MODERN SERVICE PROVIDER ROUTERS

Guy C Fedorkow
Distinguished Engineer, Juniper Networks
April 2013
THE INTERNET EXPLOSION

Exponential growth, no matter how you measure it!
The clearest indication of value delivered to end-users
WHAT’S A ROUTER

From RFC-1812

“An IP router can be distinguished from other sorts of packet switching devices in that a router examines the IP protocol header as part of the switching process. It generally removes the Link Layer header a message was received with, modifies the IP header, and replaces the Link Layer header for retransmission.”

Routers:
- Strip the L2 Header
- Decrement the TTL
- Look up the Next Hop
- Add New L2 header
- Transmit the Packet

And also
Routing topology, Routing policy, Access Control, Filters, Rate Policing, Shaping, Traffic Prioritization, Buffering, Tunnels, MPLS, LAN Emulation, v4/v6 interworking, NAT, Multicast, Subscriber Authentication, etc, etc.
Large Networks show Evolution in Action!
WHAT’S A SERVICE PROVIDER ROUTER

Core Routers
- Highest Bandwidth Capacity – Nx100GigE
- Emphasis on Route Scaling – Millions of Routes
- Modest Forwarding Feature Set

Common Features
- Common Element Redundancy
- Hot-Swap & ISSU
- BGP Control Plane
- Demanding Environmental (e.g. 55C)
- Redundant DC(!!) Power

Edge Routers
- Highest Interface Scaling – Hundreds of GigE and 10GigE interfaces
- Complex Forwarding Feature Set
- Emphasis on Subscriber Scaling
SAMPLE EDGE ROUTER PORTFOLIO

Edge Routers come in small, medium, large and larger...

<table>
<thead>
<tr>
<th></th>
<th>MX 80</th>
<th>MX 240</th>
<th>MX 480</th>
<th>MX 960</th>
<th>MX 2010</th>
<th>MX 2020</th>
</tr>
</thead>
<tbody>
<tr>
<td>10GE Ports*</td>
<td>48</td>
<td>136</td>
<td>256</td>
<td>260</td>
<td>520</td>
<td></td>
</tr>
<tr>
<td>80Gbps</td>
<td>1.4Tbps</td>
<td>2.6Tbps</td>
<td>4.8Tbps</td>
<td>5.3Tbps</td>
<td>8.8Tbps</td>
<td>40Tbps</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>80Tbps</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>34Tbps</td>
</tr>
</tbody>
</table>
KEY ELEMENTS

High-Scale Routers comprise several key elements:

• Control Plane
  - Responsible for managing routing tables, authenticating subscribers, configuring interfaces

• Packet Forwarding Engine(s) (PFE)
  - Responsible for forwarding each packet (i.e. address lookup, queues, access lists, etc)

• Fabric
  - Responsible for moving packets from one line card to another inside the router
Most high-scale routers are Fabric Based

- Multiple Line Cards, each containing PFEs
- A chassis-wide Interconnect Fabric transfers traffic from ingress to egress line cards
Each line card connects to all Fabric cards.

That’s a lot of wires on the backplane!
MX2020 PACKAGING

Craft Panel

Upper 10 I/O Slots

8 Fabric’s 2 RE’s

Lower 10 I/O Slots

CHASSIS FRONT

CHASSIS SIDE

19” W

45 RU H

920mm D
IF ONE CHASSIS IS TOO SMALL… MULTICHASSISIS Routers

What to do when one chassis isn’t big enough?

…So if that’s not enough, go to multiple fabric stages using Benes or Butterfly networks

Single Stage fabric only scales so far…

Issues to Solve:
• Expensive optical cables
• SW Scaling & Distributed Control
• Proto costs(!)

Wrap this mesh around a cylinder so these points meet…
PACKET FORWARDING ENGINE (PFE)

PFE’s do the work to move packets from Ingress to Egress

Key Functions:

- **L2 & L3 Analysis & Features**
  
  Figure out who’s packet it is, what should happen to it, and where it should go.

- **Packet Buffering**
  
  Store the Packet in DRAM until there’s room to transmit it

- **Queuing & Scheduling**
  
  Decide which packets should go in what order to achieve fairness and real-time delivery guarantees.

PFEs may be micro-programmable, table-driven or hard-coded

[It’s the old Cost/Performance/Flexibility Tradeoff Matrix…]
TRIO PFE ARCHITECTURE

10GE 10GE 10GE 10GE

IX IX IX IX
Optional Optional Optional Optional

Pre-classification Pre-classification Pre-classification Pre-classification

LU LU LU LU
Classification Classification Classification Classification
Policing Policing Policing Policing
Re-write Re-write Re-write Re-write

QX QX QX QX
Optional Optional Optional Optional

HQoS HQoS HQoS HQoS

Packet data Packet data Packet data Packet data

Fabric queuing Fabric queuing Fabric queuing Fabric queuing

MQ MQ MQ MQ

Port based queuing Port based queuing Port based queuing Port based queuing

Fabric Fabric Fabric Fabric

RRDRAM RRLDRAM RRLDRAM RRLDRAM
SRAM SRAM SRAM SRAM
Queues Queues Queues Queues

Link RAM Link RAM Link RAM Link RAM

10GE 10GE 10GE 10GE

Copyright © 2010 Juniper Networks, Inc.
SAMPLE TRIO PACKET PATH

BA Classification (LU)
Forwarding Class & Packet Loss Priority

MF Classification (LU)
Policing
Can overwrite FC & PLP

MQ: Fabric Interface
Fabric Queuing

Switch Fabric

MQ: Fabric Interface
No Schedulers

LU: Egress Filters/Policers
Rewrite Headers

QX Scheduling:
HQoS Queuing
FANCY QUEUING

Level 4

Buffer

Level 3

PIR

CIR

VLAN

Queues

Shapers

Level 2

PIR

CIR

INTERFACE-SET

PIR

CIR

INTERFACE-SET

PIR

CIR

INTERFACE-SET

Shapers

Level 1

PIR

CIR

Port

Traffic

Egress Interface

PIR

Shaper

PQ-DWRR Scheduler

WRED Scheduler

Copyright © 2010 Juniper Networks, Inc.
WHAT’S HARD ABOUT HIGH-SCALE ROUTERS?

- Power Management
  - Keeping power dissipation down
  - Getting the heat out

- Signal Integrity
  - Crosstalk from a zillion wires on the backplane

- Features
  - QoS
  - Multicast
  - Tunnels
  - Link Aggregation & ECMP Load Balancing
    -- and -- “Feature Velocity”

- Scaling the Control Plane -- FIB and Subscribers

- Service Delivery (NAT&Firewall, Content Delivery, DPI, & who knows what else)
Questions?

For further information on router functionality:
*Juniper MX Series*, Douglas Richard Hanks Jr and Harry Reynolds, published by O'Reilly