Describing the Force10 E-Series Architecture: Technology for Reliability and Scalability for High-Performance Ethernet

Overview

The Force10 Networks E-Series Architecture delivers breakthrough scalability and capacity – the result of technological advances and innovations in the design of switch fabrics, ASICs, a unique passive backplane, and revolutionary system control plane. Working together, the E-Series architecture and Force10 Operating System (FTOS™) in combination delivers the full potential and performance of 10 Gigabit Ethernet at the simplicity and low cost expected for Ethernet economics. This equates to value, unmatched by any other vendor today, to support 100% line-rate performance for massive densities of non-blocking Gigabit Ethernet and 10 Gigabit Ethernet ports—on a innovative high availability platform that offers best-in-class resiliency including zero packet loss hitless failover of all components, memory protection and modular software based on the 3-CPU Route Processor Module (RPM) design.

The E-Series architecture increases switching bandwidth to 1.68 terabits per second (Tbps) on the E-Series E1200 with throughput performance of 1,000 million packets per second (Mpps). The E600 provides a more cost-effective solution with switching capacity of 900 Gigabits per second (Gbps) and 500 Mpps throughput. The E300 is the first line rate 10 Gigabit Ethernet compact chassis-based switch/router scaling to 400 Gbps of switching capacity and supporting up to 196 Mpps throughput. The Force10 Networks Operating System, FTOS, supports a common set of features across all three E-Series products.

Force10 sets the new standard for next-generation switch/routers with unmatched scalability, line-rate forwarding with access control lists (ACLs) enabled on all ports, full L2 switching, and comprehensive L3 routing. This paper details the elements of the E-Series architecture—the building blocks for purpose-built products that provide the foundation for highly scalable network applications.

The Force10 Networks E-Series Architecture

An Ethernet Optimized Architecture

The efficiency and robustness of the E-Series comes from:

- 100% passive copper 5 Tbps backplane design
- 56.25 Gbps of per-slot switching capacity
- 1.68 Tbps switch fabric modules
- Force10 ASIC-based value-add features
- Total redundancy for critical system components
- Clean separation of data and control planes
- Multiprocessor architecture for a resilient software control plane

Features	E1200	E600	E300
Backplane Capacity	5 Tbps	2.7 Tbps	1.2 Tbps
Switch Fabric Capacity	1.68 Tbps	900 Gbps	400 Gbps
Full-Mesh Forwarding Performance	1,000 Mpps	500 Mpps	196 Mpps
Line-rate Gigabit Ethernet	672	336	132
Total Gigabit Ethernet	1,260	630	288
Line-rate 10 Gigabit Ethernet	56	28	12
Total 10 Gigabit Ethernet	224	112	48
Line-rate POS/SDH (OC – 48/12c/3c)	56	28	NA
Hardware Redundancy	Power, Route Processor, Switch Fabric		
Software Redundancy	L2/L3 Hitless Failover		
Operating System	Fully Modular Utilizing a 3-CPU Architecture		

The powerful capacity of the E-Series architecture enables the E-Series platform to deliver new levels of performance across high-density Gigabit Ethernet and 10 Gigabit Ethernet ports. The E-Series architecture maintains line-rate performance when running value-add features including: standard and extended Access Control Lists (ACLs), Quality of Service with Weighted Fair Queuing (WFQ) or Strict Priority Queuing (SPQ), bandwidth control via rate limiting and Weighted Random Early Discard (WRED), and other services simultaneously. Scaling system design to support order of magnitude performance and bandwidth increases like this is not as simple as taking an existing architecture and making it run 10 times faster. To deliver this level of scale and capacity while simultaneously driving costs down requires quantum leaps in system level design.

The switches built over the last five years converged on a similar architecture to support throughput requirements of the then "state of the art" bandwidth requirements – up to 2.5 Gbps. These switch architectures today face multiple issues in scaling to 10 Gigabit speeds and beyond.

- Backplane limitations constrain original Gigabit Ethernet-based designs to 8 Gbps
- Higher system port densities result in the sacrifice of the initially desired system line-rate and non-blocking performance characteristics
- Degrading system performance caused by single CPU-based control systems, which spike to near 100% utilization as value-add features (such as ACLs, QoS, bandwidth control, rate limiting, and others) are enabled
- Distributed switching architectures that suffer from cache exhaustion caused by interface module board-level designs are unable to scale beyond a few thousand entries

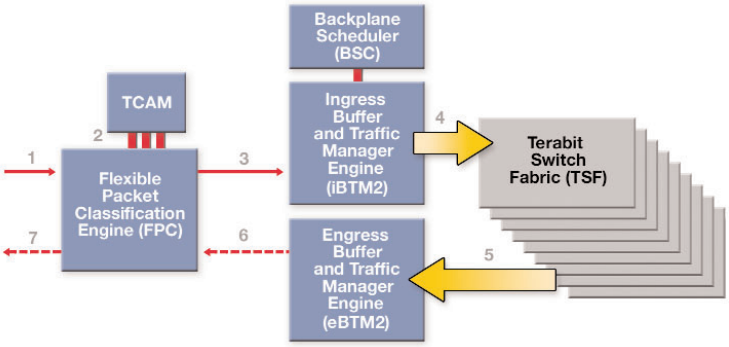


Figure 1. Interface Module Force10 ASICs

Distributed, ASIC-based Architecture

Meeting the challenges discussed above while maintaining Ethernet economics requires significant architectural advances in many areas including back-plane design, interface module construction, data plane architecture, and control plane processing. To date, Force10 Networks has filed over 47 unique patents on the architecture used in the E-Series product platforms.

Force10 ASIC Forwarding Plane

The E-Series architecture delivers line-rate forwarding with ACLs, QoS, and other advanced feature services enabled. Traffic flows through the system at speeds of more than 10 Gbps over a hardware-only data path. The Force10 ASICs working together in unison handling all packet-forwarding tasks include:

- Flexible Packet Classification Engine (FPC)
- Buffer and Traffic Manager (BTM2)
- Terabit Switch Fabric (TSF)
- Backplane Scheduler (BSC)

Packet Level Walkthrough

The Force10 ASICs have reduced the process of forwarding traffic to a minimum set of operations without sacrificing functionality. The following packet level walkthrough refers to Figure 1. Please note the walkthrough uses the functional diagrams for illustrative purposes only.

Packets enter the system (1) and are processed by the Flexible Packet Classification Engine (FPC). The FPC uses a condensed version of the header information to perform simultaneous line-rate Ternary CAM (TCAM) lookups (2) examining details that pertain to Layer 2 (MAC), Layer 3 (IP), Layer 4 (protocol), QoS and ACL related information. At this point, the FPC identifies and separates routing, control messages from data packets, thereby ensuring heavily congested user traffic cannot slow configuration or forwarding table updates to the line cards.

Once the table lookups are complete, the FPC modifies the packet header (if necessary to remark QoS, etc.) and appends the packet's egress port destination. Once packet processing is complete the packet is passed (3) to the ingress Buffer and Traffic Manager (iBTM2).

The iBTM2 enforces the policies defined by the network administrator and set during packet classification including packet rate policing, two-rate three color queuing with dual token buckets, and queuing for transmission through the TSF. The Backplane Scheduler grants access to the iBTM2 to send the packet to the destination line card through the Terabit Switch Fabric (TSF) ASICs (4).

On the egress side, the eBTM2 receives the packet from the Terabit Switching Fabric (TSF), inspects (5) and enforces egress policy (for example rate limiting or egress ACLs), performs WRED for congestion avoidance and passes the packet to the FPC (6). The FPC inspects the packet, removes internal header information, and directs the packet out the appropriate port (7).

Hardware Based Table Lookups

The FPC uses hardware lookup engines, also known as ternary content addressable memories (TCAMs) to retrieve packet-handling instructions. Based upon source or destination MAC, IP, or other combinations of addresses, Differentiated Services Control Point (DSCP) setting, type of service, protocol and/or port number fields, the FPC can retrieve all the information on how to deal with the packet in the time it takes to receive just the packet header. The E-Series architecture easily handles tens to hundreds of thousands of TCAM entries.

ASIC-Based Services

Unlike traditional architectures, Force10's E-Series architecture performs every service at line rate. Traditional architectures rely on sequentially processing instructions, the more services simultaneously applied to a packet the more instructions are required. In these older systems, adding instructions degrades overall forwarding performance and increases latency leading to unpredictable system behavior. At 10 Gbps line speeds, the magnitude of the effect of turning on each feature is dramatically increased.

Access Control Lists (ACLs) with thousands of entries, WRED for congestion avoidance, token-bucket packet classification for traffic prioritization, WFQ and SPQ for policy enforcement, and others often run simultaneously in large networks. With the E-Series architecture, Force10 provides the ability to run all services simultaneously at line rate without performance degradation. Implementing all services directly in the hardware in the forwarding path, ensures that for any combination of services, or even when all services are enabled at the same time, users will not experience any variation in system performance, latency, or jitter.

Scalable Switch Fabric

The E-Series architecture utilizes an innovative 32 x 32 virtual output queue (VoQ) crossbar switch fabric. This fabric combines multiple intelligent queues on every line card with global intelligent queue monitoring that keeps track of the priority and length of every ingress queue in the system. All queues receive efficient servicing across a simple and economical single-level fabric by sophisticated fabric scheduling algorithms. The result is a high-performance fabric that minimizes cost, fabric complexity, latency, and jitter.

High-Performance Switching Capacity

The E-Series "Total Switching Capacity" is calculated from the aggregate switching capacity of nine load-balanced and active TSF modules working in unison. Every TSF, working in conjunction with the E-Series passive copper backplane, incorporates a simple yet powerful design that contains a single Force10 ASIC – they contain no moving parts thereby granting them excellent reliability. Built with future scalability in mind, each TSF module contributes up to 187.5 Gbps of capacity, therefore, the E-Series currently supports 1.68 Tbps of Total System Switching Capacity – see Figure 2.

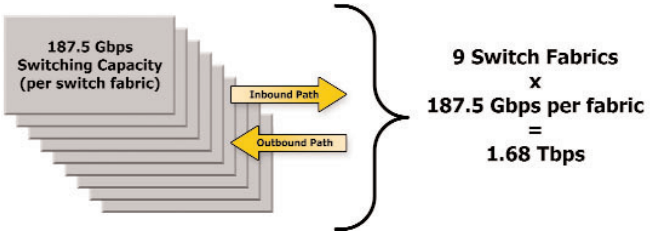


Figure 2. Switch Fabric Capacity

Per Switch Fabric Capacity

Traffic must move in and out of each TSF module through the backplane. The passive backplane and TSF clocking work together in combination to deliver total bandwidth (explained later) so the 187.5 Gbps of per switch fabric module switching capacity is divided equally between 93.75 Gbps of incoming and 93.75 Gbps of outgoing traffic – see Figure 3.

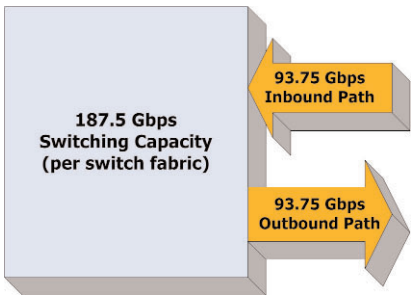


Figure 3. **Individual Switch Fabric**

Port Pipes

The concept of a "Port Pipe" is unique to Force10 Networks and the E-Series architecture. A Port Pipe acts as a backplane communication channel that provides connectivity between slots and the terabit switch fabric modules. The E-Series architecture currently supports 32 Port Pipes. The relationship between Port Pipes and Switching Capacity follows later.

There are two Port Pipes associated to each interface slot position and one Port Pipe assigned to each of the Route Processing Module (RPM) slots. While the E-Series architecture contains 32 Port Pipes, the E1200 and E600 use 30 and 16 Port Pipes, respectively.

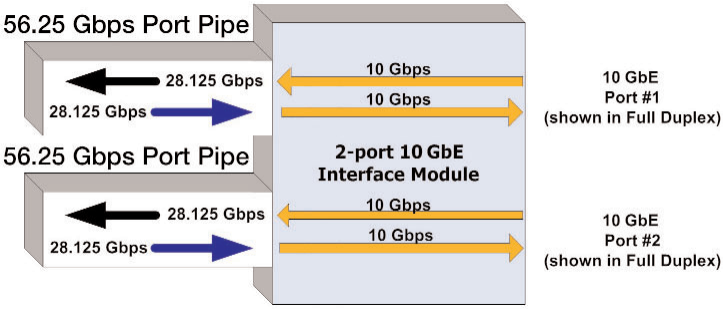


Figure 4. **Individual Switch Fabric**

Port Pipe Capacity

Dividing the 1.68 Tbps total switching capacity by the 30 Port Pipes defined for a system yields a throughput switching capacity of 56.25 Gbps in Full Duplex or 28.125 per individual Port Pipe.

Force10 designed the Port Pipe such that its individual total capacity divides equally between incoming and outgoing traffic – therefore each 28.125 Gbps Port Pipe supports 28.125 Gbps of incoming and 28.125 Gbps of outgoing traffic throughput to match the design of the TSF as illustrated in Figure 4.

Force10 Passive Copper Traces (FPCT)

Force10 Passive Copper Traces (FPCT) on the E-Series passive backplane physically connects interface module slots to the switching fabric modules. Using one single Port Pipe, this example explains how it relates back to the switch fabric modules.

A Port Pipe contains eighteen (18) FPCTs, with nine (9) FPCTs respectively supporting each 28.125 Gbps Inbound and 28.125 Gbps Outbound path. To attain 28.125 Gbps, each FPCT works in combination with the SerDes chipset built into the TSF and operates at a throughput rate of 3.125 Gbps (9 x 3.125 Gbps = 28.125 Gbps).

Within an individual Port Pipe, inbound FPCTs (displayed as solid blue lines) link to the outbound paths on the eight switching fabric modules while eight Outbound FPCTs (depicted as dashed black lines) link to the inbound paths of the switching fabric modules, as illustrated in Figure 5.

With nine active load-balancing TSF modules present in any E-Series E600 or E1200 chassis and each Port Pipe containing eighteen FPCTs operating at 3.125 Gbps, the resulting bandwidth supplied by each Port Pipe is 28.125 Gbps in each direction (56.25 Gbps total capacity). Every slot contains two Port Pipes, (with the exception of the RPM slots, which contain one each); therefore each slot provides 56.25 Gbps capacity and a completely populated E1200 chassis currently supports 1.68 Tbps of usable switching capacity [(56.25 Gbps x 14 slots) + (28.125 Gbps x 2 slots)]. Similarly, a fully populated E600 chassis supports 900 Gbps of usable switching capacity (56.25 Gbps x 7 slots) + (28.125 Gbps x 2 slots). The E-Series E300 similarly is scaled down to support 400 Gbps of usable switching capacity.

Expansive E-Series Architecture

The net effect of the E-Series architecture including FPCT grouped into Port Pipes working in conjunction with the switching fabrics is the ability to support very high-sustaining and high-performance throughput for all packet sizes with 0% loss.

High Capacity Backplane Design

Non-Blocking Fabric Connectivity

The Force10 Terabit Switch Fabric ASICs extend the virtual output queue (VoQ) model to deliver non-blocking performance at greater than terabit-per-second rates. The VoQ design overcomes the cost and complexity of multi-level or clustered fabrics by combining multiple intelligent queues on each ingress line card that are efficiently serviced across a single level fabric by sophisticated fabric scheduling algorithms. The result is a high performance fabric that eliminates head-of-line blocking while minimizing cost, fabric complexity, latency and jitter.

Low-Cost, Reliable 100% Passive Copper Backplane

The E-Series architecture's reliable and cost-efficient backplane is the industry's first high-speed, non-optical backplane to achieve 5 Tbps capacity in a single switch/router chassis. Unlike optical backplane interconnect systems or active copper backplanes, the E-Series backplane has no single points of failure and eliminates costly electrical-optical-electrical conversions. The resulting system simplicity afforded by the E-Series backplane means bulletproof reliability and minimum cost.

Force10 FTOS Software

The Force10 E-Series software and control plane architecture delivers the scalability, resiliency, and security needed to build high-performance networks. The Force10 Operating System (FTOS) consists of modular processes that together deliver carrier class features, robustness, and scalability. FTOS delivers full Layer 2 switching and Layer 3 routing functionality in the secure and protected environment of the E-Series architecture's control plane.

As with the ever-increasing amount of Internet traffic, control plane scalability is mandatory as Internet route tables are ever increasing. At the same time, control plane security is paramount as Denial of Service attacks (DoS) become more frequent, more sophisticated and are now directed at the Internet routing infrastructure. The E-Series architecture's control plane delivers both scalability and security through the multiprocessor based Route Processor Module, control packet filtering and rate-limiting mechanisms, and independent high-speed switched control paths to each system line card.

For the control plane, optionally redundant Route Processor Modules (RPM) handle all route and control processing while providing resiliency to Denial of Service attacks. The RPMs make use of an innovative multi-processor architecture with hardware-based ingress traffic rate limiting and filtering. Distributing software processes among the CPUs maximizes performance while minimizing the possibility that a fault in an individual process could take down the system.

Three independent CPU subsystems on the Force10 E-Series Route Processor Modules run independent software images. One CPU handles the local control and management functions, the second handles Layer 2 processes, and the third handles Layer 3 processes. This modularity enables the FTOS to isolate its processes, minimizing the possibility that a fault in any one process could lead to system failure. The operating system's modularity also enables it to share processes across CPUs, raising the computational power available to individual processes.

Fine Tuned Packet Classification and Control Mechanisms

The ability of the architecture to process, identify, and separate data from control packets, for example route updates or ICMP packets, is a fundamental function required by the network administrator in order to manage and block potentially harmful network traffic from propagating or flooding throughout the network. With the threat of security attacks coming from a large number of sources, the ability to apply potentially

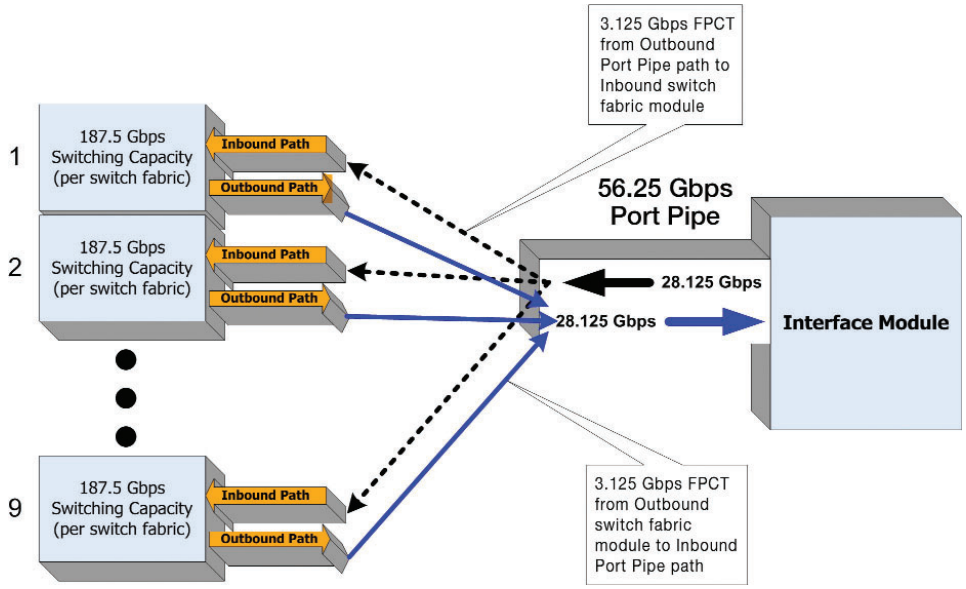
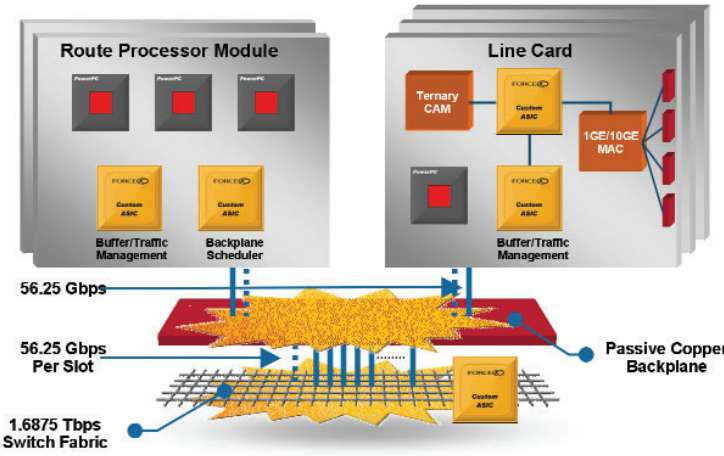


Figure 5. Port Pipe and FPCT Detail



- **3-CPU Route Processor Module — Balanced Functionality**
- **Line Rate Performance with ACL/QoS**
- **Full L2/L3 Functionality**
- **Ultimate Reliability in Hardware Design**
- **Superior Availability in Software Design**

Figure 6. Route Processor Module Multiprocessor Design

hundreds of access list criteria to each incoming control packet is essential — see figure 6. For this security, the control plane of the E-Series architecture delivers extensive ACLs that can be applied to incoming control packets, see Figure 7. With the unique line-rate ACL capability of the E-Series architecture, filtering of control packets introduces no additional latency. This minimization of latency is a crucial factor in reducing overall route table convergence time across the network.

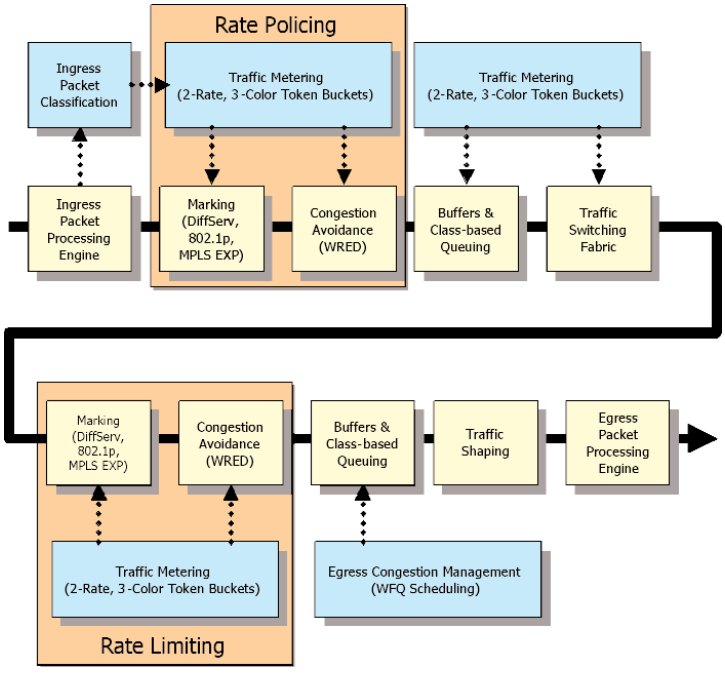


Figure 7. Packet Classification Flow Chart

Control Packet Rate Limiting

Even the most robust switch/router software and hardware implementation is vulnerable to Denial of Service attacks. DoS attacks are malicious acts designed to bring a system or network to its knees by flooding it with useless traffic disguised as specific types of control packets directed at the target control plane CPU. Distributed DoS attacks amplify the amount of useless traffic, in some reported cases to many gigabytes per second, by involving hundreds of sources. With 10 Gbps links, the volume of DoS attack traffic can enter through a single port in the system and overwhelm even the highest performing CPU.

To close this vulnerability, the E-Series architecture employs a programmable hardware mechanism on the RPM to rate limit traffic to the control plane processors. In conjunction with ACLs applied to control packets, the rate limiting mechanism can be configured to rate limit only specific traffic types, for example ICMP, while allowing all other traffic to pass. Control traffic rate limiting uses the same two-rate, three-color marking scheme and token bucket mechanisms found on the E-Series line cards.

Force10 Networks, Inc.
350 Holger Way
San Jose, CA 95134 USA
www.force10networks.com

408-571-3500 PHONE
408-571-3550 FACSIMILE

WP04 906 v3.1