

BIGIRON ARCHITECTURE TECHNICAL BRIEF

Foundry Networks BigIron® product family offers hardware-based multifunction wire-speed switching and multi-protocol routing on a single, chassis-based platform. This platform supports a variety of interfaces including 10/100 Ethernet, Gigabit Ethernet Packet over Sonet (POS), and ATM interface modules. The BigIron product family includes a four-slot chassis, an eight-slot chassis, and a fifteen-slot chassis.

Contents

Introduction	2
Introduction	2
BigIron Family Overview	2
Performance Summary.....	4
Modules Available for the BigIron Family:	4
Switch Fabric Overview	4
Crosspoint Switch Fabric.....	5
Complete Crosspoint Switch System.....	7
Interface Module Architecture	8
Physical Ports.....	8
Multi-Port MAC	8
Packet Processor	8
Shared Memory Switch Fabric (and) Buffer Pool	9
Forwarding Engine.....	9
Module Connection to Crosspoint Switch Fabric.....	9
CP and FP.....	9
Multiple Destination Output Priority Queues.....	9
<i>Unicast Traffic</i>	11
<i>Multicast Traffic</i>	11
Multiple Input "Source" Buffers per Output Port	12
Packet Level Walk-Through (overview).....	12
Packet Level Walk-Through (detailed)	12
System Management Architecture	13
System Management Interface.....	13
Management Bus.....	14
Benefits of Foundry's IronCore™ Architecture	14
Protection from Head of Line Blocking.....	15
Ultra Low Latency.....	15
Efficient Multicasting Abilities	15
Multiple Priority Queues.....	15
Intelligent Queue Management	16
Wire-Speed Performance	16
Summary	16

Figures

Figure 1 – BigIron 4000 Feature Summary.....	2
Figure 2 – BigIron 8000 Feature Summary.....	3
Figure 3 – BigIron 15000 Feature Summary.....	3
Figure 4 – Crosspoint Switch Fabric Subsystems (based on BigIron 8000)	6
Figure 5 – Crosspoint Architecture Overview (Based on BigIron 4000).....	7
Figure 6 – Four Priority Queues Established per Destination Module.....	10
Figure 7 – Per Source Input Queuing	12
Figure 8 – Management Module Subsystem Diagram.....	14

BIGIRON ARCHITECTURE TECHNICAL BRIEF

Introduction

To provide a more detailed understanding of Foundry's architecture, the following sections describe each individual system component. Although the BigIron family consists of three different chassis configurations (four slot, eight slot, and fifteen slot), they all share a common architecture. The information is organized into three major sections:

- BigIron Product Family Overview
- Switch Fabric Architecture
- System Management Architecture
- Interface Module Architecture

BigIron Family Overview

The Foundry Networks BigIron family is designed for use in LAN backbone and Gigabit aggregation in both Enterprise and Service provider networks. BigIron delivers high-density 10/100/1000 Mbps Ethernet, as well as OC3/12/48 Packet over SONET with performance of up to 178,000,000 packets per second.

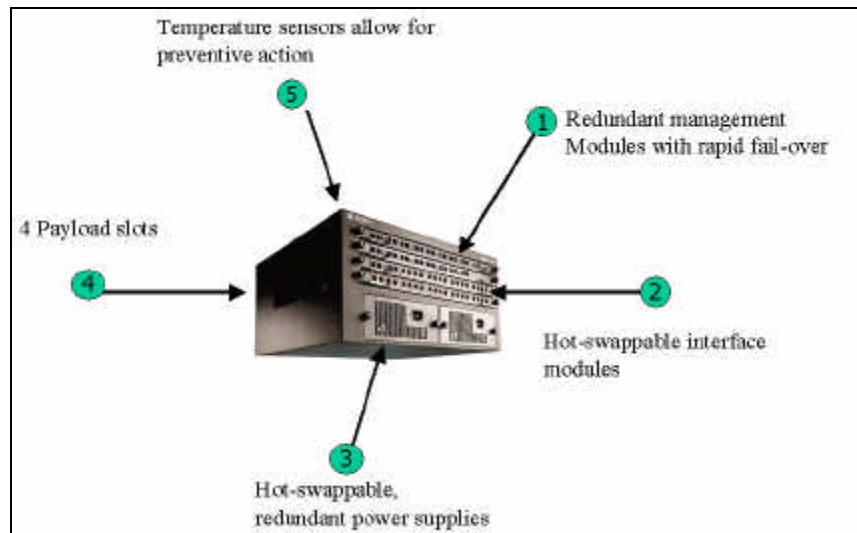


Figure 1 – BigIron 4000 Feature Summary

BIGIRON ARCHITECTURE TECHNICAL BRIEF

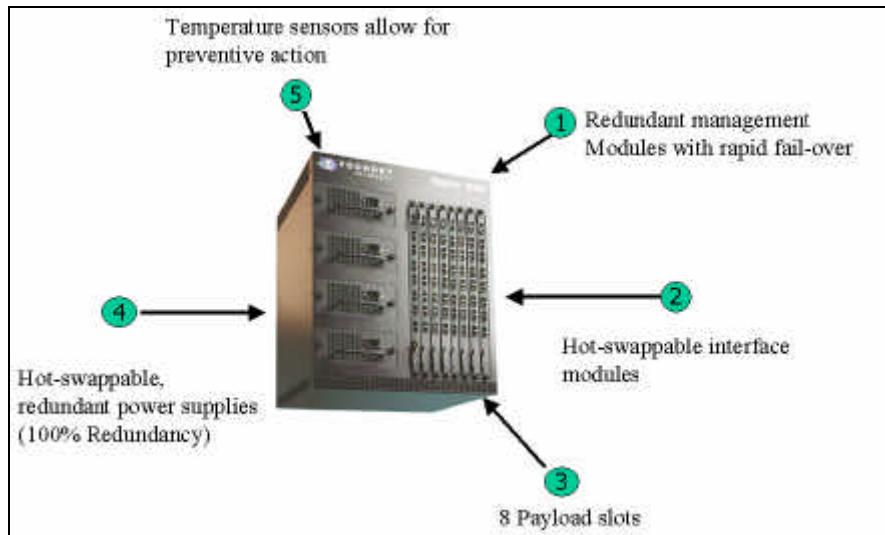


Figure 2 – BigIron 8000 Feature Summary

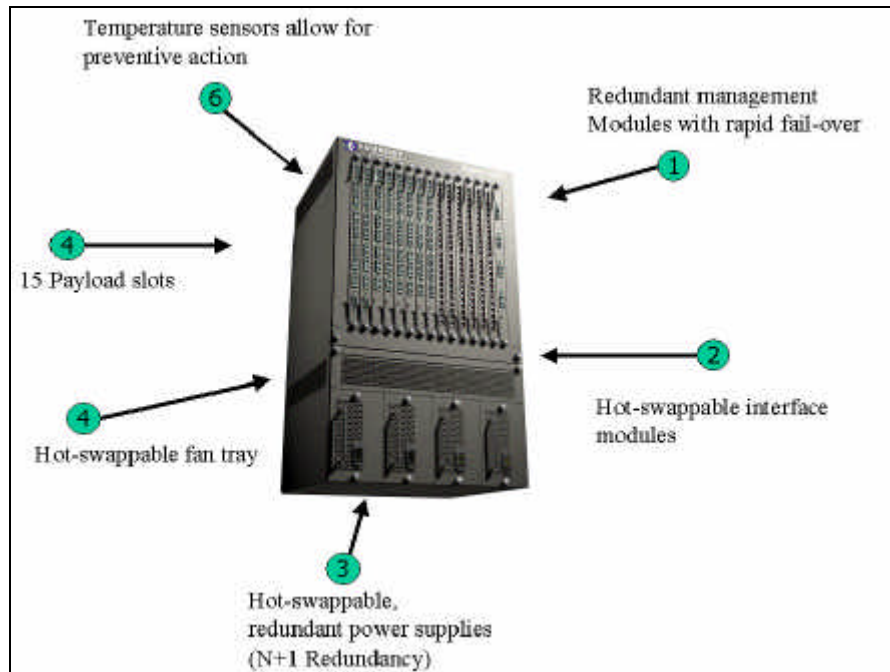


Figure 3 – BigIron 15000 Feature Summary

BIGIRON ARCHITECTURE TECHNICAL BRIEF

Performance Summary

	BigIron 4000	BigIron 8000	BigIron 15000
<i>Aggregate Switching Capacity (Gbps)</i>	120 Gbps	240 Gbps	480 Gbps
<i>Switching Throughput (based on line-rate 64 byte frames)</i>	47 Mpps	96 Mpps	178 Mpps

Modules Available for the BigIron Family:

Management IV Modules:

- 4 port GBIC based Gigabit Ethernet Management IV Module

Management III Modules:

- 0 port Management III Module
- 8 port Management III Module

Management II Modules:

- 0 port Management II Module
- 8 port Management II Module

Host Modules:

- 24 port 10/100 Mbps (RJ-45) Ethernet Module
- 24 port Fiber 100 Mbps (MT-RJ) Ethernet Module
- 8 port Mini-GBIC based Gigabit Ethernet Module
- 2 port Gigabit Ethernet Modules
- 4 port Gigabit Ethernet Modules
- 8 port Gigabit Ethernet Modules
- 2 port OC-3c Packet Over SONET (POS) Module
- 2-port OC-12c Packet Over SONET (POS) Module
- 2-port OC-48c Packet Over SONET (POS) Module

Switch Fabric Overview

The core of the BigIron architecture is an extremely efficient, high-speed parallel crosspoint switch fabric. It includes a high speed ASIC that provides a switching capacity supporting up to 178,000,000 packets per second of data throughput. Foundry's custom designed ASICs reside on each module to provide 32 Gbps of local switching capacity and connection into and out of the crosspoint fabric via eight-Gbps full-duplex data paths. Multiple quality-of-service (QoS) queues on each module and inside the crosspoint switch fabric ensure that prioritized traffic flows unencumbered from source to destination. A fully populated four-slot, eight-slot or fifteen-slot chassis delivers 128 Gbps, 256 Gbps, and 480 Gbps respectively, of total switching capacity.

The combination of local switching interface modules and non-blocking crosspoint backplane comprises Foundry's IronCore™ architecture, which delivers a two level switching system that easily scales up to 120 Gigabit Ethernet ports. To ensure compatibility with existing infrastructures and provide long-term investment protection, the IronCore architecture supports interfaces ranging from 10/100/1000 Mbps to 10 Gbps Ethernet, OC3/12/48 Packet over SONET, and OC3 ATM.

BIGIRON ARCHITECTURE TECHNICAL BRIEF

Crosspoint Switch Fabric

All Foundry products are based on a common architecture. Core to that architecture is the Crosspoint Switch Fabric. Figure 4 on page 6 illustrates Foundry's switch fabric architecture that is implemented in the BigIron 8000. As shown in the diagram, the switch fabric provides for multiple connections from each interface slot to every other slot within the chassis, which can be described as a mesh. By providing this *mesh* of multiple paths between slots, it allows for multiple simultaneous flows to be "switched" without having to arbitrate for a single path, or to be controlled by a single management function that can cause "blocking" within the switch fabric itself. Another benefit of providing a crosspoint switch fabric is that there is no central points or reserved paths that are required for specific functions such as management (that can also impact the performance of a switch fabric). This also allows for management interface modules to be inserted into any available slot within the chassis without having to be concerned with pre-defined slot reservation.

The arrows illustrated in Figure 4 show that the connection between each interface module slot and the crosspoint switch fabric is an 8 Gbps, full-duplex data path. This provides a non-blocking throughput regardless of interface module used.

Although Figure 4 illustrates the eight slot chassis (BigIron 8000), the same architecture is applied to the rest of the products within the BigIron family. The difference is in the amount of paths available within the crosspoint switch fabric and slots available. These vary depending on the chassis, as follows:

- BigIron 4000 has four (4) slots
- BigIron 8000 has eight (8) slots
- BigIron 15000 has fifteen (15) slots

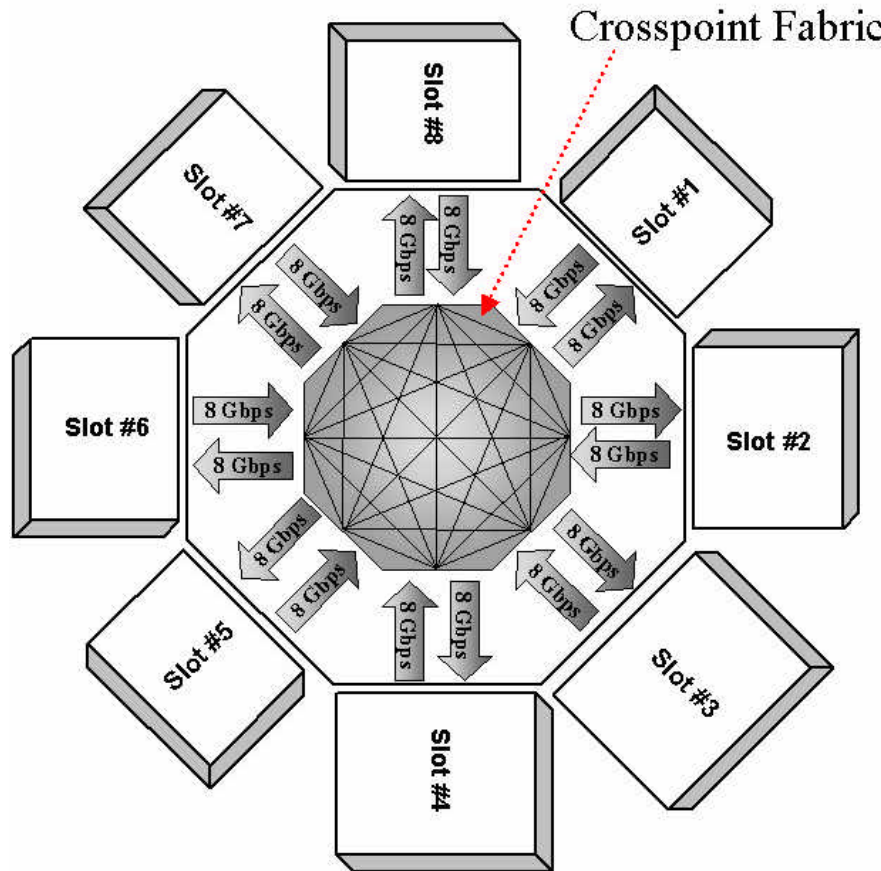
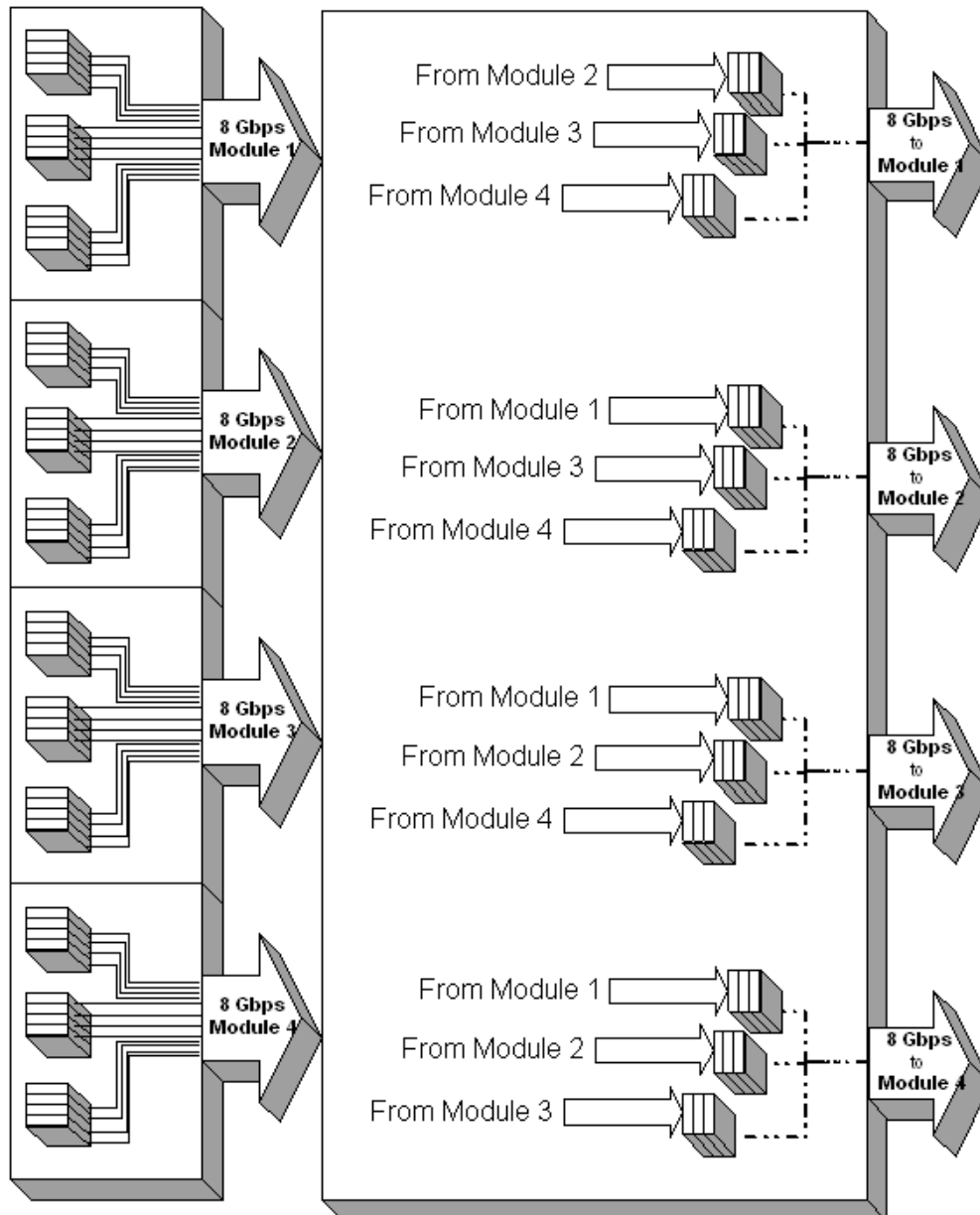


Figure 4 – Crosspoint Switch Fabric Subsystems
(based on BigIron 8000)

BIGIRON ARCHITECTURE TECHNICAL BRIEF

Complete Crosspoint Switch System

Figure 5 below illustrates the crosspoint switching fabric implemented in a fully populated BigIron 4000 architecture. Missing are the four priority queues for the local physical ports on each module (omitted to reduce the complexity of the picture). The arrows along the left side represent the four modules with output priority queues feeding the 8 Gbps connection from each slot into the switch fabric. The center picture represents the “input buffered” crosspoint switch fabric that provides fully meshed slot interconnectivity allowing parallel packet flow between slots. The arrows along the right side represent the four 8 Gbps connections out of the switch fabric to each slot.



**Figure 5 – Crosspoint Architecture Overview
(Based on BigIron 4000)**

BIGIRON ARCHITECTURE TECHNICAL BRIEF

Interface Module Architecture

Foundry Networks' BigIron modules share a common architecture that delivers leading price/performance, scalability, and QoS to both switched and routed environments. This section provides the component level definitions for interface modules including a packet level walk through for BigIron switching routers.

Physical Ports

Physical ports provide connectivity to end stations or other network devices. Foundry's products offer a full range of auto sensing, half or full-duplex switch and router port options for Ethernet, Fast Ethernet, and Gigabit Ethernet connectivity, including 10Base-T/100Base-TX, 100Base-FX, 1000Base-SX, 1000Base-LX, and 1000Base-LH, as well as Packet Over SONET and ATM interface options.

Multi-Port MAC

Multi-port media access controllers (MAC) provide continuous data flow between groups of physical switch/router ports (eight ports for 10/100 Mbps and two for Gigabit Ethernet) and the packet processor. Separate pipelines (to both the physical ports and packet processors) for transmit and receive ensure wire-speed full-duplex throughput at line rates (up to 4 Gbps) for both switching and routing.

Packet Processor

Packet processors, operating at full line rate (1.488 million packets per second per port for Gigabit and 148,800 pps per port for 10/100 Ethernet based on 64 byte frames) parse and examine packets at Layers 2, 3, and 4. Separate pipelines for transmit and receive to both the multi-port MAC and switch fabric guarantee full line rate, wire-speed operations.

The port or system-level configuration controls how deeply the packet processor looks into each packet during examination. The packet processor takes the information found during the examination and runs a comparison in its cache. A cache "hit" (or matched comparison) means that the packet processor's cache contains additional forwarding information regarding the disposition of the packet. The forwarding information, or FID, may include details such as the destination port, port mirror requirement, packet type, VLAN handling, prioritization, multicast group membership, and/or other BigIron programmable features. A cache "miss" (or non-matched comparison) means that the packet processor doesn't "know" anything about the packet, which results in a FID that directs the packet to the system manager interface for additional processing.

For ports or system-level configurations enabled for Layer 3, the FID may contain additional information regarding the IP header (including length, protocol type, and destination MAC). For Layer 3 tasking, the packet processor also executes standard routing functionality at line rate, including substitutions of next hop MAC addresses, time to live (TTL) decrement, and recalculation of the header checksums. When enabled for Layer 4, the FID may include information pertinent to a flow specification, including source and/or destination address, socket, or protocol type.

BIGIRON ARCHITECTURE TECHNICAL BRIEF

The packet processor's cache table values are managed, updated, and controlled by the system manager interface found on the management board.

Shared Memory Switch Fabric (and) Buffer Pool

Each BigIron module includes a shared memory pool and custom designed switch fabric ASIC providing 64 Gbps (a 512-bit data path operating at 133 MHz) of raw bandwidth for local on-board packet switching and storage. All the ports and packet processors on a module share and have equal access to the resources of the shared memory switch fabric and buffer pool.

The true data rate through the switch fabric is 32 Gbps, which is half the raw switching bandwidth because packets traverse the switch fabric twice (into and out of the shared memory). Separate pipelines to and from each packet processor for transmit and receive ensure full line rate, wire-speed operations.

Forwarding Engine

Based on the packet's shared memory location id (SMID) and forwarding instructions contained in each FID, this custom designed switch ASIC ultimately controls packet movements in and out of the shared memory switch fabric buffer pool. The forwarding engine uses the FID to provide the basic mechanics involved in directing unicast and broadcast traffic from a single packet located in memory. Finally, to allow for local co-ordination of FID information, the forwarding engine connects to every packet processor on a given module through a separate data path.

Module Connection to Crosspoint Switch Fabric

A full-duplex 8 Gbps backplane interface connects each module into the cross-point switch matrix, which allows for point-to-point simultaneous communications between modules. Please refer to the section called "Crosspoint Switching Backplane" of this technical brief for further information.

CP and FP

These paths represent connections between the system interface processor /central CPU and the shared memory switch fabric and forwarding engine and shared memory switch fabric. Packets that require additional processing have FIDs tagged for the CPU. The CPU accesses the packets in their memory location through this path.

Multiple Destination Output Priority Queues

Figure 6 below depicts a module in slot 1 of a fully populated BigIron 4000. The illustration is simplified by displaying only the priority queues and shared memory switch fabric. Also shown is the 8 Gbps connection to the crosspoint switch fabric.

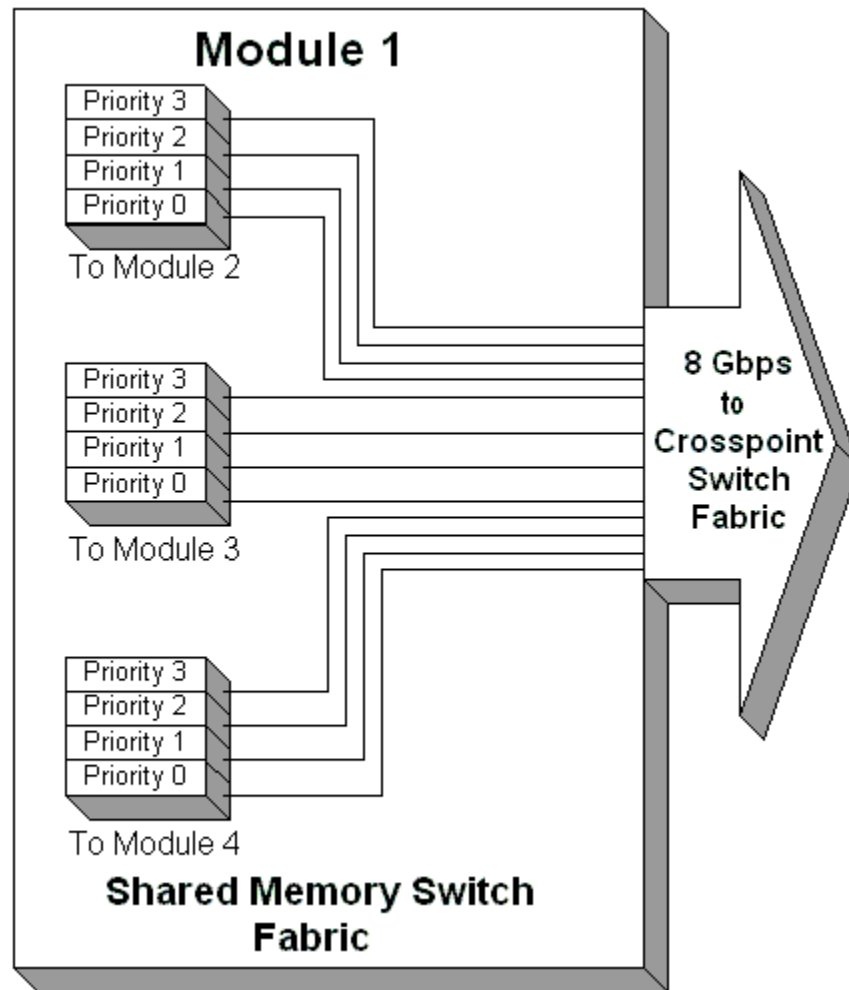


Figure 6 – Four Priority Queues Established per Destination Module

Each module has four priority output queues assigned for every possible destination output port. Destination output ports are either the local physical ports on the module or connections to other modules in the chassis. Figure 5 on page 7 represents four priority queues for the three other modules in slots two, three, and four. Omitted from the diagram are the four queues for the local physical ports.

Based on this methodology, a single BigIron module could manage up to 124 individual priority queues. A 24-port 10/100 board has 96 priority queues (24-ports x 4 queues – the other 23-ports on the module plus 1 for the CPU) plus 28 priority queues (for the seven other modules).

BIGIRON ARCHITECTURE TECHNICAL BRIEF

Unicast Traffic

Based on the information contained in the FID, the forwarding engine places the SMID into the appropriate priority queue. The packet sits in the shared memory buffer pool until its SMID reaches the head of the queue. Once there, the SMID triggers the forwarding engine to deliver the packet to the crosspoint switch fabric for delivery to the destination port or module.

At the destination module, the forwarding engine reads the FID, moves the packet into shared memory and delivers the new SMID to the destination packet processor. Based on the FID, the packet processor moves the packet out of memory and delivers it to the destination port. When delivered, the forwarding engine clears the packet from memory.

Multicast Traffic

With multicast traffic, the FID pre-identifies all members of the multicast group and their destination ports. If the members of the multicast group exist on multiple modules, the forwarding engine places one SMID (per module) into the appropriate output priority queue for each destination module. The same process takes place as previously described for unicast traffic with some minor modifications. The packet sits in the shared memory buffer pool until its SMID reaches the head of the queue, which triggers the forwarding engine to deliver the packet to the crosspoint switch fabric.

Only one copy of the packet travels to each destination module, so if multicast members exist on three modules, three copies (one per module) of the packet move through the switch fabric. Only after all modules have received their copy of the packet does the ingress module clear the packet from the shared memory switch fabric. In other words, BigIron employs a counter based on the number of modules, not ports, involved with the multicast group. When the counter reaches zero this means that all modules have received their copy of the packet and the originating module removes the packet from its shared memory switch fabric.

At the destination module (or modules), the forwarding engine reads the FID identifying the port that is a member of the multicast group. The packet is stored in the local memory (indexed by a new SMID) and the forwarding engine delivers the SMID to the destination packet processor. Based on the FID, the packet processor sends the packet to the destination port involved in the multicast group. When delivered, the forwarding engine clears the packet from memory.

In the event that *multiple ports on a module* belong to the multicast group, the forwarding engine initializes a counter equaling the number of ports for the multicast group. The packet processor copies the packet from memory (based on the SMID) and decrements the counter for each delivered packet. The counter reaches zero when all ports receive a copy of the multicast packet from the local memory. At this point, the forwarding engine clears the packet from memory.

BIGIRON ARCHITECTURE TECHNICAL BRIEF

Multiple Input “Source” Buffers per Output Port

Figure 7 shows a subsection of the crosspoint switch fabric from a fully populated BigIron 4000. For simplicity, shown is a small portion of the crosspoint switch fabric along with the 8 Gbps output port to module 1.

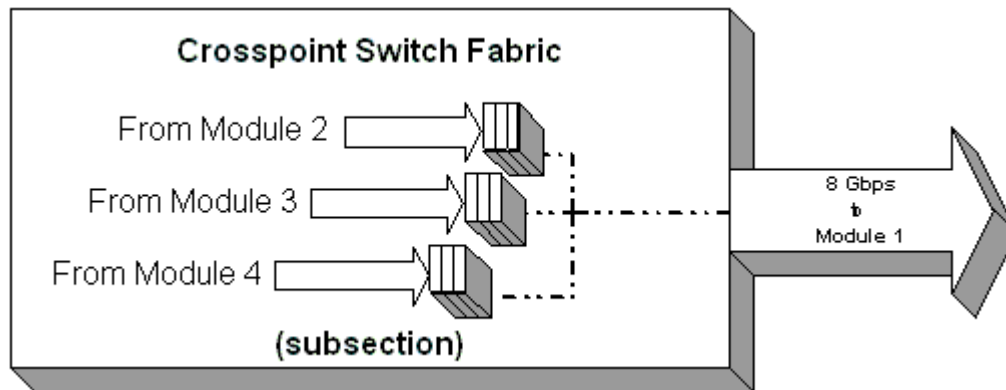


Figure 7 – Per Source Input Queuing

This portion of the crosspoint switch fabric details three input paths connecting to their respective buffers. The three buffers in turn connect to the 8 Gbps output path to module one. A “round robin” algorithm for the input buffers allows packets to move directly from one buffer at a time to the output port. This gives each module a chance to deliver a packet to its intended destination and ensures that no one module dominates an output port.

Packet Level Walk-Through (overview)

Depending on how one configures the chassis (as a Layer 2, Layer 3, Layer 4 switch, or multi-protocol router) each packet processor examines and analyzes up to the first 64 bytes of every received packet. During examination, the packet processor retrieves a forwarding identifier (FID), which contains the “handling” instructions for the packet based on the enabled feature set (configuration) for the port and system. While the packet processor examines the first 64 bytes, the packet moves into an indexed shared memory buffer location id (SMID), with the SMID, *not the packet*, moving into one of the four priority queues as requested by the “handling instructions” of the FID. Only when the SMID reaches the front of its queue does the packet move to the appropriate output port, again based on its FID.

The packet is forwarded to either one (or more) of the local ports or directly to the crosspoint switch fabric via the module’s 8 Gbps full-duplex backplane connection. All modules work independently of each other, which means that while a module sends packets to its local ports and backplane it can receive packets from a different module in parallel (further detailed in the section – Crosspoint Switching Backplane Overview).

Packet Level Walk-Through (detailed)

The packet processor parses the packet and performs a “look-up” operation. A successful look-up yields a forwarding identifier, which provides all of the “transmission” decision data for the packet. An unsuccessful look-up yields a FID that directs the packet to the CPU for additional

BIG IRON ARCHITECTURE TECHNICAL BRIEF

processing. It is the contents of the FID and not the packet itself, which is the basis for all filtering and forwarding activity. During packet examination, the packet and FID header together move into the shared memory switch fabric.

The forwarding engine reads the FID and moves the packet to a shared buffer memory location identified by a shared memory identifier (SMID). Based on the information gleaned from the FID, the forwarding engine places the SMID into one of the four priority queues for the intended output port or ports (see Multiple Destination Output Priority Queues for more information).

When the SMID gets to the front of its queue, the forwarding engine moves the packet from the shared buffer memory pool into the switching fabric for delivery to its final destination. Final destinations are either to local ports (to another port on the same module), to remote ports (to a different module) or to the central CPU (for additional processing).

With a local FID, the forwarding engine delivers the packet from its memory location to the packet processor associated with the intended destination port. With a CPU FID, the delivery process is the same as a remote FID but the packet is destined for the management module.

With a remote (or CPU) FID, the forwarding engine moves the packet (based on its SMID) from memory into the crosspoint switch (further described later in this brief). The packet arrives at the destination module and the destination forwarding engine reads the FID while the packet moves into a new shared memory buffer pool location (indexed by a new SMID). Using the SMID, the forwarding engine delivers the packet from its shared memory location directly to the packet processor associated with the intended destination port. (This final step in the destination module is exactly the same process described when handling a "local FID.")

Because packets move into and out of the shared memory buffer pool through the shared memory switch fabric, the switch fabric must provide enough bandwidth to avoid becoming a bottleneck. To ensure that the switch fabric on each module is truly non-blocking, its 32 Gbps of internal bandwidth is more than twice the aggregated total port bandwidth of the eight-port Gigabit Ethernet host module.

System Management Architecture

The System Management functions are achieved through the addition of a Management Module to the chassis, which provide the operating environment and feature set available to the user. A benefit of providing an "intelligent" switch fabric (Crosspoint switch fabric), is the ability to provide a full payload of ports on the management module (up to 8 x 1 Gbps ports) which increase the port density and versatility of the products. The following sub-sections describe the components within the Management architecture.

System Management Interface

The system management interface ASIC acts as the central "manager" of the chassis, storing the port and system-level master configuration tables, routing tables, Layer 2, Layer 3, Layer 4 address tables, and all FID registers.

Packets that require additional processing, handling, or routing, are directed to the CPU (central processing unit) portion of the system management interface. The CPU accesses the packets in the shared memory location (as noted by their shared memory identification location or SMID) through a dedicated connection into the shared memory switch fabric. Every packet processor

BIG IRON ARCHITECTURE TECHNICAL BRIEF

communicates with the system management interface (found only on a management module) through the management bus (see Management Bus below). Updated FID entries for packets that required additional handling transfer back to the originating packet processor's cache through this path.

The system management interface also provides the control logic needed to manage all of the local packet processors and acts as the hardware control interface for the CPU.

Figure 5 illustrates the sub-components of a management module, which is identical to an interface module with the exception of the Central CPU and System Management Interface.

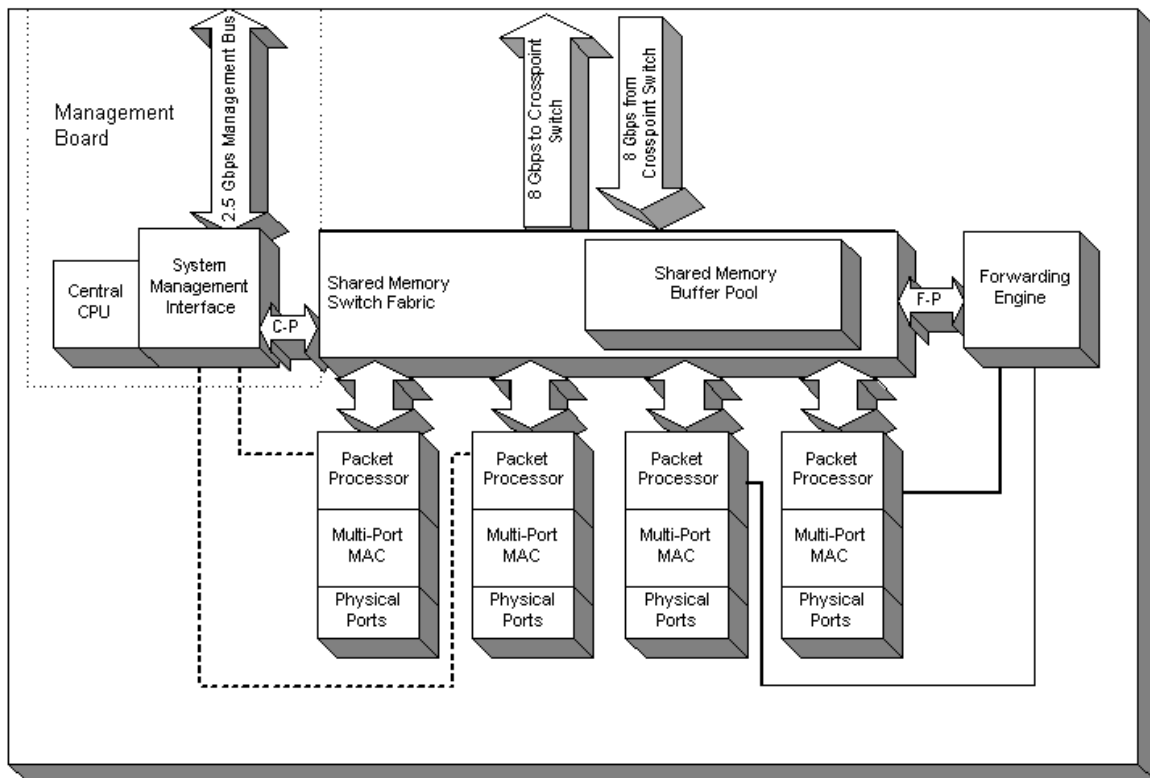


Figure 8 – Management Module Subsystem Diagram

Management Bus

A 2.5 Gbps dedicated bus for management connects the management module to all the other modules and packet processors in the chassis. It helps deliver the information associated with chassis configuration and control as well as FID updates from packets that required additional CPU intervention.

Benefits of Foundry's IronCore Architecture

The IronCore architecture, which combines Foundry's shared memory switch fabric and crosspoint switch, delivers an unbeatable, high performance and extensible foundation for your networking infrastructure.

BIGIRON ARCHITECTURE TECHNICAL BRIEF

Protection from Head of Line Blocking

Head of line blocking occurs when a source port transmits data to multiple destination ports, including one that is busy. The source is restricted from sending data to all other non-congested destinations because the congested destination holds up (or blocks) all other data streams intended for the non-congested destinations. Only when the congested destination is clear can all traffic flow (providing that one of the other destinations does not become congested). Typical crossbar architectures suffer from this problem because they do not have multiple, parallel paths between source and destination ports.

Another type of head of line blocking occurs when low priority packets whose destination is congested blocks high priority packets needing to traverse the switch fabric. Typical crossbar architectures suffer from this type of problem because they do not have multiple priority queues.

The combination of multiple destination output priority queues combined with multiple input source buffers per output port effectively avoids both these “head of line” blocking conditions. When one source port transmits to multiple destination ports, Foundry’s cross-point switch fabric provides simultaneous parallel paths from the source to destination ports. One congested destination port does not affect any of the other non-congested destination ports.

In the event that multiple source ports transmit to one destination port, the “round robin” algorithm working at the multiple input source buffer location ensures that all source ports deliver packets to the destination port. No single source port can dominate any output port.

Ultra Low Latency

BigIron’s IronCore architecture offers reduced latencies and increased efficiencies because the shared memory switch fabric copies packets into memory only once and references their SMID for later transmission. The typical local port-to-port latency is approximately five *microseconds*, while port-to-backplane-to-port is approximately 6.5 *microseconds*. Other switches and routers must copy packets from one memory location to another or from input to output port buffers, thereby consuming precious resources, decreasing available bandwidth and increasing measured latency.

Efficient Multicasting Abilities

IronCore supports effective handling of broadcast and multicast packets because multiple ports read and transmit a single given area of memory (i.e., packet), eliminating the need to make multiple copies. This reduces the amount of time and resources normally associated with multicast traffic.

Multiple Priority Queues

The four priority queues of the IronCore architecture provide better prioritization because high priority packets, flagged and identified by their FIDs always bypass low priority packets (also flagged and identified). Other systems require multiple data paths with numerous output queues to duplicate this feature, resulting in increased total cost, added system management resources, and reduced system efficiency.

BIGIRON ARCHITECTURE TECHNICAL BRIEF

Intelligent Queue Management

IronCore's advanced software queue size control mechanisms ensure that a single port does not consume all of the memory resources for a single destination port. This feature is not part of conventional crossbar architectures.

Wire-Speed Performance

The BigIron is the only product in the marketplace with sustainable wire-speed performance for all ports (10/100/1000 Mbps) on a module, both for local switching and across the backplane for Gigabit Ethernet densities up to 64 ports. For local switching, the switch fabric performance based on the 64 Gbps raw switch fabric bandwidth equates to a true data rate of 32 Gbps per BigIron module.

Summary

In conclusion, Foundry's IronCore architecture delivers the industry's highest throughput and non-blocking Gigabit Ethernet port density. The two-level switching system provides up to 178,000,000 packets per second throughput with 480 Gbps of switching capacity. It offers the highest Gigabit Ethernet port density in the marketplace, with Layer 2/3/4 switching concurrent with wire-speed routing. Packaged with many advanced features, the architecture is the first to be 10 Gigabit ready with 8 Gbps per module capacity. The BigIron product family provides the "next generation" of campus, MAN and LAN/WAN networking for Enterprise and Internet Service Providers. IronCore architecture, available only from Foundry Networks, combines price, performance, and multi-layer feature leadership ensuring QoS, security and reliability for Ironclad Network Performance™.

Foundry Networks, Inc.
Headquarters
2100 Gold Street
P.O. Box 649100
San Jose, CA 95164-9100

U.S. and Canada Toll-free: (888) TURBOLAN
Direct telephone: +1 408.586.1700
Fax: 1-408-586-1900
Email: info@foundrynet.com
Web: <http://www.foundrynet.com>

© 2001 Foundry Networks, Inc. All Rights Reserved.