling or VPNs

The ERX system supports up to eight traffic classes for each IP interface, including several low-latency classes, a low-loss class, and a best-effort class. A traffic class characterizes a specific QoS profile, such as low latency, in the chassis. In compliance with the DiffServ model, the ERX system accommodates the concept of expedited and assured forwarding.

Scheduler and QoS Profile

Egress queues are served by a three-level hierarchical scheduler, shown in Figure 3-21, that is running a WRR algorithm. This scheduler can be applied to:

- Layer 2 interfaces; for example, ATM PVC on 4-port OC3 (ATM) modules and VLAN on GE (first-level scheduler)
- IP interface (second-level scheduler)
- Traffic class or queue (third-level scheduler)

The schedulers can have a specific configuration for:

- Relative weight—how often a queue is selected relative to others, which determines minimum rate, or how much bandwidth is guaranteed
- Shaping rate—determines maximum rate, or the maximum burst size of the queue

Since the queue algorithms are intelligent, if a queue is empty, the relative weight is automatically recalculated and applied across the remaining queues of the same scheduler level.

The scheduler profile above is one part of the QoS profile. The other part is the queue profile, which describes the buffer allocated to a specific queue. Configurable parameters are:

- Queue length (buffer)
- Color threshold
- Traffic shaping

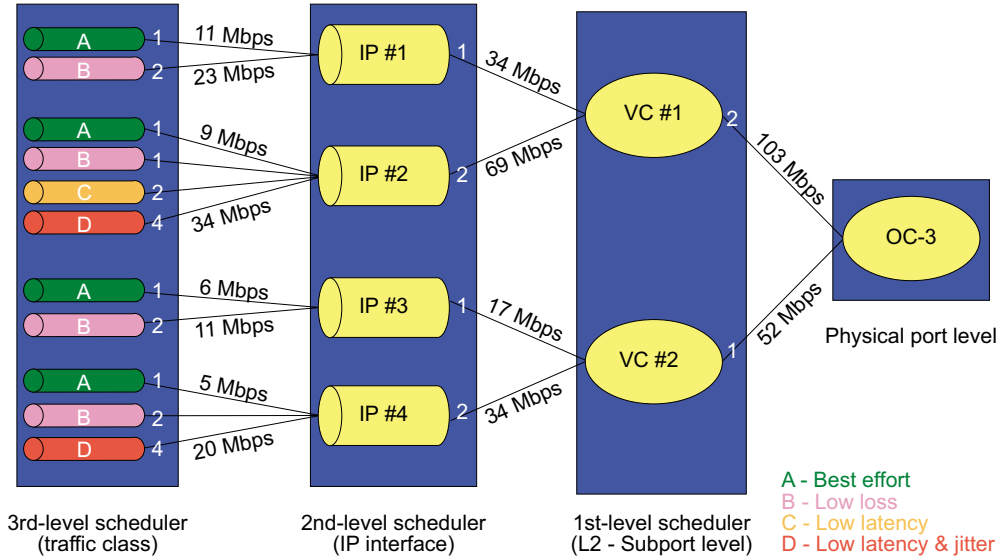


Figure 3-21 Three-level hierarchical scheduler

The QoS profile is attached to any interface, including:

- Serial, ATM, POS, Ethernet
- VC (ATM, Frame Relay), VLAN
- IP, L2TP, L2F, MPLS, tunnels
- PPP, PPPoE, ML-PPP, ML-Frame Relay

And more than one QoS profile can be configured per interface.

Queuing at the Edge

Because the ERX system is an edge router, it is specifically designed to handle the case where large numbers of ingress ports converge on a small set of egress ports. The ERX system is the only system that can protect subscriber traffic in this situation, because its QoS architecture provides the following:

- At least one fabric queue per ingress-egress line module pair.
- The WRR scheduler, which provides fairness among queues coming from different ingress line modules and destined for the same egress line modules.

- Different scheduler algorithms and weights (low-latency versus best-effort) that can be carried by each queue, representing a certain traffic class.

QoS Feature Summary

Table 3-3 lists the QoS features supported on the ERX system.

Table 3-3 QoS features supported on the ERX system

Feature	Offered
All QoS features mentioned earlier	Yes
Full packet classification	At wire speed
Rate limiting at wire speed	Per IP flow
Traffic shaping (per subscriber)	Per queue
Traffic classes (Expedited Forwarding/Assured Forwarding)	Up to 8
Queue assignment	Per subscriber and traffic class
Hardware support	Through ASICs and FPGAs
Number of queues per ASIC line module	48,000
Hierarchical schedulers	Three levels
Scheduling	WRR, SPQ

Application Scenarios

Real-Time Audio and Video

Figure 3-22 shows a real-time audio and video application using the ERX system. This application provides:

- Real-time video (TV) over DSL/FTTH plus Internet access
- TV (video) that is distributed via multicast and *not* affected by Internet access
- Two queues, including traffic shapers, per subscriber:
 - > Strict-priority queue (low latency traffic class); for example, shaped to 5 Mbps
 - > Best-effort queue
- Schedulers, which are attached to an IP interface, and traffic class

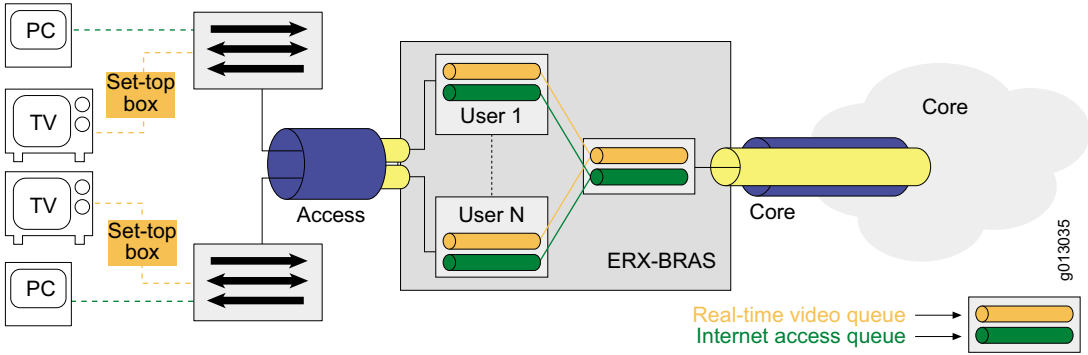


Figure 3-22 Real-time video and Internet access application

Business Versus Consumer Class Service

QoS policies on the ERX system can be used to provide different bandwidth service guarantees for different subscribers. In this application, the ERX system, as shown in Figure 3-23, is configured to have the following characteristics:

- Differentiates business users from residential users
- In case of congestion, gives business users a 100:1 priority over residential users
- Provides each subscriber with more than one queue (for example, strict-priority and best-effort queue)
- Attaches schedulers to subport level (for example, VC), IP interface, and traffic class
- Provides equal scheduler weight at IP interface and traffic class (queue) level. At subport level the queues of business customers are selected 100 times (100:1) more often than the queues of residential customers.

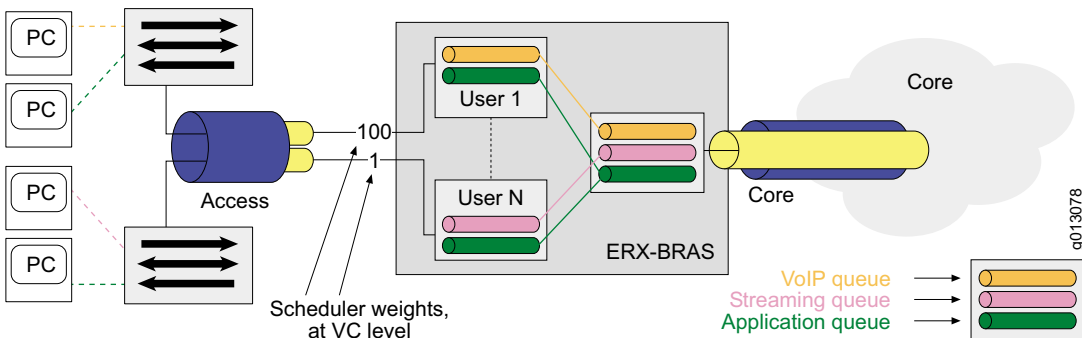


Figure 3-23 Business customers (User 1) versus residential customers (User N)

Media Gateway Aggregation

Figure 3-24 shows packetized voice over routed networks, which have the following characteristics:

- Media gateway (M-GW) with packetized voice (VoIP-GW) plus dial-up traffic (NB-RAS)
- ERX edge router aggregating multiple media gateways
- $N > 3$ queues per media gateway (for example, OC-3 ports)
- Strict-priority queue for VoIP payload (low latency and jitter) plus traffic shaper (100 Mbps)
- Low-loss queue for VoIP signaling and VPN traffic (49 Mbps)
- Best-effort queue (1 Mbps)
- Scheduler nodes attached to the port level, with queues carrying different weights

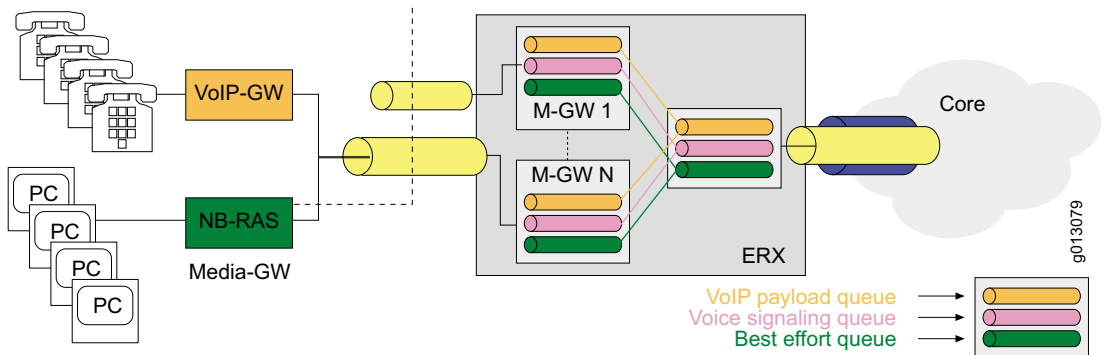


Figure 3-24 Packetized voice over routed networks

Policy Management

Policy management allows service providers to implement packet forwarding and routing specifically tailored to their customers' requirements. Using policy management, customers can implement policies that selectively affect packets.

Policy management provides the following types of services:

- Policy routing – directs packet flow to a destination port without a routing table lookup

- Packet filtering – marks or drops packets conforming to a classifier control list
- Packet forwarding – forwards packets that match a policy list
- Packet logging – allows you to log a classified packet flow
- QoS marking – stamping or labelling packets to conform to ToS, ATM, or MPLS classes
- Rate limiting – enforces line rates below the physical port line rate and provides rate limiting on a packet flow conforming to a classifier control list. Packets that do not fit the profile can be marked or dropped.
- RADIUS policy definitions – allows you to configure a policy that consists of *Filter/Forward* rules based on classified packet flows
- Traffic shaping – allows you to queue packets and transmit them at a specified rate

Policy Lists

The main tool for implementing policy management is the *policy list*. A policy list is a set of *rules* that can affect a packet. A rule is a policy command optionally combined with a *classification*.

Classifier Control Lists

A classification is specified by a *classifier control list*. Classifier control lists are used in classifying packets arriving on an interface so that different actions can be applied, based on which classifier control list the packet matches.

A classifier control list defines the values of IP packet header fields used for classification. The system applies a list of mask and match clauses to the IP header of a packet. The system supports 512 unique classification ranges per policy list.

The ERX system supports the following types of classifier lists:

- Precedence – matches the value of the precedence bits in the ToS byte of the IP packet header
- Protocol number – matches a protocol to a protocol number
- IP – matches IP protocol attributes such as source and destination IP address and IP mask
- ICMP – matches ICMP attributes such as source and destination IP address and IP mask, ICMP type and code

- IGMP – matches IGMP attributes such as source and destination IP address and IP mask, and IGMP type
- MPLS EXP – matches the EXP field (bits) of an MPLS header to classify MPLS frames and map them onto corresponding traffic classes. The assignment of EXP bits to traffic classes is configurable in the ERX system. Each LSP can have its own set of queues reflecting different traffic classes.
- TCP – matches TCP attributes such as source and destination IP address and IP mask, and source and destination TCP operator and port
- UDP – matches UDP protocol attributes such as source and destination IP address and IP mask, and source and destination UDP operator and port

Rules

Rules are created when a user specifies a policy command or combines policy commands and classifier control lists, as in a *rate-limit profile*. These rules become part of a policy list that is attached to an interface as either an input or output policy. The ERX system implements the rules in the policy list associated with the interface.

Figure 3-25 shows an example of how a rate-limit policy list is constructed.

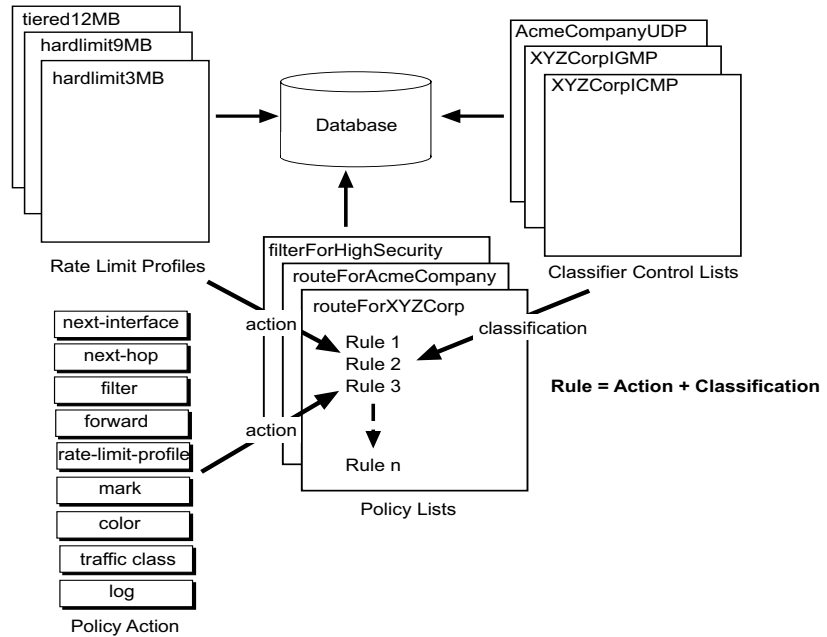


Figure 3-25 Constructing a policy list

Rate-Limit Profiles

Rate limiting is the process of limiting either a classified packet flow or source interface at a configured rate that is less than the physical rate on the port. When a policy list is created, a rule can be created that has a rate limit for an action.

A rate-limit profile is a set of bandwidth attributes and associated actions. You can create one-rate or two-rate rate limit profiles. One-rate rate limit profiles provides a hard-limit rate limiter or a TCP-friendly rate limiter mechanism, which provides better output performance for the bursty packet loss behavior of TCP traffic. Two-rate limit profiles provide a two-rate, three-color marking mechanism.

Dynamic Interfaces

Dynamic interfaces are created through some external event, typically through the receipt of data over a lower-layer link, such as an ATM virtual circuit (VC). In contrast, each layer of a *static interface* is created and configured through an existing configuration mechanism such as CLI or SNMP. Dynamic interface layers are created based on the packets received on the link and can be configured through:

- RADIUS authentication
- Profiles
- A combination of RADIUS authentication and profiles

Unlike static interfaces, dynamic interfaces are not restored through nonvolatile storage (NVS) after a reboot.

The ERX system software currently supports the following types of dynamic interfaces:

- IP over ATM (IPoA)
- IP over PPP over ATM
- IP over PPPoE over ATM
- IP over bridged Ethernet over ATM

Profiles

Dynamic interfaces can also be configured by profiles. A *profile* is a set of characteristics that act as a pattern. This pattern can then be dynamically assigned to interfaces. By using a profile, you reduce the management of a large number of interfaces by applying a set of characteristics to multiple interfaces.

When you are configuring a large number of interfaces with the same attributes at the higher layers, you can use a profile to apply all the common attributes to each layer. This action comprises one or more dynamic layers of the interface column. Once the static lower layers are defined, a profile is assigned to the highest static layer of the interface column.

You define profiles using CLI commands similar to the ones used to configure static interfaces. Profiles can be configured for IP, PPP, or PPPoE interfaces.

B-RAS Support

A separate application for the ERX system aggregates the output from DSLAMs, provides PPP session termination, enforces QoS policies, and routes traffic into the backbone. This application is called Broadband Remote Access Server (B-RAS).

For service providers to deploy xDSL services economically, they reuse OSS structures that are already in place for dial service offerings.

Therefore, the system supports a number of dial access-oriented protocols to accommodate this application.

B-RAS Features

The system features that are specifically supported for B-RAS include:

- RFC 2131 – Dynamic Host Configuration Protocol (March 1997)—where the ERX system is a DHCP proxy client
- Enhanced PPP:
 - > RFC 1877 – PPP Internet Protocol Control Protocol Extensions for Name Server Addresses (December 1995)
 - > RFC 1661 – The Point-to-Point Protocol (PPP) (July 1994)
 - > RFC 1662 – PPP in HDLC-like Framing (July 1994)
 - > RFC 1332 – The PPP Internet Protocol Control Protocol (IPCP) (May 1992)
- Password Authentication Protocol (PAP) – RFC 1472 – The Definitions of Managed Objects for the Security Protocols of the Point-to-Point Protocol (June 1993)
- Challenge Handshake Authentication Protocol (CHAP) – RFC 1472 – The Definitions of Managed Objects for the Security Protocols of the Point-to-Point Protocol (June 1993)
- Domain parsing based on destination domain (for example, @mu.edu, @company1, @isp1)
- IP address pooling
- PPP/Ethernet emerging specification
- Support for 48,000 IP interfaces in a single system
- L2TP access concentrator (LAC) and L2TP network server (LNS) capabilities to encapsulate layer 2 packets for transmission across the network via tunnels

See also the following RADIUS RFCs:

- RFC 2865 – Remote Authentication Dial In User Service (RADIUS) (June 2000)—where the ERX system is an authentication client
- RFC 2866 – RADIUS Accounting (June 2000)—to support accounting
- RFC 2867 – RADIUS Accounting Modifications for Tunnel Protocol Support (June 2000)

B-RAS Protocol Support

Figure 3-26 shows a number of protocols the system supports for B-RAS applications. These include:

- IP/PPP/ATM
- IP/PPP/Ethernet/ATM
- IP/PPP/Frame Relay
- IP/PPP/Ethernet/Frame Relay

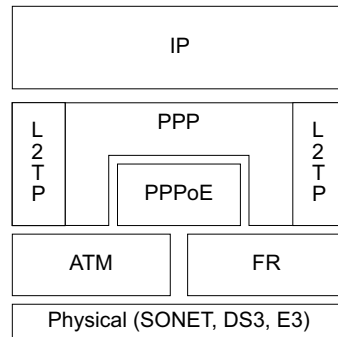


Figure 3-26 B-RAS protocol support

This protocol support allows service providers to install the ERX system as a central site router that supports the encapsulation schemes from a wide variety of CPE xDSL devices. The signal support is provided over the OC3, T3, and E3 modules.

B-RAS Data Flow

The system must accomplish several tasks for subscribers to establish their connections:

- The subscriber must be authenticated. The system can use the RADIUS authentication with PAP and CHAP to allow the subscriber to be verified.
- The connection must be assigned an IP address. The system supports IP address assignment for the end-user through DHCP, local IP pools, and RADIUS.
- The PPP session must be terminated. The system supports all varieties of xDSL encapsulation schemes to support all types of xDSL modems.
- Subscribers are matched to their routing and/or QoS policy. The system can match subscribers via the ERX classification features or via RADIUS attributes. Either of these methods can associate the

destination route and the level of service quality with the subscriber session. For example, with RADIUS, session timeout and idle timeout values can be associated with the subscriber session to support oversubscription by the service provider; or, to accommodate wholesaling applications, subscriber sessions can be directed to a specific virtual router.

- Accounting data must be collected. The system supports RADIUS accounting to allow the service provider to bill users based on call duration.

The ERX system offers a contrast to current solutions that require multiple separate network elements to provide B-RAS service. With comprehensive support for B-RAS features combined with the carrier-class product design and strong routing protocol support, the system gives service providers a single network element that can terminate large numbers of xDSL sessions, authenticate them, apply QoS policies, and route them into the core network.

Autodetection

The ERX system software performs a process called *autodetection*, also referred to as *autosensing*, to determine the layers of each dynamic interface. Autodetection is done by conditionally constructing interface layers based on the encapsulation type of the incoming packet.

Unlike static interfaces, which always allocate system resources on creation, autodetection uses system resources only on demand based on what is detected in the incoming packet. Static interfaces always consume system resources, even when the interface is quiescent. Dynamic interfaces, however, are created as a result of traffic on the interface. Dynamic interfaces may also be dynamically deleted without your intervention, thereby allowing any consumed system resources to be returned.

Zero-Touch Configuration

The combination of autodetection and autoconfiguration constitutes the zero-touch configuration capability, also called zero-touch provisioning. Zero-touch configuration enables the ERX interface to be configured without operator involvement. Subscriber connections are provisioned by:

- Autodetection of the incoming protocol
- Autoassignment of the service profile via the ERX system or RADIUS
- Autocreation of the IP interfaces

All this is achieved independent of the protocol type, allowing for streamlined operation and cost savings.

RADIUS Server Support

RADIUS is a distributed client/server system that protects networks against unauthorized access. RADIUS provides the following two functions, both supported by the ERX system:

- It authenticates operators who want to make changes to the network.
- It performs AAA (authentication, authorization, and accounting) services for subscribers who want to join the network and send traffic:
 - > Authentication – determines that a user may access a specific service or resource
 - > Authorization – associates connection attributes or characteristics with a specific user
 - > Accounting – tracks service use by subscribers

RADIUS clients running on an ERX system send authentication requests to a central RADIUS server that contains all the required user authentication and network access information. The RADIUS server is configured and managed by a RADIUS administrator. When a packet is received, the authenticating interface (either PPP or ATM 1483) establishes a session with RADIUS and passes the RADIUS server the username and password. ATM 1483 interfaces may receive configuration data from the RADIUS server in the form of *traffic-shaping* parameters.

Dynamic interfaces can be configured via RADIUS authentication. For dynamic IPoA, the RADIUS username and password are obtained from the information specified by a CLI command. The RADIUS authentication server returns a grant or deny indication. If authentication is granted, the RADIUS attributes are returned, a user login is created, and the dynamic interfaces are configured from the RADIUS attributes.

Any changes made to a RADIUS configuration for a given dynamic interface do not take effect until an existing dynamic interface configured from this RADIUS entry is recreated—that is, deleted and then dynamically created.

The RADIUS client in the ERX system uses the router ID of the virtual router as the source IP address regardless of the interface from which RADIUS packets are sent.

Basic Client-to-Server Interaction

The ERX system can support multiple RADIUS server targets. The order in which a server is queried determines which servers the system submits requests to on behalf of clients. The first configured authentication or accounting server is the primary authentication or accounting server, the next the secondary, and so on.

The ERX system offers two options by which servers are accessed:

- Direct – treats the first configured server as primary for all users
- Round-robin – treats the first configured server as a primary for the first request, the second server configured as primary for the second request, and so on. When the system reaches the end of the list of servers, it starts again at the top of the list.

The ERX system allows the service provider to specify configuration parameters for managing RADIUS client-to-server interaction:

- Maximum number of times the system retransmits a RADIUS packet to an authentication or accounting server
- Interval (in seconds) before the system retransmits a RADIUS packet to an authentication or accounting server
- Maximum number of outstanding requests allowed to a RADIUS server
- Amount of time (in minutes) that a RADIUS server is marked as unavailable if a request times out for the configured retry count
- Accounting interval between updates
- Algorithm for RADIUS server access (direct or round-robin)
- Encryption of the primary, secondary, and tertiary RADIUS authentication server secret
- Enabling of the duplicate accounting records feature to specify if duplicate records are sent to the RADIUS accounting server for a virtual router
- Collection of RADIUS statistics

The interaction between the ERX system and the RADIUS server is powerful and can be used by the service provider to centralize all subscriber profiles. For example, the RADIUS server can return all information required to start the subscriber session, including IP address, service profile, and authorization. This interaction streamlines operations across multiple ERX systems.

L2TP

Related to B-RAS is the Layer Two Tunneling Protocol (L2TP). L2TP allows you to encapsulate layer 2 packets, such as PPP, for transmission across a network. In an L2TP relationship, an L2TP access concentrator (LAC) forms a client-server relationship with a destination, known as an L2TP network server (LNS), on a remote network. L2TP supports an LAC to LNS tunnel using an ATM or Ethernet interface.

L2F

Layer Two Forwarding (L2F) provides a method for virtual dial-up service over the Internet. The traditional method for a remote user to access a company's network is through remote access equipment that is directly attached to the corporate network. This method requires a significant investment in equipment and support in addition to the cost of telephone charges for remote workers calling into the access equipment.

By employing L2F, a service provider can provide local access for remote workers and forward their data traffic through a tunnel to the corporate network. This method allows a company to outsource the investment in remote access equipment to the service provider, while retaining full control over access to the corporate network. In particular, L2F allows leveraging multiple protocols and private addressing across the existing Internet infrastructure.

In the Juniper Networks implementation of L2F, the ERX system, configured as a network access server (NAS), receives packets from a remote client and forwards them through a tunnel to a home gateway on a remote network.

MPLS Support

The ERX system is designed to support the powerful, new Multiprotocol Label Switching (MPLS) protocol as an integrated IP technology for improving traffic engineering, adding VPN functions, and scaling capabilities for IP networks. We are closely following the work on MPLS and is an active participant in the IETF and external MPLS user forums.

MPLS specifies a means to generate and attach a "label" to individual packets. This label, rather than the information contained in the IP header (such as the source or destination address), is then used to make forwarding decisions about the packet. This process is a departure from current routing practice, where every router node in the network handles an IP lookup to make a forwarding decision, resulting in a hop-by-hop

approach to packet control. The IP packets are labelled at the edge of the network by the ERX edge router. The rest of the network then relies on the label to forward the packet over an appropriate path. As the packet leaves the network, the MPLS label is removed.

Service providers can use MPLS to set up sophisticated subscriber-based or traffic-based policies at the network edge, while easing the lookup and transport requirements on the backbone. Service providers can use the ERX system to perform a multifield search on the IP header to determine the type of MPLS label to apply. Once the label is applied, the rest of the network can use the labels to handle lookup, rather than performing lookups against the entire route table.

The system supports MPLS for both traffic engineering and scalable VPNs.

MPLS Features

The ERX system supports the following MPLS features:

- LDP, CR-LDP, or RSVP-TE
- ATM, POS, Gigabit Ethernet, and Fast Ethernet
- Integration with extended IGP protocols (OSPF, IS-IS) to support traffic engineering via constraint-based explicit routes
- Label edge routing (initiate/terminate LSP) or core LSR (forward only based on label)
- Label stacking
- CAC for bandwidth accounting and service enforcement
- Downstream-on-demand, ordered control label distribution and downstream-unsolicited, independent control label distribution
- Topology-driven LSPs for LDP hop-by-hop MPLS networks, which enable the LDP to learn all IGP routes from IP and automatically create LSPs for all routes, resulting in a fully meshed MPLS network
- Label filtering for topology-driven mode, which determines whether and where incoming labels are distributed
- Differentiated services support by setting EXP bits corresponding to matched traffic class/color and setting traffic class/color corresponding to matched EXP bits
- BGP and MPLS VPNs, which support RFC 2547
- Inter-AS services, which support VPNs that cross AS boundaries

- Martini draft support for layer 2 services, such as Frame Relay and Ethernet/VLAN, over MPLS

MPLS Applications

MPLS supports the following applications:

- Traffic engineering
- VPNs (including MBGP support)
- QoS/CoS
 - > Classifier association of traffic with LSP
 - > Ability of all LSPs to make use of QoS features

BGP/MPLS VPNs

The ERX system supports the BGP multiprotocol extensions, which enable BGP to exchange routing information for different types of address families. The VPN-IPv4 address family enables the outsourcing of IP backbone services via the configuration of virtual private networks across the IP backbone. BGP/MPLS VPNs are scalable and flexible and enable service providers to offer value-added services. BGP carries routing information for the network and MPLS labels, while MPLS transports the data traffic. Figure 3-27 shows a typical scenario.

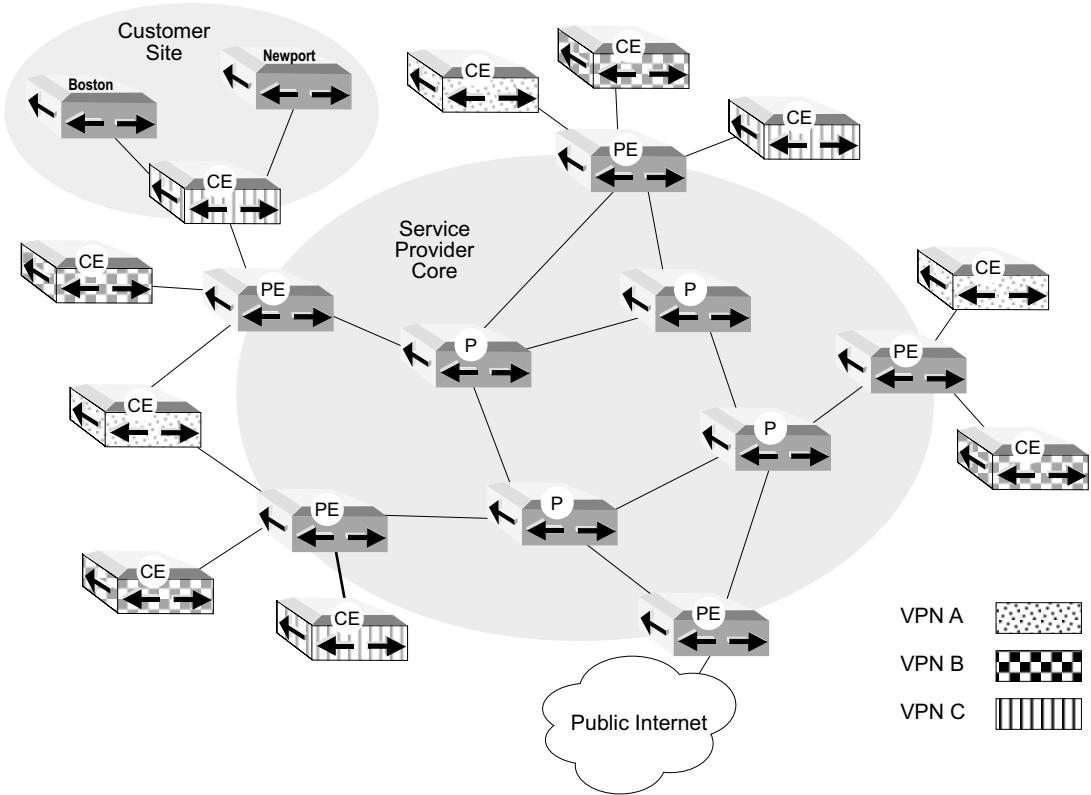


Figure 3-27 BGP/MPLS VPN scenario

The backbone through the service provider core comprises two types of routers:

- Provider edge routers (PEs) sit at the edge of the service provider core and connect directly to customer sites. These routers must run BGP-4, including the BGP/MPLS VPN extensions. They must also be able to originate and terminate MPLS LSPs.
- Provider core routers (Ps) connect directly to PEs or other Ps and do not connect directly to customer sites. These routers must be able to switch MPLS LSPs. It is not necessary to run BGP-4 on the P routers to be able to exchange routing information for VPNs. The P routes do not need to possess any information about customer sites.

PEs communicate with customer sites via a direct connection to a customer edge (CE) device that sits at the edge of the customer site. The CE can be a single host, a switch, or a router. If the CE is a router, it is a routing peer of all directly connected PEs, but it is not a routing peer of CEs at any other site. The link between the CE and the PE can employ any

type of encapsulation. It is not necessary to use MPLS. In Figure 3-27, each PE connects to multiple CEs and at least one P. Although only one customer site is shown, each CE lies within a customer site.

Each customer site is a network that can communicate with other networks in the same VPN. A customer site can belong to more than one VPN. Two sites can exchange IP packets with each other only if they have at least one VPN in common.

Each customer site that is connected to a particular PE is also associated with a forwarding table, known as a VPN routing and forwarding table (VRF), as shown in Figure 3-28.

You assign one or more (sub)interfaces to a given VRF. If multiple customer sites are members of the same set of VPNs, they can share a VRF. The system looks up a packet's destination in the VRF associated with the interface on which the packet is received. The VRFs are populated by BGP as it learns routes from the VPN. If a customer site is a member of multiple VPNs, the routes learned from all those VPNs populate the VRF associated with the site.

VRFs exist within the context of a virtual router (VR). A given virtual router can have zero or more VRFs, in addition to its global routing table (which is not associated with any VPN, CE, or customer site). An ERX system can support up to 1,000 forwarding tables; that is, up to a total of 1,000 VRs and VRFs per module or per chassis.

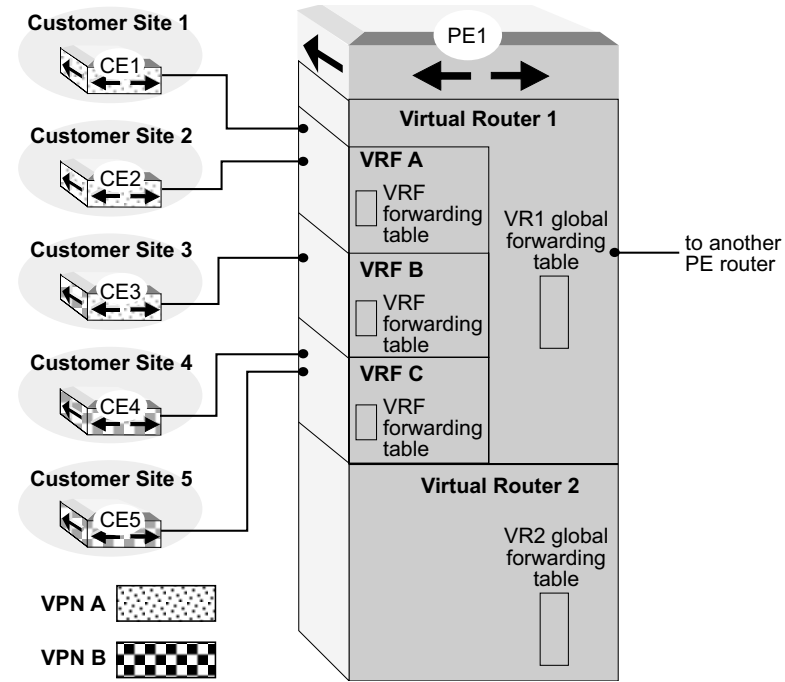


Figure 3-28 BGP/MPLS VPN components

Encapsulating Layer 2 Services

Many Internet service providers are concerned with the expense and implementation time to transition from mature layer 2 networks, such as Frame Relay or ATM, to optical networks based on IP and/or MPLS. By emulating circuits or services, service providers can use their existing equipment base and move to a packet-based optical infrastructure. Emulation allows existing circuits and/or services to be carried across the new infrastructure without your having to employ new services, thus allowing the internetworking of disparate networks.

Encapsulating layer 2 services such as Frame Relay in GRE tunnels leverages use of the public Internet while continuing to provide end-to-end Frame Relay congestion control and management plane functionality that Frame Relay users expect. By encapsulating all traffic received for a given Frame Relay data-link connection identifier (DLCI) into a GRE tunnel, a service provider must own both endpoints of the tunnel. If both ends are ERX systems, intervendor compatibility of a Network-to-Network Interface (NNI) is not a concern.

Layer 2 services over IP/MPLS allow service providers to emulate the legacy layer 2 network links and services over their IP/MPLS-based

network. Layer 2 services over MPLS are especially desirable because MPLS provides features such as traffic engineering and congestion control. These MPLS features can be used to emulate certain layer 2 service characteristics.

DHCP Local Server

You can configure the system to allow remote access to non-PPP clients. This configuration is particularly useful for broadband (cable and DSL) environments or environments that use bridged Ethernet.

Using PPP in these environments requires a PPPoE client for each subscriber's computer. Using the embedded DHCP local server in the ERX system for these environments requires no additional software for subscribers' computers, because the system provides IP addresses to the computers. It is easier for network operators to support a central system than to maintain software on subscribers' computers.

The ERX system offers an embedded DHCP server, known as the DHCP local server. The DHCP local server has two modes: equal-access and standalone.

- In equal-access mode, the DHCP local server works with the Service Deployment System (SDX) and authentication, authorization, accounting, and address assignment (AAAA) utility to provide an advanced subscriber configuration and management service.
- In standalone mode, the DHCP local server provides a basic DHCP service.

Equal-Access Mode

In equal-access mode, the main purpose of the system is to enable access to non-PPP users. The DHCP local server performs the following functions in equal-access mode:

- Assigns a temporary IP address, which enables the subscriber to access the SDX or the HTTP local server.
- Communicates with the SDX or the HTTP local server.
- Communicates with the RADIUS server.
- Supports RADIUS accounting attributes. This feature allows you to use RADIUS start and stop attributes to track user events such as the lifetime of an IP address.

- Assigns an IP address with a long lease time, which allows the subscriber to access services.
- Provides the relay of DHCP requests from cable modems to DHCP servers that support DHCP requests from cable modems.

SDX Application

The SDX application provides a Web-based interface that allows subscribers to access services, such as the Internet, an intranet, or an extranet. For more information about the SDX product, see *Chapter 4, Element and Network Management*.

HTTP Local Server

The system offers an embedded Web server, known as the HTTP local server. You can configure one HTTP local server per virtual router.

The sole purpose of the HTTP local server is to allow user login and authentication without the SDX application.

Local Pool Selection and Address Allocation

In equal-access mode, clients are authenticated by AAA. For authenticated clients, the DHCP local server selects a DHCP pool from which to allocate an address using the framed IP address, pool name, domain name, and gateway address (giaddr) parameters.

Linking Local Address Pools In equal access mode, you can link a DHCP local pool to another local pool. The linked pool serves as a backup pool. If there are no addresses available in a pool, the DHCP local server attempts to allocate an address from the linked pool.

Cable Modem DHCP Relay

The DHCP local server does not support DHCP requests from cable modems. However, you can set up the DHCP local server to relay DHCP packets that originate from a cable modem to an external DHCP server that can provide the cable modem with the configuration it requests.

Standalone Mode

In standalone mode, the DHCP local server operates as a basic DHCP server. Clients are not authenticated and standalone mode does not use token addresses. The DHCP local server receives DHCP client requests for addresses, selects DHCP local pools from which to allocate addresses,

distributes public addresses to the clients, and maintains the resulting DHCP bindings in a server management table.

Local Pool Selection and Address Allocation

In standalone mode, the DHCP local server selects a pool to allocate a public address for a client by matching the local pool's network address to the giaddr or the received interface IP address.

VLAN Support

The ERX system supports VLANs (IEEE 802.1q) on Fast Ethernet and Gigabit Ethernet interfaces. IEEE 802.1q describes two key Ethernet fields that have been added to the Ethernet frame to identify and prioritize traffic: VLAN ID (virtual LAN identifier) and user priority values (802.1p).

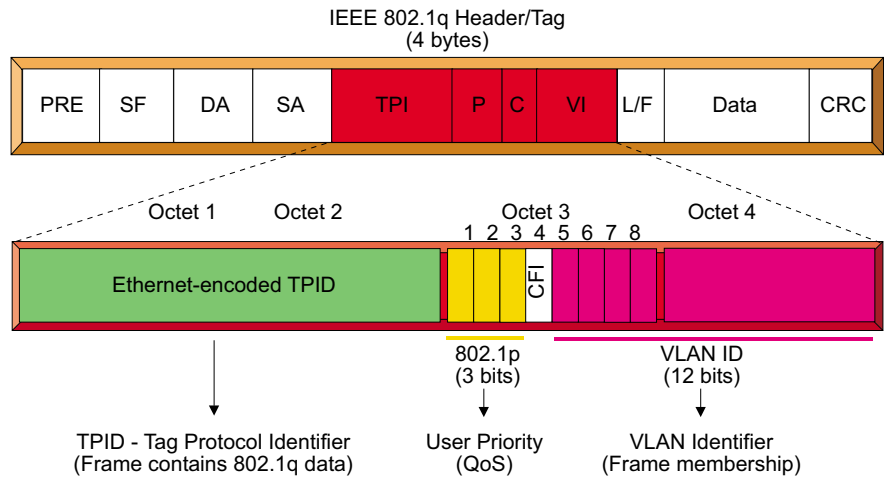


Figure 3-29 Ethernet frame with 802.1q fields (VLAN)

Ethernet frames using this standard can be tagged and viewed in the same way that ATM VCs (VPI/VCI) are represented on an ATM circuit. When an Ethernet frame is sent over a shared medium (for example, a GE line), it is logically defined as belonging to a specific virtual LAN and provisioned in the same way as a PVC on an ATM circuit. The Tag Protocol Identifier (TPID) in the Ethernet header indicates that this Ethernet frame contains 802.1q data. The VLAN ID indicates the virtual

LAN (virtual circuit) to which this frame belongs. The three bits in the 802.1p field can be used to determine the priority of this frame.

On a given physical link, Ethernet frames may be tagged or untagged. The VLAN ID in the 802.1q tag explicitly identifies the VLAN for tagged frames; the ID may be 0. Untagged frames are implicitly in VLAN 0, which is considered to be the default VLAN. This implementation is different from ATM, where all frames are required to have an explicit VPI/VCI.

The VLAN ID field is 12 bits, allowing 4096 VLAN IDs. Tagged frames are identified by the value 0x8100 in the Ethernet protocol type field; the actual protocol type of the frame then follows the tag. Some attached devices may not accept 802.1q-tagged frames, and therefore can reside only in VLAN 0.

Conversely, some devices may accept only tagged frames, requiring that even frames in VLAN 0 be tagged. Ports that have been explicitly added to VLAN 0 will tag frames unless the untagged option is specified. Ports that do not have VLANs configured will not tag frames.

On ingress to a VLAN port, the Ethernet protocol type field of a frame is examined to determine if it is tagged. If so, the VLAN ID is extracted from the tag, and the actual protocol type field following the tag is used to identify the protocol. If not, the frame is in the default VLAN, and the type field is used to identify the protocol.

On egress, a frame destined for the default VLAN may be sent untagged or tagged (depending on whether the untagged option was specified). A frame destined for any other VLAN is sent tagged.

S-VLAN

As specified in IEEE Standard 802.1q, the twelve-bit VLAN identifier's tagged frames enables the construction of a maximum of 4,096 distinct VLANs in an Ethernet B-RAS application environment; however, this VLAN limit is inadequate. A stacked VLAN (S-VLAN) provides a two-level VLAN tag structure, which extends the VLAN ID space to over 16 million VLANs.

Creating an S-VLAN requires the use of a second encapsulation tag. The ERX system performs decapsulation twice, once to get the S-VLAN tag and once to get the VLAN tag. Using this "double tagging" approach allows for over 16 million address possibilities, which more than satisfies the scaling requirement for Ethernet B-RAS applications.

VLAN and S-VLAN subinterfaces can coexist over the same VLAN major interface. S-VLANs and VLANs are configured similarly.

Interface Modes

Ethernet interfaces are configured to operate in one of the following modes, depending on the combination of protocol and VLAN usage:

- Single protocol non-VLAN (that is, IP only)
 The Ethernet interface supports only one upper-layer protocol interface (without VLANs). Such support is backward compatible with the initial implementation of Ethernet interfaces.
- Single protocol VLAN (that is, VLANs and IP)
 The Ethernet interface supports a single upper-layer protocol on each of one or more VLANs.
- Multiprotocol VLAN (that is, VLANs, PPPoE, and IP)
 The Ethernet interface supports multiple upper-layer protocols, one upper-layer interface per protocol, on each of one or more VLANs.

The first upper-layer interface configured on an Ethernet interface determines the mode. Switching to a different mode requires first removing the entire stack above the Ethernet interface. The supported upper-layer protocols are IP and PPPoE. Figure 3-30 shows the interface stacks for the different modes.

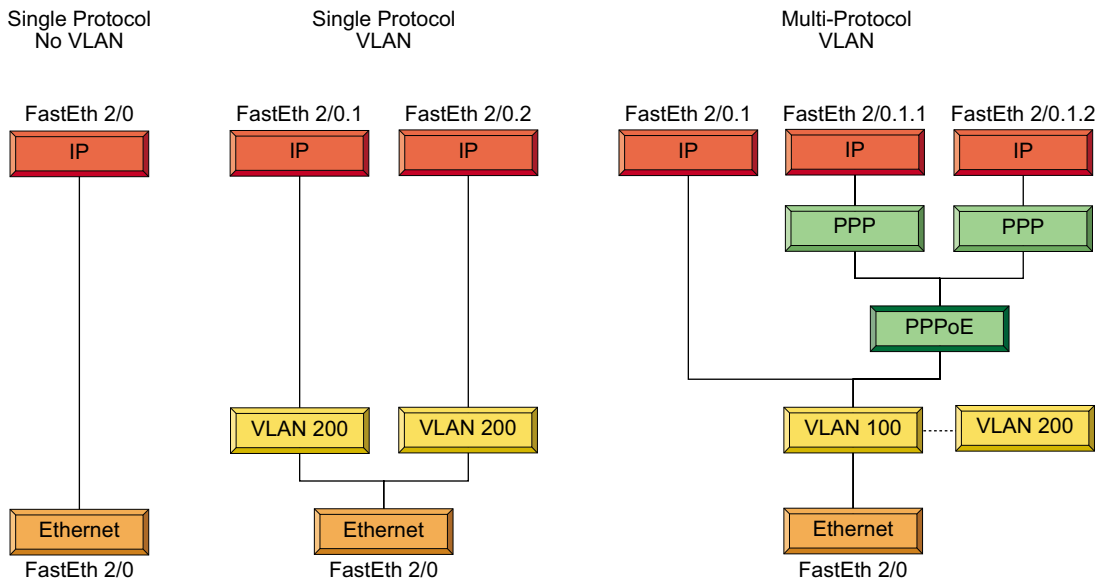


Figure 3-30 Interface stacks for VLANs

