Router Architecture Overview

What’s inside a router?

Philipp S. Tiesel – philipp@inet.tu-berlin.de
Slides credits to James Kempf & Anja Feldmann
What does a Router Look Like?

- **Ericsson SSR 8020 BNG/BRAS/PGW**
  - Maximum 16 Tbit/s

- **Cisco CRS-1 Core Router**
  - 2.2Tbit/s single chassis
  - Up to 322Tbit/s for multichassis

- **Juniper T4000 Core Router**
  - 3.8 Tbit/s single chassis
<table>
<thead>
<tr>
<th>Model</th>
<th>Description</th>
</tr>
</thead>
</table>
| Dlink DIR-615 Wireless N 300 | - LAN: 4x 10/100Mbit/s Ports  
                              | - WAN: 1x 10/100Mbit/s Port  
                              | - Up to 300 Mbit/s throughput  
                              | - WiFi support                |
| Belkin (formerly Cisco, Linksys) N600 DB Wireless Dual-Band N+ Home Router | - LAN: 4x 10/100Mbit/s Ports  
                              | - WAN: 1x 10/100Mbit/s Port  
                              | - Up to 300 Mbit/s throughput  
                              | - WiFi support                |
Who Makes Core Routers?

- Cisco
  - CRS (Carrier Router Series)
- Juniper
  - T-series
- Alcatel-Lucent (soon to be Nokia)
  - XRS (Extensible Routing System)
- Huawei
  - Netengine
- Others manufacture aggregation/access networking gear for edge deployments
Router architecture overview

two key router functions:
- run routing algorithms/protocol (RIP, OSPF, BGP)
- **forwarding** datagrams from incoming to outgoing link

Incoming reachability information via RIP/OSPF/IS-IS/BGP used to compute routing information Base (RIB)

Routing, management control plane (software)

Forwarding data plane (hardware)

Forwarding information base (FIB) computed, pushed to line cards

Incoming datagrams forwarded to outgoing link by high-speed switching fabric

Router line cards on input side

Router line cards on output side
RIB

- Router can contain many different RIBs
  - One for each routing protocol
  - Usually consolidated into one global RIB or into FIB
  - End system IP addresses (/32s) populated through ARP for default gateway MAC

- Minimum contents
  - Network id of destination subnet
  - Cost/metric for hop
  - Next hop gateway or end system

- Other information
  - Quality of service, e.g. a U if the link is up
  - Access control lists for security
  - Interface, such as eth0 for first Ethernet line card, etc.

Network + Netmask = network id (192.168.0.0/24 in this case)

<table>
<thead>
<tr>
<th>Network Destination</th>
<th>Netmask</th>
<th>Gateway</th>
<th>Interface</th>
<th>Metric</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0.0.0</td>
<td>0.0.0.0</td>
<td>192.168.0.1</td>
<td>192.168.0.100</td>
<td>10</td>
</tr>
<tr>
<td>127.0.0.0</td>
<td>255.0.0.0</td>
<td>127.0.0.1</td>
<td>127.0.0.1</td>
<td>1</td>
</tr>
<tr>
<td>192.168.0.0</td>
<td>255.255.255.0</td>
<td>192.168.0.100</td>
<td>192.168.0.100</td>
<td>10</td>
</tr>
<tr>
<td>192.168.0.100</td>
<td>255.255.255.255</td>
<td>127.0.0.1</td>
<td>127.0.0.1</td>
<td>10</td>
</tr>
<tr>
<td>192.168.0.1</td>
<td>255.255.255.255</td>
<td>192.168.0.100</td>
<td>192.168.0.100</td>
<td>10</td>
</tr>
</tbody>
</table>
FIB

- FIB contains optimized next hop forwarding information
  - Exact format depends on the line card hardware (ASIC, CAM, etc.)

- RIB compiled into FIB by the router control processor when routes change
  - Example:
    - If routing ASIC uses btrees, then RIB compiled into btrees
  - Usually contains information in a form needed for getting a packet out fast
    - Example:
      - Replace IP address of outgoing interface by hardware address on dedicated switch fabric

- Installed into the line card by route processor

- A packet that experiences a FIB miss on fast path will incur a substantial performance penalty
  - Must be transferred to route control processor and processed on the slow path
Input port functions

decentralized switching:
- given datagram dest., lookup output port using forwarding table in input port memory ("match plus action")
- goal: complete input port processing at 'line speed'
- queuing: if datagrams arrive faster than forwarding rate into switch fabric
Switching fabrics

- transfer packet from input buffer to appropriate output buffer
- switching rate: rate at which packets can be transfer from inputs to outputs
  - often measured as multiple of input/output line rate
  - N inputs: switching rate N times line rate desirable

- three types of switching fabrics

  memory
  - memory
  - bus
  - dedicated fabric
Switching via memory

*first generation routers:*

- traditional computers with switching under direct control of CPU
- packet copied to system’s memory
- speed limited by memory bandwidth (2 bus crossings per datagram)
Switching via a bus

datagram is switched
→ from input port memory
← to output port memory
via a shared bus

bus contention: switching speed limited by bus bandwidth

e.g. Cisco 5600: 32 Gbps bus sufficient speed for access and enterprise routers
Switching via Dedicated Fabric

- overcome bus bandwidth limitations
- banyan networks, crossbar, other interconnection nets initially developed to connect processors in multiprocessor
- advanced design:
  - fragmenting datagram into fixed length cells.
  - append hardware address of output line card to front of cell
  - switch cells through the fabric.
- Cisco 12000: switches 60 Gbps through the interconnection network
- **buffering** required when datagrams arrive from fabric faster than the transmission rate
- **scheduling discipline** chooses among queued datagrams for transmission
Output port queueing

- buffering when arrival rate via switch exceeds output line speed
- *queueing (delay) and loss due to output port buffer overflow!*
How much buffering?

- RFC 3439 rule of thumb: average buffering equal to “typical” RTT times link capacity C
  - for C = 10 Gpbs link, RTT = 250 msec: 2.5 Gbit buffer

- recent recommendation: with $N$ flows, buffering equal to
  
  \[ \frac{\text{RTT} \cdot C}{\sqrt{N}} \]
Input port queuing

- fabric slower than input ports combined -> queueing may occur at input queues
  - *queueing delay and loss due to input buffer overflow!*
- Head-of-the-Line (HOL) blocking: queued datagram at front of queue prevents others in queue from moving forward

output port contention: only one red datagram can be transferred.
*lower red packet is blocked*

one packet time later: green packet experiences HOL blocking
Summary

- Routers have evolved from dedicated workstations/PCs into heavy duty industrial equipment costing millions of €s
  - Capable of switching 300+ Tb/s
- Dedicated line cards separate fast path from slow path
  - Slow path through route processor primarily for control plane traffic
- Different hardware approaches to accelerating fast path forwarding
- Queuing and buffering can help or hinder traffic flow