BigIron RX architecture brief

The BigIron® RX Series is the industry's most powerful family of Layer 2/Layer 3 Ethernet switches. IP networks now play a critical role in ensuring satisfactory end-user experience, irrespective of the application being run on the network. It is therefore imperative to ensure non-stop operation, high performance, low latency and QoS guarantees for the respective applications that are concurrently running on the network. Further, the ability to adapt to evolving technologies such as 40-Gig or 100-Gig interfaces in the not too distant future has imposed a need for administrators to install future-proof systems. The BigIron RX is a cost-efficient and future-proof solution that has been specifically architected to address the build-out of enterprise and service provider networks around today's technologies such as 10 Gigabit Ethernet, and IPv4/IPv6 while allowing growth into evolving technologies such as 40-Gig or 100-Gig interfaces. Its robust system architecture, versatile feature set and availability in 3 different sizes allows network designers to standardize on a single product family for needs ranging from aggregation to backbone switching.

Designed with state of the art packet processing technology, the BigIron RX has a non-blocking switching capacity of up to 3.84 Tbps and routing performance of up to 1.14 billion packets per second. Its advanced distributed hardware architecture with fine-grained QoS support allows uncompromised full-duplex, wire-speed performance to be achieved for any mix of IPv4, IPv6 and Layer 2 services. These capabilities are made possible by an innovative system architecture that has several distinguishing characteristics:

- Clos-based[1] self-routing, distributed, non-blocking architecture provides the foundation for a robust, scalable platform
- Distributed packet processing and advanced QoS capabilities across the system allow a rich set of features to be implemented at wire-speed rates
- High availability architecture with a clear separation between control and data planes
- Fully redundant architecture with redundant power supplies, management modules, fan trays and switch fabric modules to avoid any single point of failure

The BigIron RX is available in three different configurations:

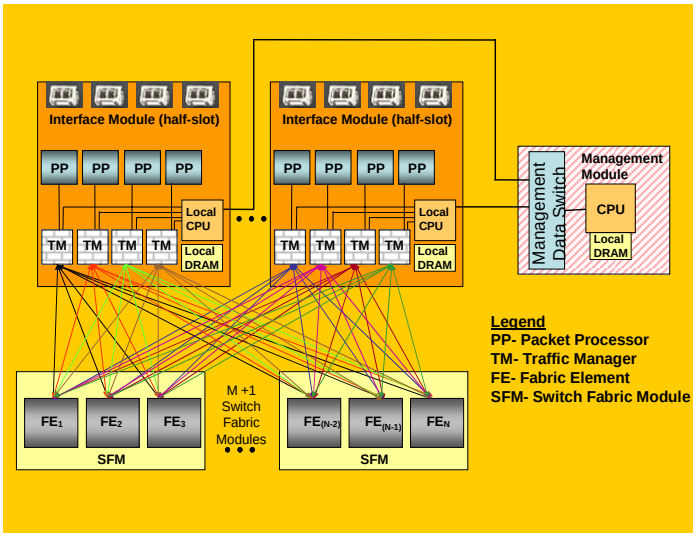
- BigIron RX-16, a 14 RU, 16 interface-slot system
- BigIron RX-8, a 7 RU, 8 interface-slot system
- BigIron RX-4, a 4 RU, 4 interface-slot system

All modules on the BigIron RX are hot pluggable. The management and interface modules can be interchangeably used across any of these systems, thereby decreasing inventory and maintenance costs for network administrators.

Industry-leading density

The BigIron RX is scalable to an industry leading density of 64 non-blocking 10 Gigabit Ethernet ports or 768 non-blocking Gigabit Ethernet ports in a single chassis. In a standard 7' rack, the BigIron RX can support up to 192 10 Gigabit Ethernet ports or 2,304 Gigabit Ethernet ports. These capabilities give the BigIron RX an extraordinary density of 55 Gigabit Ethernet ports per rack unit (RU) and 110 Gbps of full-duplex switching capacity per RU.

Scalable Clos fabric architecture



The BigIron RX uses a Clos fabric architecture that provides a high level of scalability, redundancy and performance. As shown in the Figure above, there are multiple switch fabric

[1] Named after the groundbreaking work by researcher Charles Clos, the Clos architecture has been the subject of much research over several years. The resiliency of this architecture makes it the ideal building block in the design of high availability, high performance systems.

modules (SFMs) in the system. An SFM has multiple fabric elements, each of which has multiple connections to every interface slot.

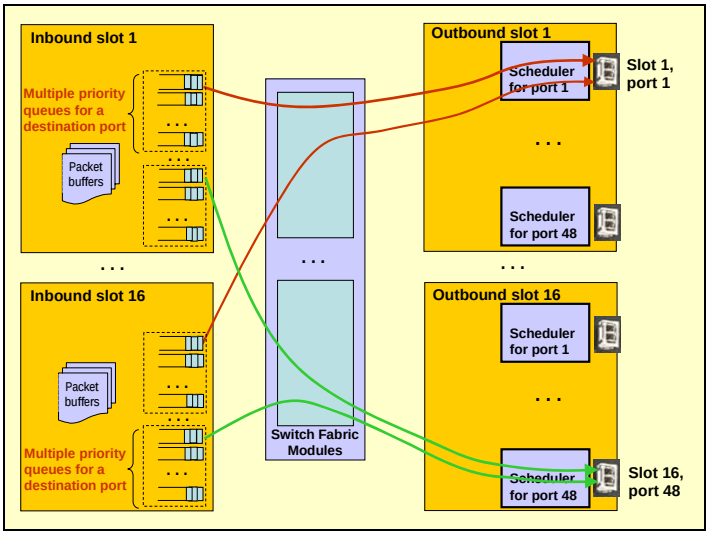
The Clos architecture uses data striping to ensure optimal utilization of fabric interconnects at all times. This mechanism always distributes the load equally across all available links between the input and output interface modules. By using fixed-size cells to transport packets across the switch fabric, the BigIron RX's switching architecture ensures predictable performance with very low and deterministic latency and jitter for any packet size. The presence of multiple switching paths between the input and output interface modules also provides an additional level of redundancy.

There are several advantages of a Clos architecture over a traditional architecture:

- **Common architecture across the product family** as the same fabric elements are used on all three chassis of the BigIron RX Series. This demonstrates the superior scalability of the architecture from a small 4 slot system to a large 16 slot system.
- **No head-of-line blocking** at any point irrespective of traffic pattern, packet size or type of traffic.
- **Optimal utilization of switch fabric resources** at all times. The data striping capability ensures that there is fair utilization of the switch fabric elements at all times without overloading of any single switch fabric element.
- **"Intra-SFM" redundancy**: An SFM can withstand the failure of some of the fabric elements and yet continue to operate with the remaining fabric elements. This unique capability provides a very high level of redundancy even within an SFM.
- **Exceptional high availability**: The BigIron RX SFMs have (N+1) redundancy. This allows the BigIron RX to gracefully adapt to the failure of multiple switch fabric elements. Moreover, because there are multiple fabric elements within an SFM, the failure of a fabric element does not bring down the entire SFM.

Distributed queuing for fine-grained QoS

A distinguishing characteristic of the BigIron RX architecture is the use of a distributed queuing scheme that maximizes the utilization of buffers across the whole system during congestion. This scheme marries the benefits of input-side buffering (Virtual Output Queuing) with those of an output-port driven scheduling mechanism. Input queuing using virtual output queues ensures that bursty traffic from one port does not hog too many buffers on an output port. An output-port driven scheduling scheme ensures that packets are sent to the output port only when the port is ready to transmit a packet. Each interface module maintains multiple, distinct priority queues to every output port on the system. Packets are "pulled" by the outbound interface module when the output port is ready to send a packet. Switch fabric messaging is used to ensure that there is tight coupling between the two stages. This closed loop feedback between the input and output stages ensures that no information is lost between the two stages. The use of such "virtual output queues" maximizes the efficiency of the system by storing packets on the input module until the output port is ready to transmit the packet. In all, there are 512K virtual output queues on the BigIron RX chassis that are distributed across the system.



Complementing the large number of virtual output queues is a rich set of scheduling algorithms that can be applied to each Egress port. These mechanisms can either be used individually or in combination to deliver tiered QoS guarantees for several applications on the same port:

- Strict priority
- Enhanced strict priority scheduling
- Weighted Fair Queue (WFQ) destination-based scheduling

- Weighted Fair Queue (WFQ) source-based scheduling
- Maximum rate-based scheduling
- Minimum rate-based scheduling

For example, multiple queues can be serviced in the order of decreasing priority in this architecture. These advanced QoS capabilities make the BigIron RX architecturally superior to existing legacy architectures.

Congestion avoidance is handled by applying Weighted Random Early Discard (WRED) or tail-drop policy.

The QoS subsystem on the BigIron RX has extensive classification and packet marking capabilities that can be configured:

- Prioritization based on Layer 2 (802.1p), TOS, or DSCP of an input packet
- Mapping of packet/frame priority from Ingress encapsulation to Egress encapsulation
- Remarking of a packet's priority based on the result of the 2-rate, 3-color policer.

IPv6 capabilities

The increasing migration to converged network architectures and the ubiquity of IP-enabled user devices makes the eventual migration to IPv6 inevitable. To ease this migration, the BigIron RX supports IPv6 over IPv4 tunnels. Every interface module of the BigIron RX has native hardware support for dual-stack IPv4/IPv6 routing with a route information base capacity of 200K IPv6 routes. The hardware FIB per interface module allows wire-speed, full-duplex IPv6 performance to be achieved even at full capacity. Each interface module also supports IPv6 ACLs in hardware. IPv6 routing protocols that are supported today include MP-BGP-4, OSPFv3, IS-ISv6 and RIPng. These capabilities make the BigIron RX an ideal system for IPv6 networks.

High Availability

Both the hardware and software architecture of the BigIron RX are designed to ensure very high Mean Time Between Failures (MTBF) and low Mean Time To Repair (MTTR). Cable management and module insertion on the same side of the chassis allows ease of serviceability when a failed module needs to be replaced or a new module needs to be inserted.

The ability to handle the failure of not only an SFM but also elements *within* an SFM ensures a robust, redundant system ideal for non-stop operation. The overall system redundancy is further bolstered by redundancy in other active system components such as power supplies, fans, and management modules. The passive backplane on the BigIron RX chassis increases the reliability of the system. Self-adjusting, variable-speed fans in the BigIron RX chassis help to maintain the optimal operating temperature. Cards that cross a pre-set temperature threshold will be powered off by the management module so as not to completely disrupt switch function.

The BigIron RX also supports the ability to gracefully shut down a switch fabric module with zero packet loss for a scheduled maintenance event. When this facility is invoked, the system will not use the links between the interface modules and the decommissioned SFM.

The modular architecture of IronWare operating system has several distinguishing characteristics that differentiate it from legacy operating systems:

- Industry-leading cold restart time of less than a minute
- Support for hitless software upgrade
- Hitless Layer 2 and Layer 3 failovers
- Sub-second switchover to the standby management module if a communication failure occurs between active and standby management modules.
- Support for graceful OSPF restart and graceful BGP restart

Distributed forwarding for wire-speed performance at any packet size

The BigIron RX has a distributed forwarding architecture that combines state of the art packet processing technology with a very fast switch fabric to ensure uncompromised, full-duplex, wire-speed performance at any packet size. These packet processors implement a variety of features such as access control lists, policing, classification, and multicast replication. The use of fast packet processors on each interface module allows wire-speed performance to be maintained, independent of the features that have been enabled.

There are several capabilities that have been implemented in the IronWare® operating system software to facilitate distributed packet forwarding and security:

- **Distributed maintenance of the Layer 2 MAC address table on each interface module**: The management module maintains all the learned MAC addresses and distributes the information to be locally maintained on the interface modules. Each interface module locally handles aging of its local MAC addresses and updates the management module in order to keep the MAC table consistent across the entire system.
- **Foundry Direct Routing (FDR)** technology stores the entire forwarding table in each interface module to allow for hardware forwarding of all traffic.
- **Distributed Access Control List (ACL) maintenance**: Each interface module has hardware support for input ACLs on all ports.

Hardware assists for a robust, scalable infrastructure

The BigIron RX has several innovative hardware assists that enables network designers to create a scalable and stable network. These include:

- **CPU protection:** Layer 2 services require support for efficient replication of packets to the entire broadcast domain. For example, traditional architectures handle Ethernet frames with unknown MAC address by sending them to a processor to replicate the packet to the broadcast domain. The involvement of the CPU makes the system vulnerable to a potential denial of service attack. In contrast, the BigIron RX handles this scenario very efficiently by performing such flooding in hardware.
- **Complete separation of control and data plane**: The isolation of traffic to control plane ensures that control packets to be processed by the system's CPU are efficiently processed even during periods of high data traffic.
- **sFlow**: By performing sFlow collection in hardware, L2-L7 flows can be precisely monitored. This facilitates rapid trouble-shooting and accurate isolation of faults in a network. The BigIron RX supports sFlow v5.

Traffic policers and ACLs

All interface modules support a large number of inbound traffic policers in hardware. Both single-rate 3-color and 2-rate 3-color policers are supported. The single-rate, 3-color marker meters flows that exceed a configured compliant (CIR) rate while the 2-rate, 3-color policers meter subscriber flows by classifying them into compliant (CIR) rates or Peak (PIR) rates. These capabilities are especially useful when mixing traffic flows with different characteristics on the same port.

Input ACLs (Access Control Lists) are supported by the system on all interface modules.

Separation of control and data plane

The BigIron RX has a dedicated out-of-band management link between each interface module and the management module to isolate control traffic from data traffic. Multiple queues to the management module allow different types of control traffic to be prioritized. These capabilities, together with secure management (via SSH, SCP and SNMPv3) and ACLs, are immensely useful in securing the system from potential DoS attacks in the network.

Spatial multicast support

The BigIron RX architecture has native support for spatial multicast, a critical requirement for offering video services in a network. The input interface module sends one copy of an incoming multicast packet to the switch fabric. The switch fabric then replicates the packet within itself to multiple output interface modules in the system, which in turn replicate the multicast packet to the destination ports and VLANs.

100-Gig ready slots

The BigIron RX has a future-proof architecture. Every interface slot has over 48 Gbps of full-duplex switching bandwidth available. The divider between two adjacent interface slots can be removed to convert the half slots into a full slot. The full slot is equipped to handle 100 Gbps of full-duplex bandwidth from the backplane. This makes the BigIron RX chassis capable of migration to 40 Gigabit Ethernet or 100 Gigabit Ethernet interfaces in future.

Industry-leading feature set

Foundry Networks has built on the cumulative experience gained in powering enterprise and service provider networks for over 7 years to create the IronWare software that runs on BigIron RX. The software complements the BigIron RX architecture to offer the following capabilities:

- Support for BGPv4, OSPF, IS-IS and RIP routing protocols in IPv4 networks
- Support for IPv6 including MP-BGP-4, OSPFv3, IS-ISv6 and RIPng routing protocols
- IGMP, MLD, PIM-SM/-DM, and DVMRP support to power multicast applications
- Layer 3 redundancy protocols such as Virtual Router Redundancy Protocol (VRRP), and Virtual Router Redundancy Protocol- Extended (VRRP-E)
- Layer 2 redundancy protocols such as Virtual Switch Redundancy Protocol (VSRP)
- Support for MAC layer service protection protocols such as Metro Ring Protocol (MRP), Rapid Spanning Tree Protocol (RSTP)
- Support for secure management via SSH (v1 and v2), SCP (v1 and v2) or SNMPv3
- Support for 802.1x authentication
- sFlow-based L2-L7 traffic monitoring of activity on the node with underlying hardware support for reliable packet sampling

The BigIron RX architecture allows both Layer 2 and Layer 3 services to be offered on the same device and the same port **concurrently**. This ability gives unprecedented flexibility to the network administrator in tailoring the system to meet end user needs.

Conclusion

The BigIron RX Series is the industry's most powerful IPv4/IPv6 Ethernet switch family. Its robust, scalable architecture coupled with the feature-rich IronWare networking software make it an ideal infrastructure solution for a wide range of customer environments including enterprise backbone, data centers, service provider infrastructure and high performance and cluster computing centers. Featuring the highest density wire-speed Gigabit and 10 Gigabit Ethernet solution in a single system and designed with a redundant switch fabric supporting 100 Gbps capacity per full slot, the BigIron RX delivers the scalability, reliability and performance necessary to meet the needs of the most demanding networking environments.

Document version 2.6

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