Lecture 17: Router Design

CSE 123: Computer Networks
Alex C. Snoeren

HW 3 due WEDNESDAY
Lecture 17 Overview

- BGP relationships
- Router internals
  - Buffering
  - Scheduling
Business Relationships

- Neighboring ASes have business contracts
  - How much traffic to carry
  - Which destinations to reach
  - How much money to pay

- Common business relationships
  - Customer-provider
    - E.g., Princeton is a customer of USLEC
    - E.g., MIT is a customer of Level3
  - Peer-peer
    - E.g., UUNET is a peer of Sprint
    - E.g., Harvard is a peer of Harvard Business School
Customer/Provider

- Customer needs to be reachable from everyone
  - Provider tells all neighbors how to reach the customer
- Customer does not want to provide transit service
  - Customer does not let its providers route through it

Traffic **to** the customer

Traffic **from** the customer
Multi-Homing

- Customers may have more than one provider
  - Extra reliability, survive single ISP failure
  - Financial leverage through competition
  - Better performance by selecting better path
  - Gaming the 95th-percentile billing model
Peer-to-Peer Relationship

- Peers exchange traffic between customers
  - AS exports *only* customer routes to a peer
  - AS exports a peer’s routes *only* to its customers
  - Often the relationship is settlement-free (i.e., no $$$)

Traffic to/from the peer and its customers
Tier-1 Providers

- Make up the “core” of the Internet
  - Has no upstream provider of its own
  - Typically has a national or international backbone
- Top of the Internet hierarchy of ~10-20 ASes
  - E.g., AT&T, Level 3 (Global Crossing), NTT/Verio, Century Link (formerly Qwest), Sprint, Verizon
  - Full peer-peer connections between tier-1 providers
The Internet Hierarchy

- Flatter and much more densely interconnected Internet
- Disintermediation between content and "eyeball" networks
- New commercial models between content, consumer, and transit

Settlement Free

Pay for BW

Pay for access BW

CSE 123 – Lecture 17: Router Design
BGP Summary

- Interdomain-routing
  - Exchange reachability information (plus hints)
  - BGP is based on path vector routing
  - Local policy to decide which path to follow

- Traffic exchange policies are a big issue $$$
  - Complicated by lack of compelling economic model (who creates value?)
  - Can have significant impact on performance
What’s in a Router?

● Physical components
  - One or more input interfaces that receive packets
  - One or more output interfaces that transmit packets
  - A chassis (box + power) to hold it all

● Functions
  - Forward packets
  - Drop packets (congestion, security, QoS)
  - Delay packets (QoS)
  - Transform packets? (Encapsulation, Tunneling)
Router Functions

1. Receive incoming packet from link input interface
2. Lookup packet destination in forwarding table (destination, output port(s))
3. Validate checksum, decrement ttl, update checksum
4. Buffer packet in input queue
5. Send packet to output interface (interfaces?)
6. Buffer packet in output queue
7. Send packet to output interface link
Functional architecture

Control Plane
- Complex
- Per-control action
- May be slow

Data plane
- Simple
- Per-packet
- Must be fast
Interconnect architecture

- Input & output connected via switch fabric

- Kinds of switch fabric
  - Shared Memory
  - Bus
  - Crossbar

- How to deal with transient contention?
  - Input queuing
  - Output queuing
First Generation Routers

- CPU
- Route Table
- Buffer Memory

Shared Bus(es)

Line Card

MAC

Line Card

MAC

Line Card

MAC

Single CPU and shared memory;

All classification by main CPU
Second Generation Routers

- CPU
- Route Table
- Shared Bus(es)
- Direct DMA on cache hit
- Line Card
  - Buffers
  - Forwarding Cache
  - MAC
- Cache of recent routes
Third Generation Routers

- Shared interconnect (frequently crossbar)
- Centralized scheduler
- Full forwarding table in line card
- Fixed cells

CSE 123 – Lecture 17: Router Design
Output queuing

- Output interfaces buffer packets

- Pro
  - Simple algorithms
  - Single congestion point

- Con
  - $N$ inputs may send to the same output
  - Requires *speedup* of $N$
    - Output ports must be $N$ times faster than input ports
Input queuing

- Input interfaces buffer packets

- Pro
  - Single congestion point
  - Simple to design algorithms

- Con
  - Must implement flow control
  - Low utilization due to Head-of-Line (HoL) Blocking
Head-of-Line Blocking
IQ + Virtual Output Queuing

- Input interfaces buffer packets in per-output virtual queues

- Pro
  - Solves blocking problem

- Con
  - More resources per port
  - Complex arbiter at switch
  - Still limited by input/output contention (scheduler)
Virtual Output Queues
Switch scheduling

- Problem
  - Match inputs and outputs
  - Resolve contentions, no packet drops
  - Maximize throughput
  - Do it in constant time…

- If traffic is uniformly distributed it's easy
  - Lots of algorithms (approximate matching)

- Seminal result (Dai et al, 2000)
  - Maximal size matching + speedup of two guarantees
  - 100% utilization for most traffic assumptions
For Next Time

- Read P&D 6.2
- Get started on P2 initial tasks
- Homework 3 due Wednesday