New Generation ClearChannel Architecture for Catalyst 1900/2820 Ethernet Switches

Executive Summary

This white paper describes the ClearChannel architecture and salient features of the new Catalyst® 1900 and 2820 Ethernet switches. Cisco developed the ClearChannel architecture to address performance requirements of bandwidth intensive networks at the workgroup and desktop level. The new Catalyst 1900/2820 Ethernet switches are based on third generation ClearChannel architecture.

Switches are integral components of a large and robust LAN, they have four essential elements: buffering and filtering, forwarding mechanisms, network management, and backplane bus architecture. The Catalyst 1900/2820 Ethernet switches have been designed to optimize performance along each of these dimensions at a very low cost.

Buffering and Filtering

The Catalyst 1900/2820 Ethernet switches integrate a unique shared memory buffering architecture and intelligent filtering mechanism to offer industry-leading price/performance among Ethernet switches.

The Catalyst 1900/2820 Ethernet switches use shared memory buffering that enables optimal utilization of packet memory. This architecture uses memory efficiently by allowing buffers to be dynamically allocated to ports as needed and by avoiding packet duplication during multicast and broadcast transmission. Packets are never copied or moved to another memory location. Because the total memory capacity is shared among all ports, shared memory buffering allows much larger burst packet capacity, virtually eliminating the possibility of dropped packets—even during peak loads. This feature makes the switch ideally positioned to handle bursty network traffic. A unique hardware implementation of the shared memory architecture allows for low latency, wire-speed switching across the network. This architecture completely eliminates a phenomenon that affects port based buffering architectures in which packets destined for one congested port can block packets going to other ports in the switch. Widely known as 'head of line blocking,' this problem results in underutilization of the switch.

With increasing use of audio/videoconferencing packages and video playback that use broadcast and multicast traffic, even high capacity networks encounter bandwidth congestion. Switches that proliferate network traffic without intelligent filtering can seriously degrade network performance. The Catalyst 1900/2820 Ethernet switches efficiently limit multicast flooding by interoperating with intelligent routing software to restrict transmission to only ports interested in receiving a particular multicast. This is accomplished by the switch communicating with routers through the Cisco Group Management Protocol (CGMP), which has been designed to interoperate with the router based IGMP (Internet Group Management Protocol). CGMP is a layer 3 enhancement that provides value-added functionality and tighter integration between the switch and traditional router-based Cisco Internetwork Operating System (Cisco IOS™)software components.

The switches incorporate multicast address packet filtering in addition to CGMP. As a result, the switch can handle both IP-based multicast and MAC address-based multicast protocols in the network. Multicast filtering can be further combined with source port filtering—packet forwarding using both the destination address and the source port—to achieve load balancing across ports. The switches also incorporate broadcast storm control—a Cisco IOS feature—to filter out broadcast traffic from a port that is receiving unacceptable levels of broadcasts without affecting unicast and multicast traffic.

Forwarding Flexibility

The type of forwarding mechanism needed in a switch depends on the network. The Catalyst 1900/2820 Ethernet switches can be configured to use any of three forwarding mechanisms: FastForward™ cut-through, FragmentFree™ cut-through or standard store and forward. For high speed low latency networks, cut-through switches are the best suited. For congested networks with high collision rates, FragmentFree mode provides better error checking than FastForward cut-through with practically no increase in latency. The same switch at the desktop level might use FragmentFree mode but as a workgroup switch aggregating shared media hubs, use store and forward mode for enhanced error checking. Flexibility in choosing forwarding mechanisms allows positioning of Catalyst 1900/2820 Ethernet switches at different points in the network without sacrificing efficiency.

Catalyst 1900/2820 Ethernet switches integrate full-duplex 100BaseT ports, which can be used for high-bandwidth connections between switches, or switch to router or switch to server. The switches also allow for extended network diameters through full duplex 100-Mbps over fiber cabling with single links up to 2.2 Kilometers.

Network Management

Network management is critical to configuring and monitoring complex switched networks. As the complexity of the network increases, so does the demand on network management. Catalyst 1900/2820 Ethernet switches incorporate powerful network management capabilities including SNMP management, embedded RMON support and port mirroring. The switch supports IEEE 802.1D Spanning-Tree Protocol for redundant backbone connections and loop free networks that simplify network configuration and improve fault tolerance.

As networks grow in size and complexity, so does the need for increased security to protect access to corporate resources. The Catalyst 1900/2820 Ethernet switches incorporate port based security mechanisms to prevent unauthorized users from accessing the network. Any unauthorized attempt at accessing the switch can be detected from a network management console. In addition, source port filtering can be used to further restrict access to network resources.

Backplane Architecture

The backplane architecture determines the forwarding rates attainable by the switch. Two critical elements are the capacity of the backplane bus and the mechanism for scheduling data transfer over it. The Catalyst 1900/2820 Ethernet switches use a 1-Gbps high capacity backplane bus. In addition, access to the bus is scheduled in a pipelined fashion so that no bandwidth is lost to arbitration delays. This enables the switch to implement wire-speed throughput on all its ports. The forwarding engine is implemented entirely in hardware, which, combined with a shared memory architecture, results in a true internally nonblocking switch.

The EtherSwitch switches have aggressively driven the cost down through ASIC integration. Four Ethernet controllers are integrated onto a single ASIC; two high speed port controllers are also integrated into one ASIC. Eight Ethernet transceivers are integrated in a single ASIC permitting only three ASICs to handle 24 ports. This setup has resulted in fewer components, simplified hardware configuration, reduced, cost and higher reliability.

Organization

This paper is organized into four sections.

The first section gives a broad overview of the switch architecture. It briefly describes the functions of each of the components that make up the ClearChannel architecture.

The second section describes two integral switch processes, forward processing and address learning. Basic forward processing—during which the destination port is determined and the packet forwarded to it- and source address learning—extracting and storing the association between an address and a port—are examined.

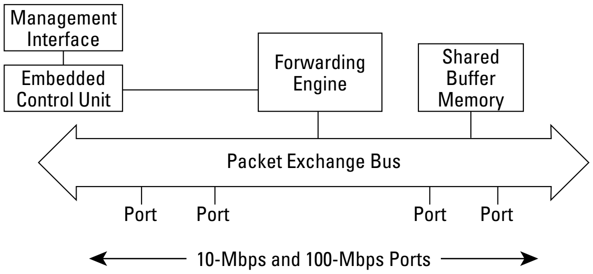
The third section describes the hardware layout of the switch. The principal ASICs—that constitute the switch, Ethernet Controller, Transceiver, Forwarding Engine and the Master Scheduler—are described.

New switch features that are visible to the end user are described in the final section. CGMP, Cisco Discovery Protocol (CDP), multicast filtering, port security, and RMON support are discussed.

Architecture Overview

In order to satisfy the performance requirements of bandwidth intensive networks and meet the needs of workgroup computing, Cisco has developed the ClearChannel switch architecture. The Catalyst 1900/2820 Ethernet switches are third generation switches based on the ClearChannel architecture. Figure 1 gives a block level view of the switch architecture.

Figure 1 ClearChannel Architecture



Forwarding Engine

The heart of the switch, the forwarding engine, is responsible for the central function of the switch: examining packets from incoming ports, looking up the destination address, and queuing them to the appropriate port for transmission. The forwarding engine is implemented entirely in hardware to ensure low latency and higher throughput through the switch. This implementation also reduces complexity, increases reliability and consequently lowers the switch cost.

In addition to packet processing, the engine also collects and maintains switch statistics. By monitoring the packet exchange bus, it is able to count packet lengths, throughput, errors, and exceptions. This data is collected by network management stations to construct switch statistics tables and monitor traffic patterns on the networks which can in turn be used to design efficient network architectures. Forwarding is discussed in detail in the next section.

Packet Exchange Bus (Xbus)

The primary bus connecting the key functional units within the switch is a high speed bus called the Xbus. Data is transferred within the switch through the Xbus. The bus sees all of the traffic passing through the switch. Access to the bus is prioritized since several components could attempt to put data on the bus concurrently. Access is sequenced according to transaction priority and time of arrival. Transaction requests have different priorities. A transaction requesting buffer memory for a packet, for example, has higher priority than one sending status at the end of a transmission. Xbus is also used to transmit signals between switch components for initiating transactions associated with receiving and transmitting packets. Access to the bus is controlled by a separate master scheduler.

Xbus is a 53-bit-wide bus running at 20 MHz. This high bandwidth enables the switch to be completely non-blocking, that is, it is able to handle wire-speed on all ports concurrently.

Embedded Control Unit

Most of the per-packet processing in the switch is done in hardware but network management and related activities are handled by software in the embedded control unit (ECU) subsystem. This clear demarcation allows for regular switch forwarding functions to be performed at wire-speed with minimal latency in hardware, leaving the ECU to handle more complicated scenarios. The ECU subsystem comprises the following modules:

- Embedded central processing unit (CPU)
- 512-KB DRAM (for CPU)
- 1-MB Flash for firmware, configuration data and statistics

Flash memory is partitioned into three areas—switch software image, switch configuration data, and boot sector. Most of the CPU Flash is used to store the switch software image. This is also the area that changes during software upgrades. The area above this area in Flash stores the switch configuration data. This area contains information about the packet switching mode, switch IP address and subnet mask, broadcast storm control, full-duplex configuration, statically configured addresses and other switch configuration parameters. Whenever a new switch configuration is written, data in this section of the Flash change.

The boot sector located at the top of the flash is write-protected. The switch software image is not write-protected so that software upgrades can be made without opening the switch enclosure. The boot sector contains a back-up software image load capability, allowing the user to reload a replacement software image should the primary image become corrupted, such as by an interrupted upgrade. The boot sector ensures that there is at least one section of the Flash that always retains its data integrity, allowing the switch to boot up correctly every time.

The ECU subsystem is responsible for diagnostics and error handling, switch configuration, Spanning-Tree Protocol, in-band and out-of-band management, statistics reporting, and control of the front panel display. In-band management refers to switch management through telnet or SNMP using an application such as CiscoWorks.™ Out-of-band management refers to management through a menu-based console connected to the serial port. The ECU also contains an embedded RMON software agent that provides enhanced manageability, monitoring, and traffic analysis to a network management station.

Shared Buffer Memory

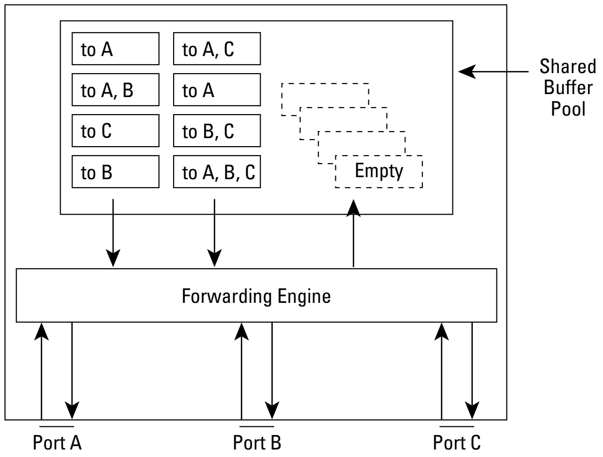
Catalyst 1900/2820 Ethernet switches employ a shared memory buffering scheme using 3 MB of dynamic RAM, providing a large packet buffer as a shared system resource for dynamic allocation of packet buffer memory to individual ports. Using a memory allocation mechanism, incoming packets are deposited into "empty" buffers that are kept in a common pool available to all ports. Buffer identifiers, or "pointers" are then dynamically "linked" into the appropriate transmit queues and the packets are transmitted when allowed by the destination ports. These are logical queues which are associated with each port. A packet remains in the same, single memory location until it has been correctly forwarded by all destination ports. Packets are never copied or moved to another memory location. In the cut-through switching mode (described in the "Switch Features" section), leading bytes of a packet are fetched and transmitted before the entire packet has been received into the buffer.

This architecture optimizes buffer utilization by allowing buffers to be dynamically allocated to ports as needed and by avoiding packet duplication for multicast and broadcast packets. The forwarding engine maintains a map of ports on which a packet buffer needs to be transmitted, and retains the memory until the buffer has been forwarded through all appropriate destination ports. Because the total memory capacity is shared among all ports, shared memory buffering allows much larger burst packet capacity virtually eliminating the possibility of dropped packets—even during peak loads.

The shared memory architecture in Catalyst 1900/2820 Ethernet switches completely eliminates a problem in port-based buffering architectures known as head of line blocking which can lead to underutilization of a switch. In receiver-based and bi-directional port-buffered switches, packets are stored in queues associated with incoming ports or both incoming and outgoing ports. A packet is transmitted once all other packets ahead in the queue have been successfully transmitted. As a result, a single packet that cannot be transmitted because of a busy destination port can block all packets in the queue behind it even if other destination ports are completely free. Shared memory switches do not have receive queues and are not subject to head of line blocking.

In order to prevent one single port from monopolizing all available buffers and starving out other ports, the switch enforces a 1.5-MB limit on the amount of packet buffer memory that may be queued for any one port at any instant of time. This set-up assures that, at most, half the available buffer pool may be utilized by any one port, regardless of circumstances.

Figure 2 Shared Memory Buffering



The switch is unique in its hardware implementation of a shared buffer memory architectures. Simulations, supported by more than three years of extensive field experience, demonstrate that the 3-MB shared buffers are more than sufficient to prevent any packet loss, even under the heaviest asymmetric loads found in client/server configurations and multimedia applications.

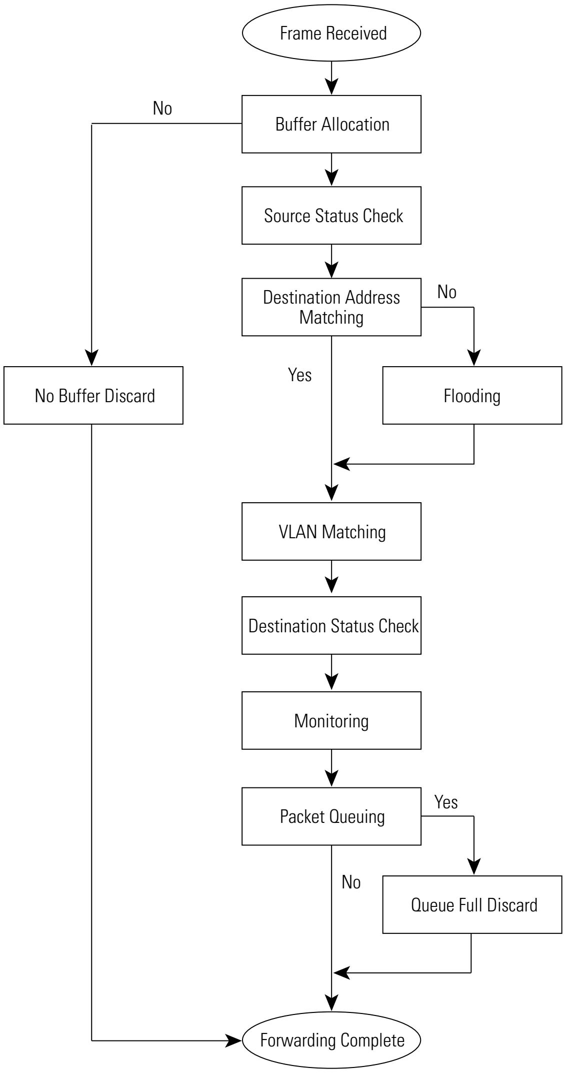
Forwarding Engine

Upon receipt of the packet, the switch processes it along two concurrent paths. The *forwarding **decision*** is a function of the destination address, VLAN configuration, port monitoring, and source address. The output of this decision is a list of destination ports on which the packet is scheduled for transmission. *Address learning* occurs concurrently with the forwarding decision. The forwarding engine learns the port on which a particular end station resides by looking at the source Ethernet address in the packet. The engine makes and stores the association between a port and the Ethernet address of an end station if it has no prior knowledge of it.

Forwarding Decision

The following flow-chart in figure 3 outlines the steps involved in making a packet forwarding decision.

Figure 3 Packet Forwarding



Buffer Allocation

An incoming frame is stored into an empty packet buffer. If no more empty buffers are available, a *no buffer discard* error is generated for the port and the packet is dropped. This condition should never occur under normal switch operation.

Source Status Check

The forwarding engine next checks the source port status, which can be in one of three states: active, suspended or disabled. An active port is in a ready state to receive packets, the normal state for the port. A port can get suspended by Spanning-Tree Protocol to break a topology loop in the network, by address violation caused when a secure port receives a packet from an unauthorized station or through management control. Although a suspended port does not forward the packet, it is monitored by the ECU subsystem for diagnostic purposes. A suspended port gets re-enabled by Spanning-Tree Protocol or by encountering a valid packet after an address violation. A port gets into a disabled state upon encountering an address violation or by management control. A disabled port can only be re-enabled through management intervention.

Destination Address Matching

The forwarding engine next looks up the destination address in the content addressable memory (CAM) table. The CAM table maintains an association between MAC addresses and ports. If the address is found in CAM, the packet is flagged for delivery to the corresponding ports. If the address is not found in the CAM table, the packet is flagged for delivery to all ports except the source port. This process is called flooding and it happens every time the switch receives a packet for an address it is unaware of. Note that an entry (address, port) is added to the CAM during the source address learning process described later.

VLAN Matching

The Catalyst 1900/2820 Ethernet switches address the scalability issues and network management complexities associated with flat switched network topologies through port-based VLANs. The Catalyst 1900/2820 Ethernet switches support four port-based VLANs per switch. After obtaining the list of destination ports from the CAM table, the forwarding engine uses the VLAN map at this point to further prune the list of ports on which this packet would be forwarded. Each VLAN has its own bridge MIB and spanning tree.

The port-based VLANs in Catalyst 1900/2820 Ethernet switches are restricted to a single switch without a VLAN trunk port. Both the source and destination ports for a packet must be in the same VLAN. If the destination port is outside the source VLAN, then the packet is treated the same as a packet with an unknown address. The ports belonging to the source address VLAN are flooded with the packet.

Destination Status Check

The switch maintains a map of destination ports for each packet in buffer memory. If any destination port is in a suspended or disabled state, it does not receive the packet and is removed from the map. If all destination ports are suspended or disabled, the switch frees the buffer.

Monitoring

The last check before the packet can be queued for transmission is to determine if the destination port is being monitored. If so, the switch adds the monitoring port—called the Switch Port Analyzer or SPAN port—to the map of ports associated with the packet. Usually the network manager configures one of the high speed ports as a SPAN port.

In order to analyze the functioning of a switch, an administrator can configure any port in the switch to be a monitoring port. This port can be set up to see traffic destined for any port or group of ports in the switch. An RMON probe or sniffer can attach to a monitoring port to analyze network traffic. The switch does not mirror broadcast packets or spanning-tree packets to the monitoring port—spanning-tree packets are used to spread information about network topology to the switches and is intended to be purely inter-switch traffic. This ensures that the monitoring port does not see duplicate broadcast packets or control packets that do not originate from a host in the network. SPAN port provide a powerful and flexible way to monitor switch performance.

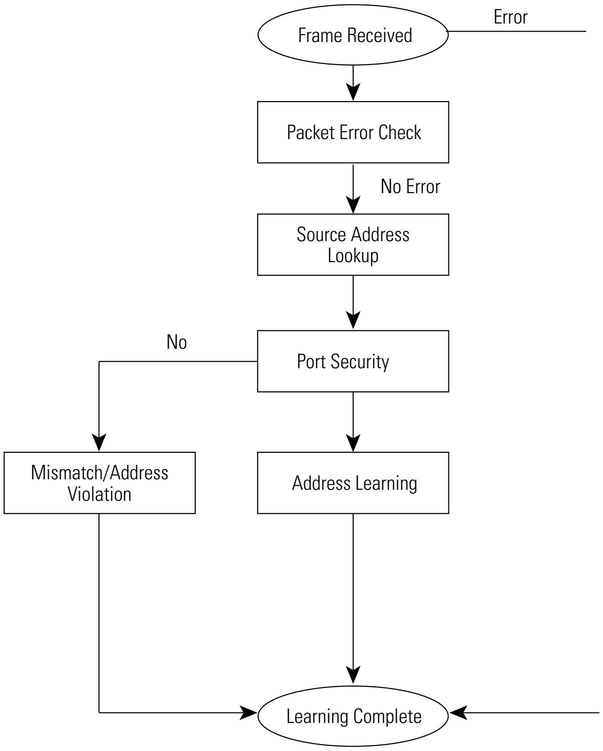
Packet Queuing

The packet is now forwarded to each destination port in the map, starting with the port with the highest priority. In order to ensure that no single port occupies all available shared memory, the switch imposes a 1.5-MB limit on each destination port. If a destination port already has more than this amount of buffer memory allocated to it, a *queue full discard* error is generated and the buffer is freed. Once the packet is successfully transmitted through all destination ports, the buffer is freed. This completes forwarding processing.

Source Address Processing

The Flow-chart in figure 4 outlines the steps involved in the source address learning process.

Figure 4 Source Address Processing



Packet Error Check

Source address processing begins as soon as the source Ethernet address is read from the incoming packet. The packet is checked for errors. In case an error is found in the packet, the switch does not learn information about the source address. But since forwarding in cut-through mode begins as soon as the destination address is read, the erred packet continues through the switch like a regular packet. In store and forward mode, the switch drops the packet if it encounters an error.

Source Address Lookup

If the packet is error free, the CAM table is examined to see if the source port already exists in it. The source port would exist in the CAM table if the switch had previously learned this address. This CAM table is the same table in which the destination address is looked up during frame processing. If the address is not found in the CAM table, the switch goes through the address learning process. First it must check to see whether port security has been enabled.

Port Security

Port security can be set to prevent unauthorized users from accessing the network. Each port can have an individual address or a group of addresses that represent stations permitted on that port. A secure port is allowed to have a maximum of 132 addresses associated with it. If the source address of a received packet does appear in the CAM table, but is associated with a different port than the one on which it was received, it is called a duplicate address violation. If the address does not appear in the CAM table, it is added to the list of secure addresses associated with that port. In the case when the per-port limit of secure addresses is reached (defaults to132), the address is not added and *a mismatch violation* is generated. The administrator can specify the action upon detecting a violation—disable or suspend a port, issue a trap or ignore the violation.

If port security is not enabled, then the packet goes from CAM table lookup straight into address learning.

Address Learning

The switch "learns" an address when it adds an entry (address, port) to its forwarding database by saving the source address of each LAN transmission that it receives, along with the port identifier for the port on which it was received. Addresses are dynamically learned and stored in the CAM table. If a source address is not found in the CAM table, it is learned (stored) for future reference. A time stamp is recorded with the address which allows the address to "age." Once a predetermined age is reached, the address expires and is erased from CAM. Each time an address is referenced, its time stamp is updated and the aging process begins again. In the case where the address exists in CAM, but causes a mismatch address violation, the time stamp is not updated. Address aging in Catalyst 1900/2820 Ethernet switches conforms to the IEEE 802.1D specification.

Addresses can also be entered manually in the switch. These are static addresses which do not age. Static addresses remain in CAM until manually removed.

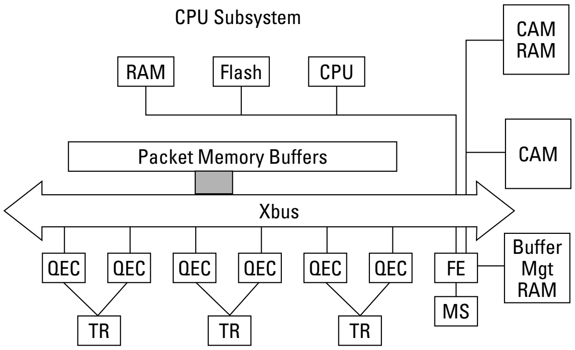
If there is no more space in CAM—a situation that occurs after the CAM has l024 addresses for EtherSwitch 1220 and 2048 or 8096 addresses for the EtherSwitch 1420—the switch skips the address learning process for an incoming packet. The number of nodes in the network can be greater than the size of the CAM. If a switch with a CAM space of 1024 is placed in a network with a greater number of nodes, and despite address aging the CAM becomes full, it will continue to forward the packets correctly but will resort to flooding to forward packets it cannot match in CAM tables—resulting in some performance degradation. Address aging ensures that the switch uses the CAM most efficiently, retaining only addresses that have been active over a certain time period.

This completes the source address learning process.

Switch Hardware

The switch implements the forwarding engine entirely in hardware that allows for low latency, wire speed switching at a high level of reliability and low cost. The switch functionality is embodied in custom ASICs consisting of ethernet controllers, bus master scheduler, forwarding Engine and the 10BaseT transceivers. The CPU subsystem and the switch software is responsible for initialization, configuration management, Spanning-Tree Protocol, control of the front panel display, diagnostics, and error handling.

Figure 5 Hardware Layout



Quad Ethernet Controller ASIC

Each Quad Ethernet Controller (QEC) ASIC contains four 10-Mbps Ethernet controllers. The QEC implements the Ethernet carrier sense multiple access collision detect (CSMA/CD) protocol. It is responsible for the initial receipt of an Ethernet packet in the switch and the final transmission out of the switch.

The QEC receives the packet and deposits it on to the Xbus—the central data highway within the switch. The Xbus data path is 48 bits wide. The controller assembles the data into 48 bit words before sending it to packet memory over the bus. The QEC communicates with the master scheduler described below to schedule data transfer with the bus. After the switch has processed the packet and is ready to transmit, the QEC reads the data from the Xbus, assembles it into appropriate frames and sends it out of the switch. If a collision is detected during transmission, the QEC informs the Forwarding Engine that the packet has experienced a collision. It is up to the Forwarding Engine to resubmit data for transmission. All transmit attempts are identical for the QEC; it does not distinguish between original transmission attempts and collision retries.

Six QEC ASICs reside on the board that control 24 ports.

10BaseT Transceiver ASIC

A transceiver performs physical level 1 functions such as electrical signaling, monitoring line states, encoding, and decoding data from the wire. Eight 10BaseT transceivers plus one 10BaseT or AUI drop cable configurable transceiver are integrated into a single transceiver ASIC (TR). The TR, which interfaces directly with the wire reads the data from the wire and passes it on to the appropriate QEC. During transmission, the TR receives data from the QEC and sends it out on the wire. Three TR ASICs handle 25 ports and interface with the six QEC and MS ASICs.

High Speed Ethernet Controller ASIC—Master Scheduler

This Master Scheduler ASIC (MS) contains one additional 10-Mbps Ethernet controller and two 100-Mbps controllers. In addition to performing standard controller functions similar to QEC, it also orchestrates all data movement across the Xbus. The MS generates Xbus cycles at the appropriate times to transfer data into and out of the Xbus.

Forwarding Engine ASIC

The Forwarding Engine ASIC (FE) is the heart of the switch. As the name suggests, it makes all the packet forwarding decisions within the switch.

The FE is first made aware of the packet reception by the MS. In response to an indication by the MS, the FE obtains a free buffer and transfers the incoming packet to buffer memory. The FE monitors the transfer and notes the destination and source address fields from the packet. The FE submits the destination address for lookup in the CAM table and determines—based on the feedback from the CAM—which ports should ultimately transmit the packet. The source address field is used for the address learning process.

The FE creates a map containing a list of destination ports for each packet—multicast and broadcast packets and packets with unknown source addresses have multiple ports associated with them. The FE initiates a transmit process for each port in the map. Once the transmit process has been initiated, the FE waits for transmit status to be returned by the QEC. The QEC returns a status indicating successful transmission or a packet collision.

If the packet is successfully transmitted, the FE clears the bit in the packet buffer map that corresponds to the port that just completed its transmission. If all bits in the map are clear, implying there are no more destination ports, the buffer is returned to the free pool. If the QEC reports an indication that a collision has occurred, the FE initiates a new transmission attempt according to the Ethernet protocol.

In addition to packet processing, the FE also collects and maintains switch statistics.

Processor—CPU

The switch uses an embedded processor as the central processing unit. The most important task of the CPU is handling network management for the switch. All in-band and out-of-band management is handled by the CPU. When a management session is established with the switch through Telnet or SNMP, the CPU displays the switch statistics, takes requests for configuration changes and takes the appropriate action to implement the changes. Normal forward processing completely bypasses the CPU.

The CPU does see two kinds of packets (broadcast packets and in-band management packets)—that are destined for the switch rather than individual nodes.

Interswitch communication, such as information about the network topology using Cisco Discovery Protocol (CDP), sometimes occurs through broadcast transmission. The CPU processes the broadcast packets in order to act on this information. In-band packets such as the Telnet and SNMP packets that are destined for the switch are also processed by the CPU. These packets contain configuration information.

The FE forwards packets that require further processing to the CPU. The packets are not actually transmitted to the CPU but merely flagged in the receive buffer as of interest to the CPU. The CPU is then able to read data directly from the packet buffer. For transmitting a packet, the CPU requests FE for a buffer allocation. The CPU forms the packet within this buffer and creates a map that contains the list of ports on which this packet would be transmitted. The FE then transmits the packet through the destination ports.

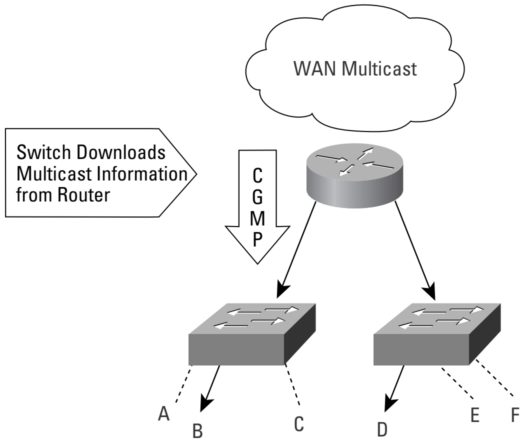
Switch Features

This section gives details of the new features available in the Catalyst 1900/2820 Ethernet switches.

Cisco Group Management Protocol

The use of multicast traffic has increased with the delivery of new applications such as audio/video conferencing packages, video playback, and real time financial data delivery services. In these applications, a packet sent from one workstation is proliferated over the switched network to all other workstations that want to participate in a particular conference or broadcast. Although routers selectively send the multicast down relevant links only, the switch forwarding engine sends multicasts to every switched port configured within that VLAN, resulting in unnecessary bandwidth use within the switch as well as on the attached LANs. CGMP efficiently limits flooding by restricting transmission to only the ports interested in receiving the multicast.

Figure 6 IP Multicast through CGMP



CGMP prevents the flooding of IP multicast packets among all switched ports within a VLAN. CGMP relies on a Cisco router to download the identity of the multicast clients within the switched network to the switch. Using this layer 3 information, the switch then programs the Forwarding Engine to switch these multicast packets at wire speed to only those ports interested in this traffic. The router exchanges and obtains multicast information with other routers using the standard Internet Group Management Protocol (IGMP). CGMP inter-operates with hosts running IGMP version 1 and version 2, to obtain the multicast information. In the figure above all nodes (A through F) are in the same broadcast domain. Only hosts B and D, however, are interested in the multicast. CGMP ensures that the multicast traffic is switched only to B and D and does not interrupt activity on the rest of the nodes.

CGMP is a mechanism for handling layer 3 multicast or IP multicast through tighter integration between the switch and traditional router-based Cisco IOS software components.

Broadcast Storm Control

In some cases the level of broadcast traffic can become so severe that virtually no bandwidth remains for application data. In this case new network connections cannot be established and existing connections may be dropped. This situation is often referred to as a “broadcast storm.” Since broadcast storms usually result from faulty end stations that begin generating unacceptable level of broadcasts, the probability of broadcast storms increases as the switched internetwork grows.

Catalyst 1900/2820 Ethernet switches incorporate Broadcast storm control—a Cisco IOS feature—that allows an administrator to automatically suspend a port that is receiving an unacceptable level of broadcast traffic. Each port can be configured with a broadcast threshold. When the number of broadcast packets per second exceed this limit, the switch disables forwarding of any broadcast packets that originate from this port. Unicast and multicast packets are not affected; they continue to be forwarded as before. Broadcast forwarding is resumed when the broadcast rate falls below a low watermark on the port. By default, broadcast storm control is disabled.

Multicast Address Packet Filtering

The Catalyst 1900/2820 Ethernet switches provide the ability to filter multicast traffic by MAC addresses. Multicast packets are identified by a multicast group number. In order to receive a multicast packet, a host has to belong to that multicast group. The switch maintains a mapping between multicast group addresses and port numbers (not host MAC addresses).

Table 1 Multicast Group Address Information in CAM

Multicast Group Address	Port Numbers
4	5, 2, 4
7	4, 6
1	2, 6, 1

The switch maintains a table—as part of the CAM table—that maps a multicast group address to a set of ports. When a multicast packet is received, the switch sends the packets out to all the ports that belong to that multicast group. In Table 1, if a packet is addressed to multicast group 7, it will be sent out to ports 4 and 6 in the switch. Note that the switch will go through the regular checks before queuing the packets for transmission at the destination ports (checking port status and VLAN membership). If a multicast packet is received that does not correspond to any group address registered within the switch, it is treated as a standard multicast packet. The switch floods the unregistered multicast on all the ports except the source port. In addition, flooding of unregistered multicast packets can be disabled on a per-port basis. Registering ports into multicast groups is accomplished manually through switch configuration.

Multicast address packet filtering can be combined with source port filtering, to do load balancing. If there are two servers on the network that provide the same service and belong to the same multicast group, then the network manager can add source port filtering to ensure that certain ports reach one server and other ports reach the other server. Hence load is distributed between the servers while keeping the process totally transparent to the clients.

While CGMP offers the capability to extend IP multicast's layer 3 based filtering into the switched domain, multicast address packet filtering—in which multicast addresses are entered manually in a switch—allows MAC address based multicast protocols to be efficiently filtered in the network even in the absence of IGMP support in routers.

Spanning-Tree Protocol

A network wiring infrastructure often provides more than one path for a packet to go from source node to a destination node. Existence of redundant paths in a network gives rise to a problem. A switch can see a packet originating from a host on two different ports—implying a topology loop in the network. Indeterminate forwarding behavior can then result, and broadcast packets may traverse loops indefinitely. To prevent this, a protocol called the Spanning-Tree Protocol is executed between the switches to detect and logically remove redundant paths from the network. A spanning-tree protocol essentially establishes a root node and constructs a network topology such that there is exactly one path for reaching any node. Network devices exchange messages with each other to detect loops, and then remove the loops by shutting down selected interfaces. The protocol also ensures that in case of failure of an intermediate node, the redundant paths are utilized to construct a new tree that circumvents the failed node and maintains connectivity with nodes that lie downstream from it. The Catalyst 1900/2820 Ethernet switches conform to the IEEE 802.1D Spanning-Tree Protocol specification.

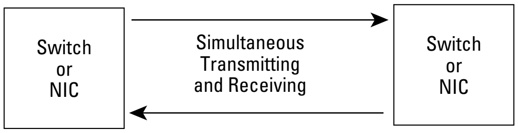
Each configured VLAN has its own instance of a spanning tree. Since the switch supports a maximum of four VLANs, it can participate in up to four spanning trees.

The Catalyst 1900/2820 Ethernet switches also provide a "fast awakening" extension to spanning tree operation. In this extension, direct switched links to end stations are enabled into the spanning tree in a few seconds after attached stations are powered on. This prevents the problems with initial downtime on a network connection experienced by some desktop switches when spanning tree is enabled on them.

Full Duplex Support

Full duplex provides the ability to transmit and receive data simultaneously. Since standard Ethernet is a shared media it is half-duplex by definition. The use of UTP for Ethernet cabling that has separate pairs of wires for transmitting and receiving, and the advent of switching, which allows for transmission channels not shared among multiple users, have permitted full duplex to be implemented over Ethernet.

Figure 7 Full Duplex Ethernet



The Catalyst 1900/2820 Ethernet switches supports full duplex operation on the fast Ethernet 100BaseT ports. The total available throughput on these ports is 200-Mbps. In order to achieve this throughput, both ends of the segment must be configured for full duplex. Full duplex 100BaseT ports can be used for high-bandwidth connections between switches, or switch to router or switch to server. The switch also allows for extended network diameters through full duplex 100-Mbps over fiber cabling up to 2.2 km. Full duplex ports cannot be used to connect to a repeater since a repeater is a shared bandwidth device and is inherently half duplex.

Switching Modes

Catalyst 1900/2820 Ethernet switches support three types of switching modes. The latency for packet forwarding through the switch depends on the choice of switching modes. The faster modes trade off error checking for low forwarding latency. Switch throughput is not affected by the choice of switching modes; it is always at wire speed. Switching modes are set through the configuration menu.

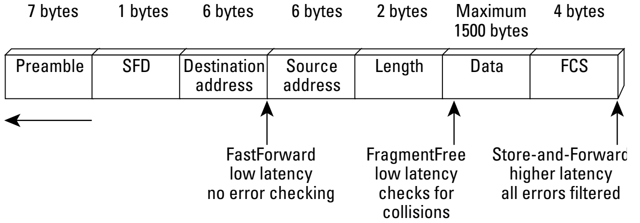
Cut Through: The switch supports two types of Cut-through modes: FastForward and FragmentFree. In FastForward mode, forward processing begins as soon as the destination address is recognized—in the first 6 bytes of an Ethernet frame.

FragmentFree or modified cut-through mode holds on to an incoming packet until the first 64 bytes have been received. If a packet has an error, it almost always occurs within the first 64 bytes. FragmentFree mode provides better error checking than FastForward mode with practically no increase in latency. This mode incurs a 70-microsecond latency.

Store and Forward: Store and forward receives the entire packet before initiating forwarding. All packet errors are detected before forwarding. Since the time to forward in this mode depends on the length of the packet, store and forward latency is measured from the time the last bit is received to the time the first bit is transmitted out of the switch. Store and forward latency is seven microseconds, excluding the length of the packet.

All packets going from 10-Mbps ports to 100-Mbps ports are sent store and forward. Since the destination port is receiving data at a faster rate than the rate at which the data is arriving from the source port, cut-through operation would cause the transmitter to prematurely run out of data, resulting in a packet error. Hence when establishing a virtual connection from a low speed port to a higher speed port, the switch has to buffer (store) the packet before forwarding. All broadcast packets are sent store and forward as well.

Figure 8 Store and Forward



Cisco Discovery Protocol

One of the problems that arises after installing switches in a multi-layer hierarchy is that network managers lose their visibility of the network, traffic, and users. Often, managers have no idea of the amount of traffic flowing through the network. They are unable to keep track of changes being made by other administrators and lack complete information on how the network is connected.

While there are several network management tools that reveal the router-to-router connections, they are inadequate at discovering and mapping switches and how there are linked within the network. Understanding these connectivity relationships is fundamental to managing multilayer switched networks. It enables configuration checking and analysis at the physical layer, including the links between switches, routers, and hubs.

Cisco IOS software uses the CDP to obtain and maintain a comprehensive understanding of the network topology. CDP defines a protocol to exchange information between devices in the network. The CDP management agent is part of the Cisco IOS software and functions as a lower-layer protocol that interrogates adjacent devices for information. CDP agent gathers information about the type of devices in the network, configuration information about the links connecting those devices, and the number of interfaces within each device. A network management application, such as VLANDirector,™ imports data provided through CDP and uses that to construct a low-level connectivity model of the network and a graphical topology map.

Security

As a network grows, the possibility of unauthorized access to critical data increases. A switch based network makes it relatively easy for a rogue device to enter by attaching itself on to a port in the network. Catalyst 1900/2820 Ethernet switches provide port based security mechanisms to prevent unauthorized users from accessing the network.

The network administrator can associate an individual address or a group of valid addresses to a port. Only the addresses that appear on the list are allowed to send data through that port. If an address that does not appear on the list sends a packet to the port, the switch recognizes it as an address violation and raises an exception. Depending on the type of administrative action desired upon an address violation, the switch can entirely ignore the violation, send an alert, disable the port or suspend the port. The switch allows security to be set on a per port basis.

RMON Support

In traditional Ethernet implementations that used a completely shared media, such as 10BaseT, thinnet or thicknet coaxial cables, network management was relatively easy. A traffic analyzer attached to the segment could see all traffic passing on the network. Any event on one end of the network could easily be seen at all other points of the network.

In a switched network a traffic analyzer can only provide information about the segment it is connected to; it cannot see events on other ports. In order to do that, network management tools use Simple Network Management Protocol (SNMP) to communicate and collate information among managed elements of a network. SNMP messages can only be sent and received by SNMP-enabled devices. The information used in this communication is stored in a Management Information Base (MIB). Under this paradigm, the network elements are responsible for keeping their own statistics in a MIB compliant database. Statistics and management information for network devices are collected in an RMON MIB.

Using an RMON MIB, a network management tool can monitor traffic coming in and out of distant ports. An RMON MIB defines nine groups containing information pertinent to managing a switched network. An RMON based network management application can graphically display statistics and event generation based upon information stored in groups of the RMON MIB. The Catalyst 1900/2820 Ethernet switches have embedded support for four RMON MIB groups. Table 2 describes the relevant groups.

Figure 9 RMON Groups Supported in Catalyst 1900/2820 Ethernet switches

Group Number	Group Name	Function	Description
1	Statistics	Measures utilization and error statistics for each monitored device	Reports statistics for an entire device, such as how many erroneous packets a switch has seen on all ports combined
2	History	Reports statistical samples with a given time period	Allows statistics to be viewed in 30-second or 30-minute intervals.
3	Alarm	Generates an external event if a given threshold is exceeded	Generates alarms based on preprogrammed thresholds; if the number of CRC errors, for example, exceed a predefined limit, an alarm signal is generated; an RMON manager can receive this alarm and determine if it is important.
9	Event	Controls the generation of events based on network information	A special SNMP RMON event packet can be generated based on statistics and information obtained from other groups

Any network management application can obtain information about the four RMON groups by using SNMP to query the switch. An external RMON probe can obtain information beyond the four groups by monitoring traffic on the SPAN port. The switch mirrors traffic going through all other ports on to the SPAN port. The external RMON probe connects to the SPAN port to construct the desired information.

FDDI Transparent Bridging

FDDI packets have a different frame format than Ethernet packets. The FDDI module for the Catalyst 2820 uses transparent bridging to accomplish translation between Ethernet and FDDI. After reading the packet, the module strips off the layer 2 information from it. It constructs a new packet by adding FDDI layer 2 information—destination address, source address, packet type, control bits and so on. The module thus translates an Ethernet packet into an FDDI packet before transmission to the destination address. The packet looks exactly the same as if it were transmitted from a FDDI host. Since the module processes the packet before transmission, switching between Ethernet and FDDI can only be performed in store-and-forward mode.

FDDI IP Fragmentation Support and MTU Discovery

Not only are the FDDI and Ethernet frame formats different, the maximum transmission units (MTU) for the two protocols are different as well. FDDI allows for a maximum size of 4500 bytes whereas the Ethernet MTU is only 1518 bytes. This creates problems when packets originating from FDDI cannot fit into the largest permissible Ethernet packet.

The module handles this situation by fragmenting the FDDI packet. IP fragmentation splits FDDI frames too large to transmit over Ethernet into two, three, or four smaller frames. A switch is a layer 2 device. It does not look at layer 3 or IP information during frame processing. For fragmentation, however, the module needs to have knowledge of layer 3 information. IP allows a packet to be sent in fragments as long as the fragmentation information is included among the IP headers. A FDDI packet is broken up into smaller pieces, each of which can be accommodated in an Ethernet packet. The module analyses the IP headers, sets fragmentation information in the IP control bits—fragment sequence numbers—and transmits the packet on the Ethernet port. The fragments are reassembled by the destination IP layer. For IP fragmentation to take place, the IP header's don't fragment flag should not be set on the source FDDI packet. Since the Ethernet MTU is smaller than that for FDDI, no fragmentation is necessary when switching packets from Ethernet to FDDI.

The module also supports MTU discovery. In the event the packet has to be fragmented in order to be successfully transmitted over Ethernet, but the IP header has the don't fragment flag set, the FDDI module returns a Destination-Unreachable message to the source. The module also indicates the size of the maximum transmission unit (MTU) as 1500 bytes in the same message. The source, having determined the size of the MTU for the next hop, can then suitably break up the packet and retransmit.

FDDI Automatic Packet Recognition and Translation

Some layer 3 protocols, allow more than one packet format to exist over the physical layer. IPX for instance allows four packet formats over Ethernet. AppleTalk allows two formats over Ethernet. The number of formats, however, are not constant over Ethernet and FDDI. There are only two legal formats for IPX over FDDI and one for AppleTalk over FDDI. The table below summarizes the packet formats.

Table 2 IPX and AppleTalk frame formats

	Ethernet	FDDI
IPX	—IEEE 802.2 —Ethernet II —IEEE 802.3 —IEEE SNAP	FDDI 802.2 FDDI SNAP
AppleTalk	Ethernet II IEEE SNAP	AppleTalk SNAP

When translating IPX packets from Ethernet to FDDI, the module converts all IEEE 802.2 packets to FDDI 802.2 packets. The module converts packets in the other three Ethernet formats to the FDDI SNAP format. When translating from FDDI to Ethernet, however, the module needs to know which one of the three Ethernet formats it should translate into. The module uses incoming frames to learn which format to use for a given Ethernet host. If the module has seen packets coming in from an Ethernet host in IEEE 802.3 format, then it will translate the FDDI SNAP packet destined for that host into an IEEE 802.3 packet.

Table 3 FDDI to Ethernet Translation table for IPX

FDDI	Ethernet
FDDI 802.2	IEEE 802.2
FDDI SNAP	Ethernet II
	IEEE 802.3
	IEEE SNAP

If the module sees a FDDI SNAP packet destined for an Ethernet host whose packet format it is unaware then it translates the incoming packet into all three formats—Ethernet II, 802.3 and SNAP—before transmitting. This situation can arise if the module has not seen any packets coming in from the destination Ethernet host before receiving the FDDI SNAP packet. Alternatively, the module provides the option of defining a default Ethernet frame format to be used.

When translating AppleTalk packets, the module converts IEEE SNAP packets on Ethernet to AppleTalk SNAP packets. Ethernet II and SNAP are incompatible at the Network level (layer 3), hence Ethernet II hosts cannot communicate with SNAP hosts. While the AppleTalk Ethernet II host cannot talk to an AppleTalk FDDI SNAP host, it can still use a AppleTalk FDDI backbone to communicate with another Ethernet II host. The module supports this by encapsulating Ethernet II packets according to the IEEE 802.1H specification before sending it over on FDDI—a process known as tunneling. The packets are decoded back to Ethernet II by the device connecting the FDDI backbone to the Ethernet network.

The FDDI module supports a Single Attachment Station (SAS) in which the module is connected only to the primary ring and a Dual Attachment Station (DAS) in which the module is connected to both the primary and the secondary FDDI rings. The second redundant counter-rotating ring offers added reliability. In the event that the primary ring fails, the module shifts to the secondary ring, isolating the failure and retaining ring integrity. The module also offers a CDDI interface—FDDI SAS over unshielded twisted pair (UTP) cabling.

Summary

The next generation Ethernet switches from Cisco Catalyst 1900/2820 Ethernet switches—are designed to optimize performance in bandwidth intensive networks and provide a rich feature set at an affordable price. The ClearChannel architecture offers wirespeed switching, extremely low latencies and nonblocking performance on all ports. The switches include Cisco IOS technology that permits tight integration with routers to offer CDP, broadcast storm control and layer 3 enhancements such as CGMP. The switches contain a high level of ASIC integration providing increased reliability at reduced complexity and cost. Based on third generation ClearChannel architecture—with a unique shared memory design and a forwarding engine implemented entirely in silicon—the Catalyst 1900/2820 Ethernet switches offer industry-leading price/performance value for desktop switching applications.

CISCO SYSTEMS

Cisco Systems
Corporate Headquarters
Cisco Systems, Inc.
170 West Tasman Drive
San Jose, CA 95134-1706
USA
World Wide Web URL:
http://www.cisco.com
Tel: 408 526-4000
800 553-NETS (6387)
Fax: 408 526-4100

European Headquarters
Cisco Systems Europe
s.a.r.l.
Parc Evolic-Batiment
L1/L2
16, Avenue du Quebec
BP 706-Villebon
91961 Courtaboeuf Cedex
France
Tel: 33 1 6918 61 00
Fax: 33 1 6928 83 26

Intercontinental Headquarters
Cisco Systems, Inc.
170 West Tasman Drive
San Jose, CA 95134-1706
USA
Tel: 408 526-7660
Fax: 408 526-4646

Latin American Headquarters
Cisco Systems, Inc.
790 N.W. 107th Avenue
Suite 102
Miami, FL 33172
Tel: 305 228-1200
Fax: 305 222-8456

Japanese Headquarters
Nihon Cisco Systems K.K.
Fuji Building
3-2-3 Marunouchi
Chiyoda-ku, Tokyo 100
Japan
Tel: 81 3 5219 6000
Fax: 81 3 5219 6010

Cisco Systems has over 190 offices in the following countries. Addresses, phone numbers, and fax numbers are listed on the Cisco Connection Online Web site at http://www.cisco.com.

Argentina • Australia • Austria • Belgium • Brazil • Canada • Chile • China (PRC) • Colombia • Costa Rica • Denmark • Finland • France • Germany Hong Kong • India • Indonesia • Ireland • Italy • Japan • Korea • Malaysia • Mexico • The Netherlands • New Zealand • Norway • Philippines Portugal • Singapore • South Africa • Spain • Sweden • Switzerland • Taiwan, ROC • Thailand • United Arab Emirates • United Kingdom • Venezuela