Lecture 17: Router Design

CSE 123: Computer Networks

Alex C. Snoeren

Eample courtesy Mike Freedman

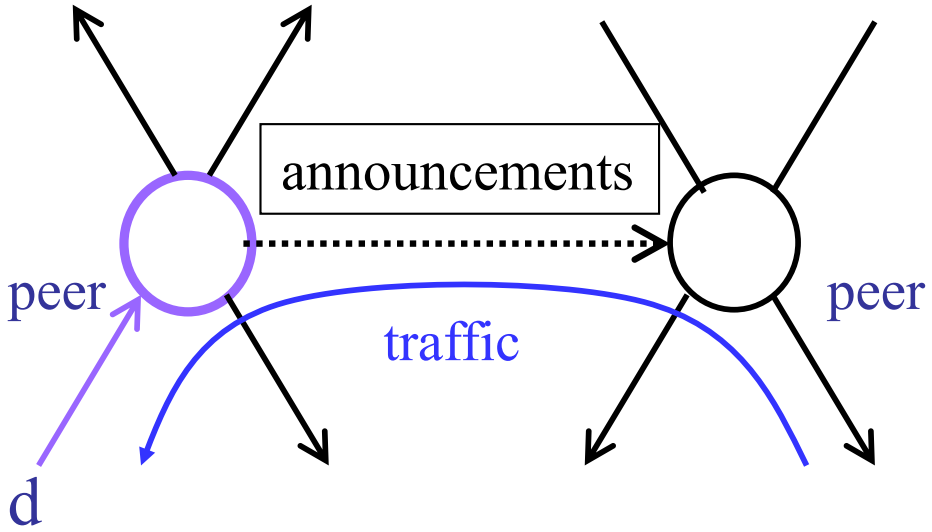
Lecture 17 Overview

- Finish up BGP relationships
- Router internals
  - Buffering
  - Scheduling

Peer-to-Peer Relationship

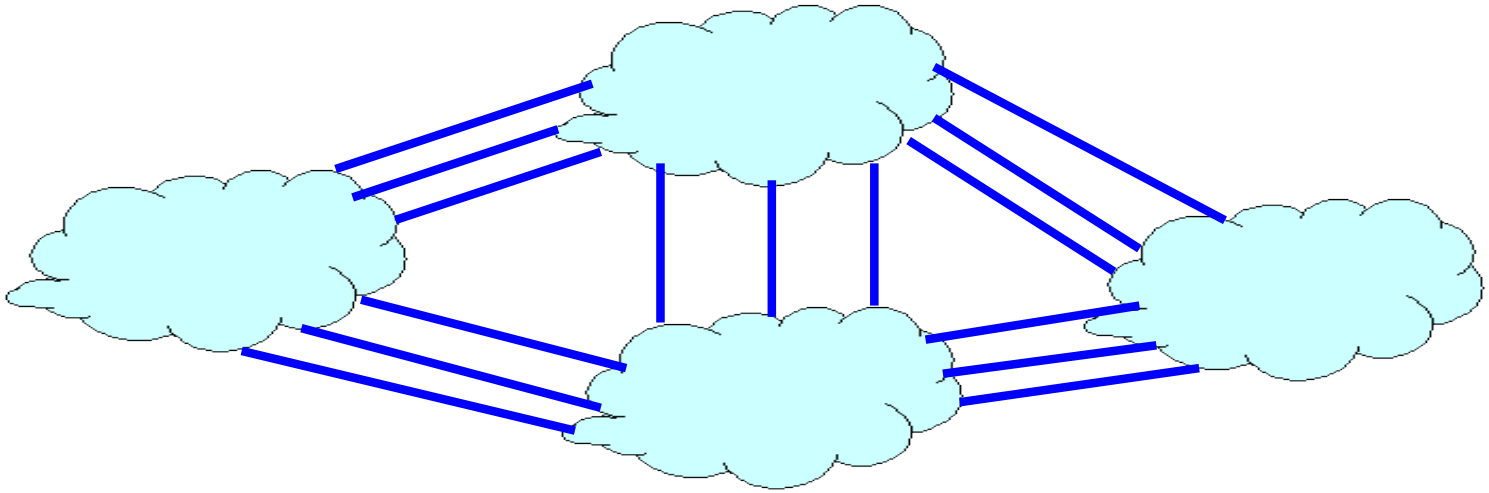
- Peers exchange traffic between customers
  - AS exports *only* customer routes to a peer
  - AS exports a peer's routes *only* to its customers
  - Often the relationship is settlement-free (i.e., no $$$)

Traffic to/from the peer and its customers

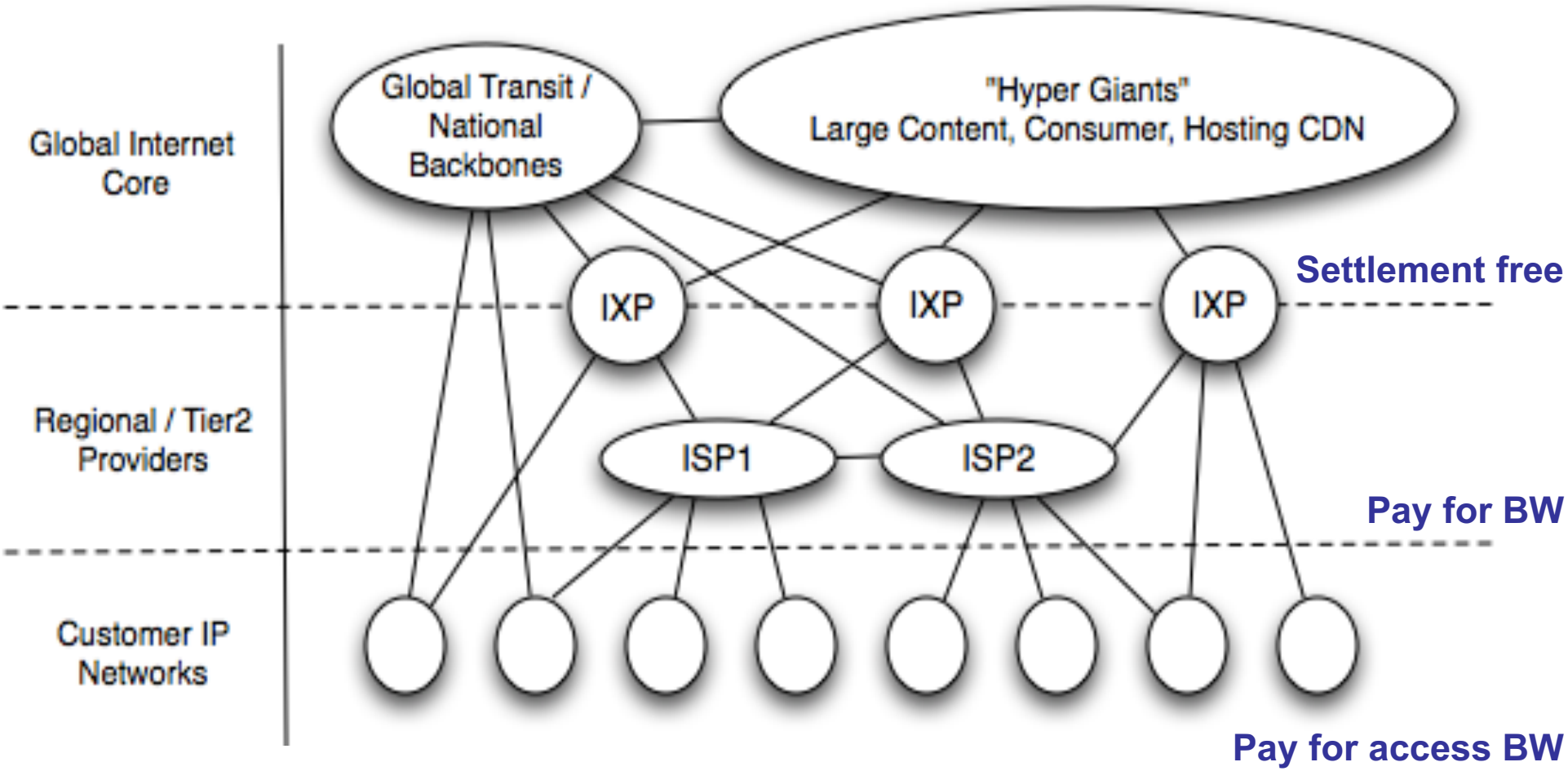


Tier-1 Providers

- 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



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?

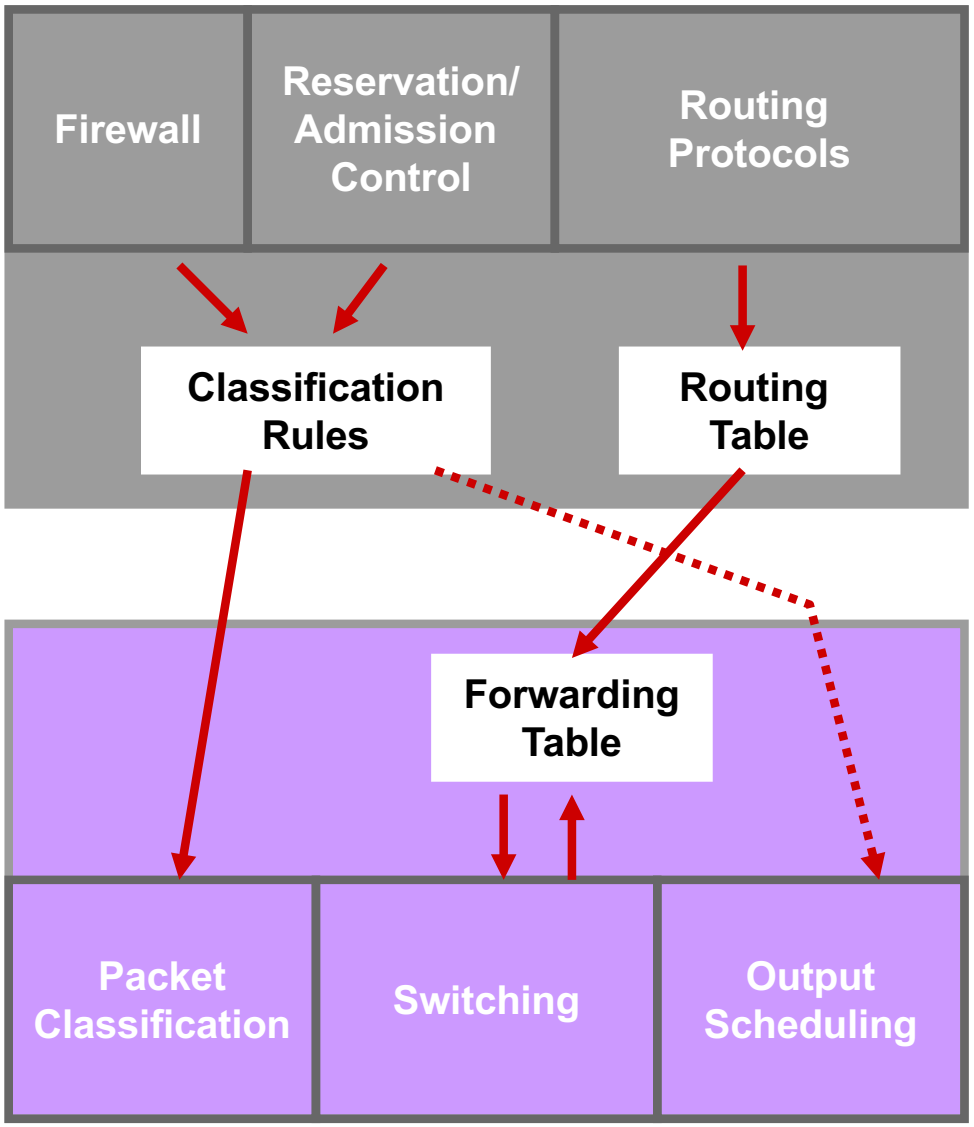
- Physical components
  - One or more input interfaces that receive packets
  - One or more output interfaces that transmit packets
  - A chassis (box + power) to hold it all

- 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

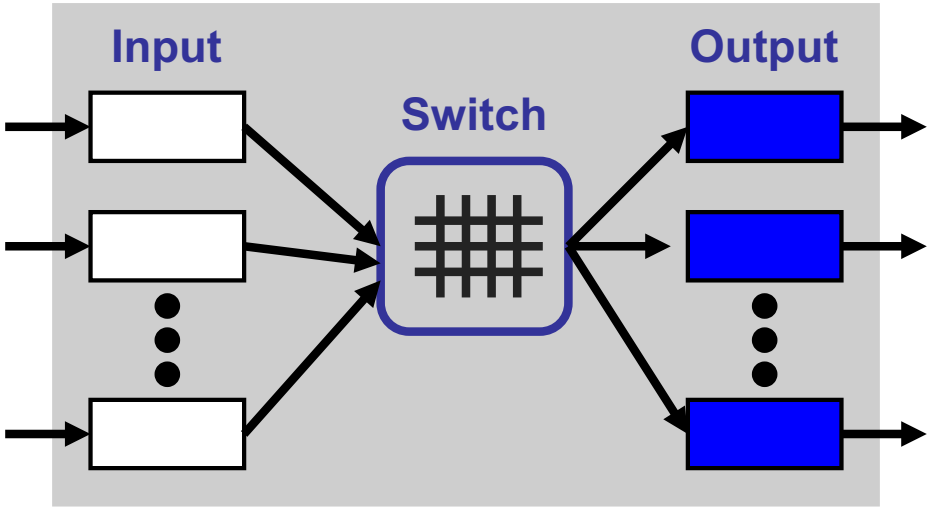


Control Plane
- Complex
- Per-control action
- May be slow

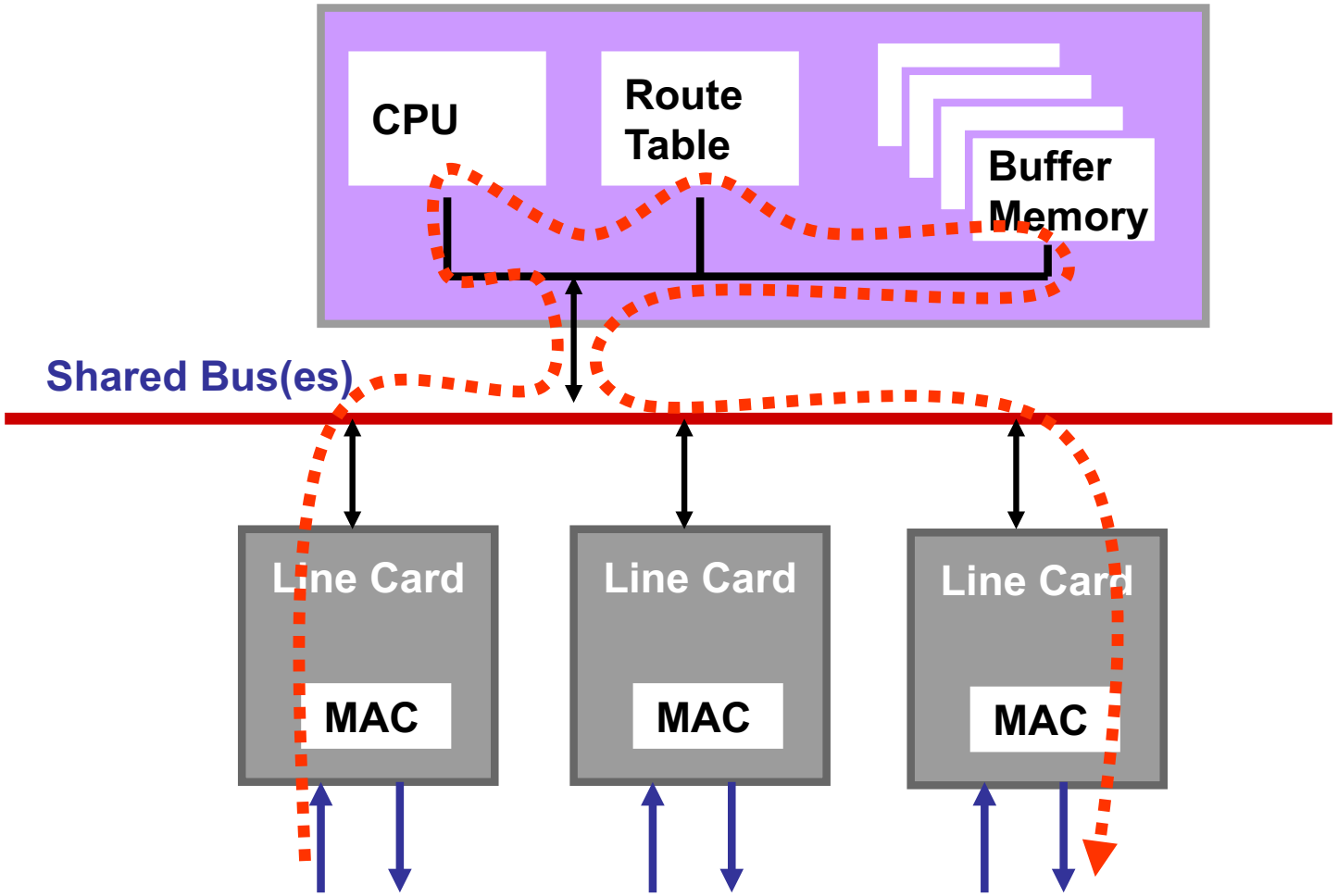
Data plane
- Simple
- Per-packet
- Must be fast

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



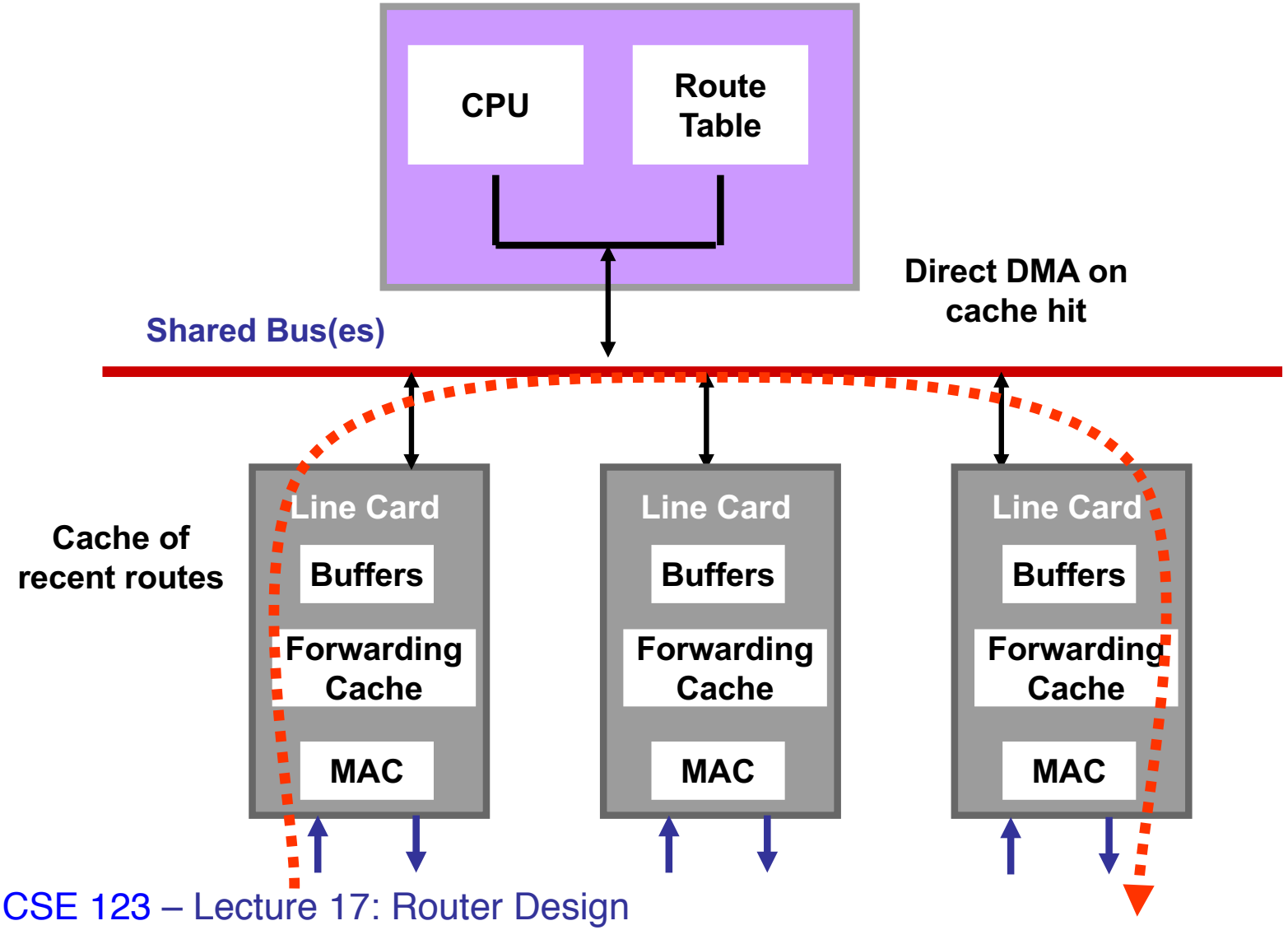
First Generation Routers



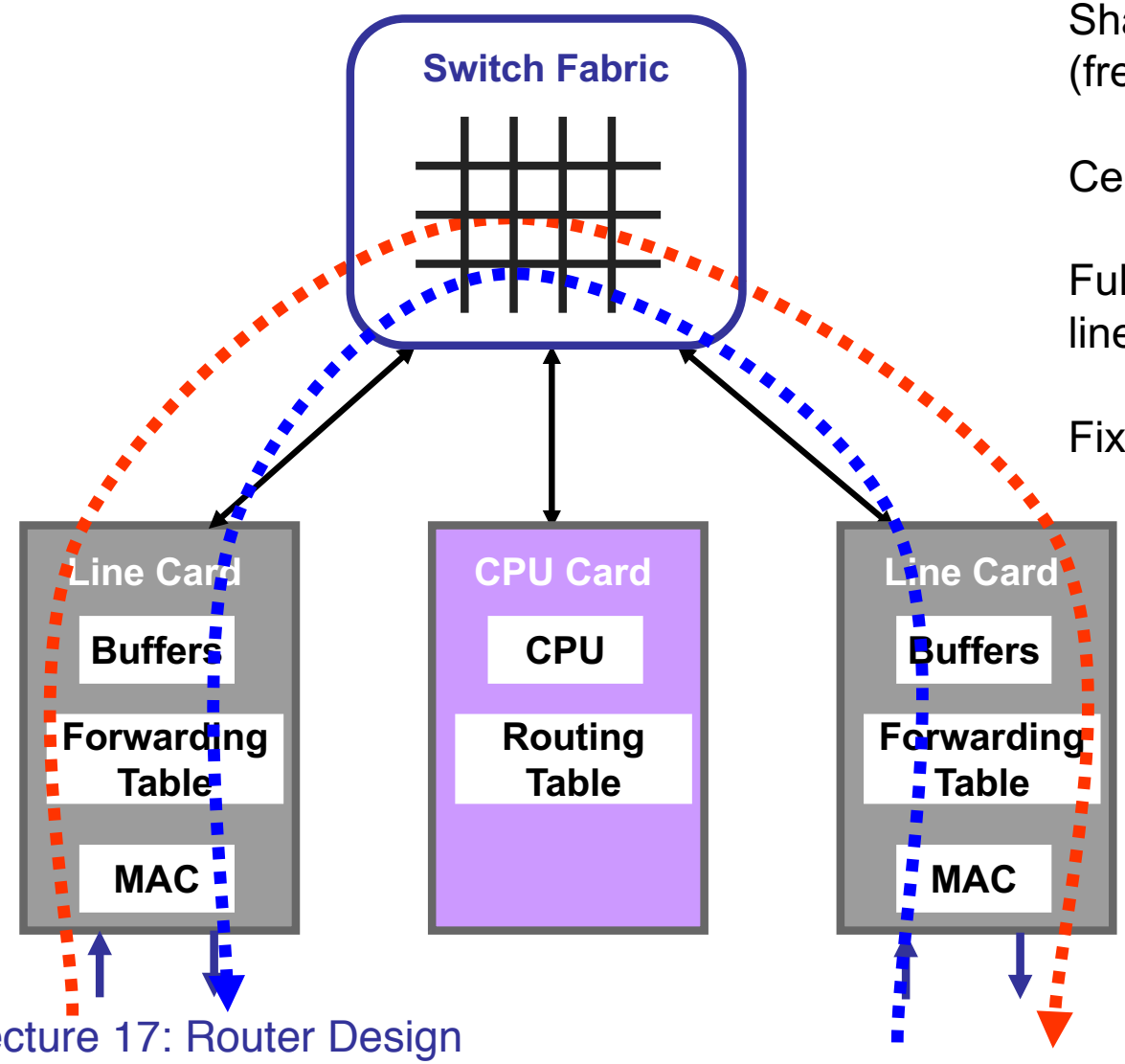
Single CPU and shared memory;

All classification by main CPU

Second Generation Routers



Third Generation Routers



Shared interconnect (frequently crossbar)

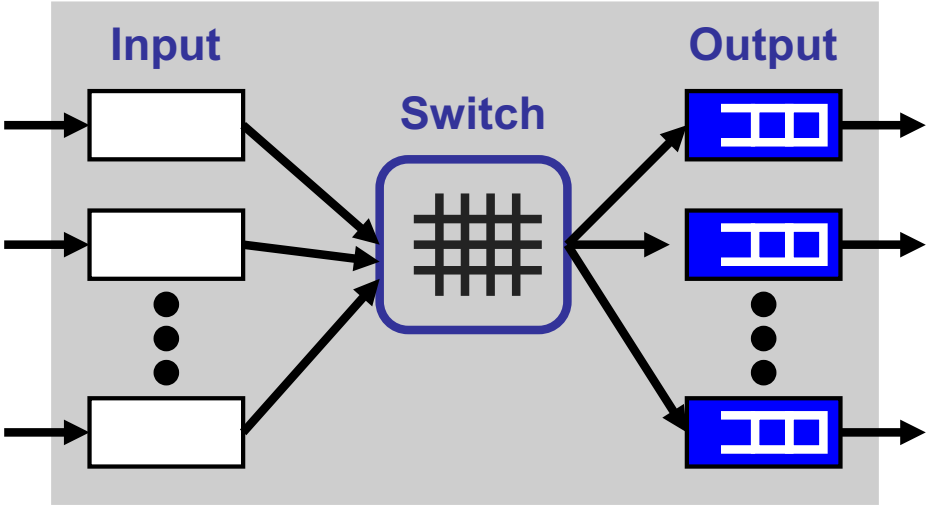
Centralized scheduler

Full forwarding table in line card

Fixed cells

Output queuing

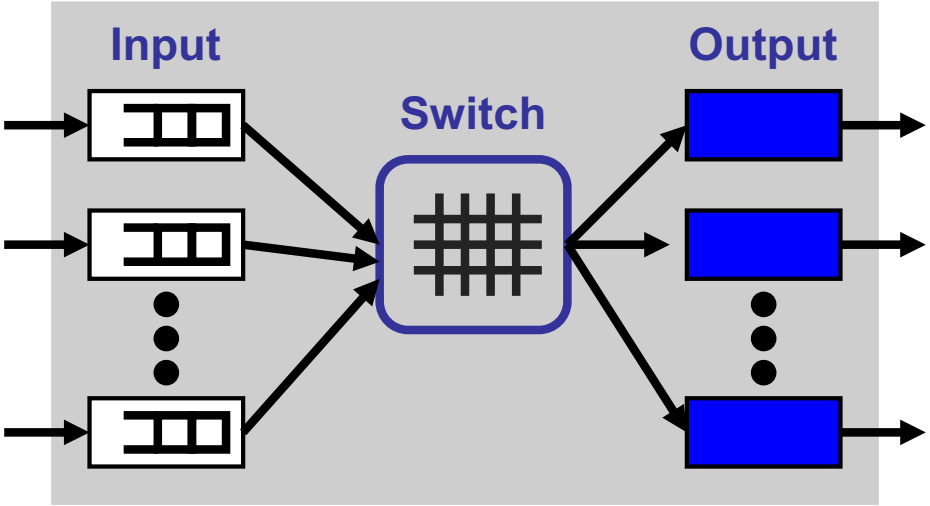
- Output interfaces buffer packets



- 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

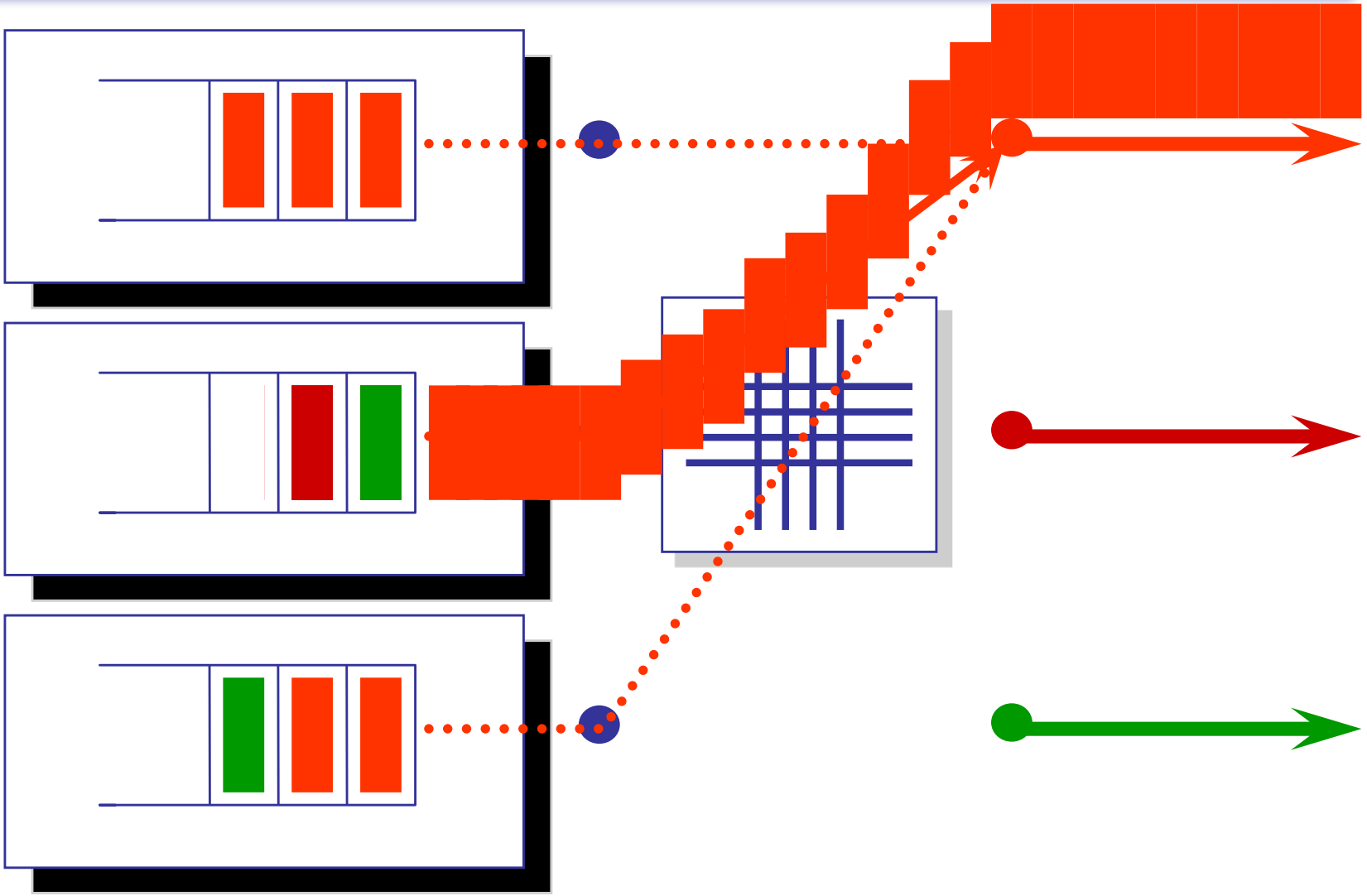
Input queuing

- Input interfaces buffer packets



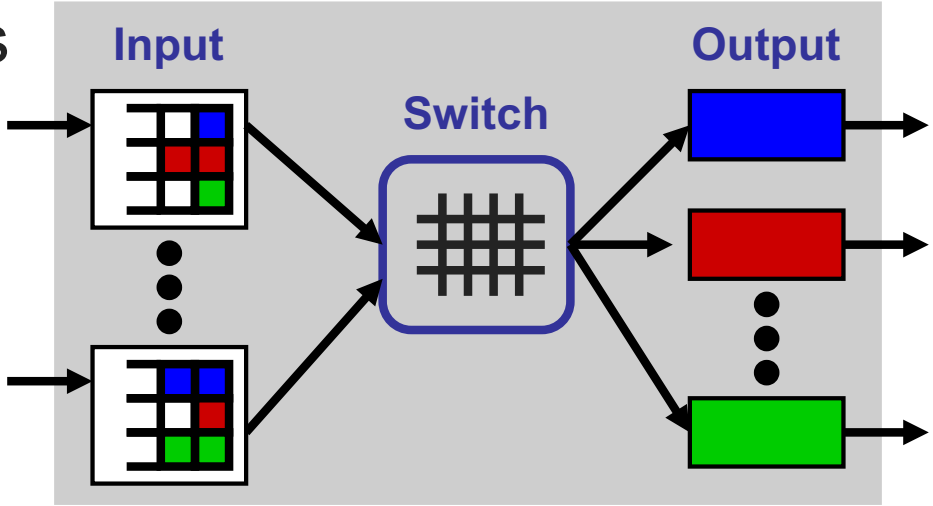
- Pro
  - Single congestion point
  - Simple to design algorithms
- Con
  - Must implement flow control
  - Low utilization due to Head-of-Line (HoL) Blocking

Head-of-Line Blocking



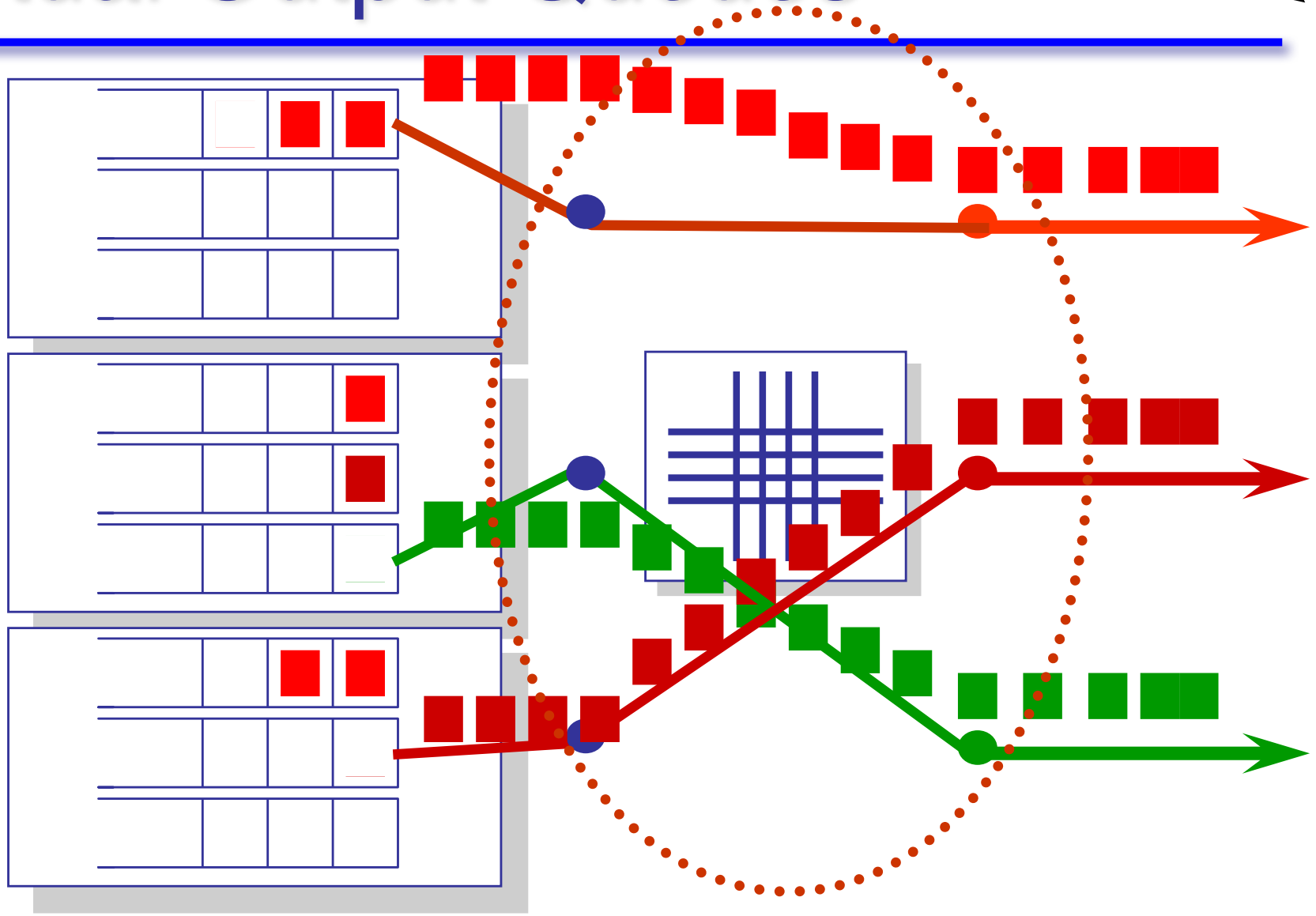
IQ + Virtual Output Queuing

- 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 its 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