ProCurve Networking
HP Innovation

Proposal to improve expedited forwarding

Paul Congdon
IEEE 802.1 Interim - Berlin
May 12, 2005

hp invent

Why consider this…

ProCurve Networking
HP Innovation

Number of requests on 802.1 to improve the way a bridge forwards differentiated traffic

- Congestion management
- Residential Ethernet
- Industrial Ethernet
- Provider Bridging

Most bridges shipping today do more than what 802.1D defines (expedited forwarding), yet there is no standard way to manage them

Why NOT consider this…

ProCurve Networking
HP Innovation

This is a can of worms…

Many different methods with differing implementation implications

Fully addressing Quality of Service concepts requires more than just packet scheduling in bridges…classification, metering, marking, shaping, dropping, etc…

What we have today

ProCurve Networking
HP Innovation

802.1D-2004 specifies

- Between 1 and 8 Traffic Classes
- Egress queues (7.7.3)
- Strict Priority Queuing (PQ) (7.7.4)

802.1Q also specifies

- Metering of flows identified by:
  - Destination MAC Address
  - VLAN
  - Priority
- Recommends Metro Ethernet Forum algorithm for metering (MEF5)
- No specification to configure the meters
- Meters can drop frames or mark drop precedence (802.1ad)
- Ability to drop frames from the queue to manage depth, but no specific queue management algorithm (8.6.7)
- Frames with drop precedence set shall have a higher priority of being dropped by the queue management algorithm (802.1ad)

802.1D Forwarding Process

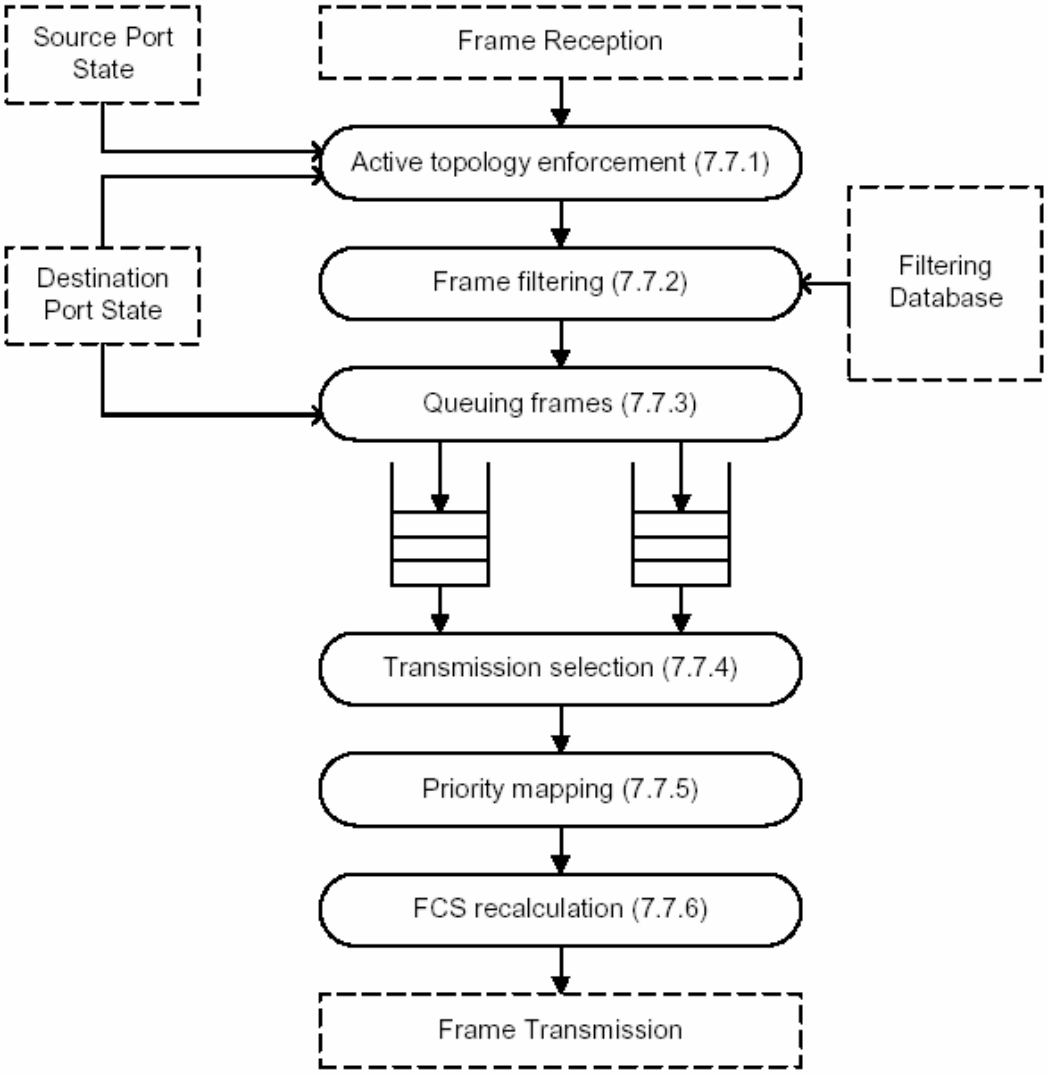


Figure 7-8—Operation of the Forwarding Process

802.1Q Forwarding Process

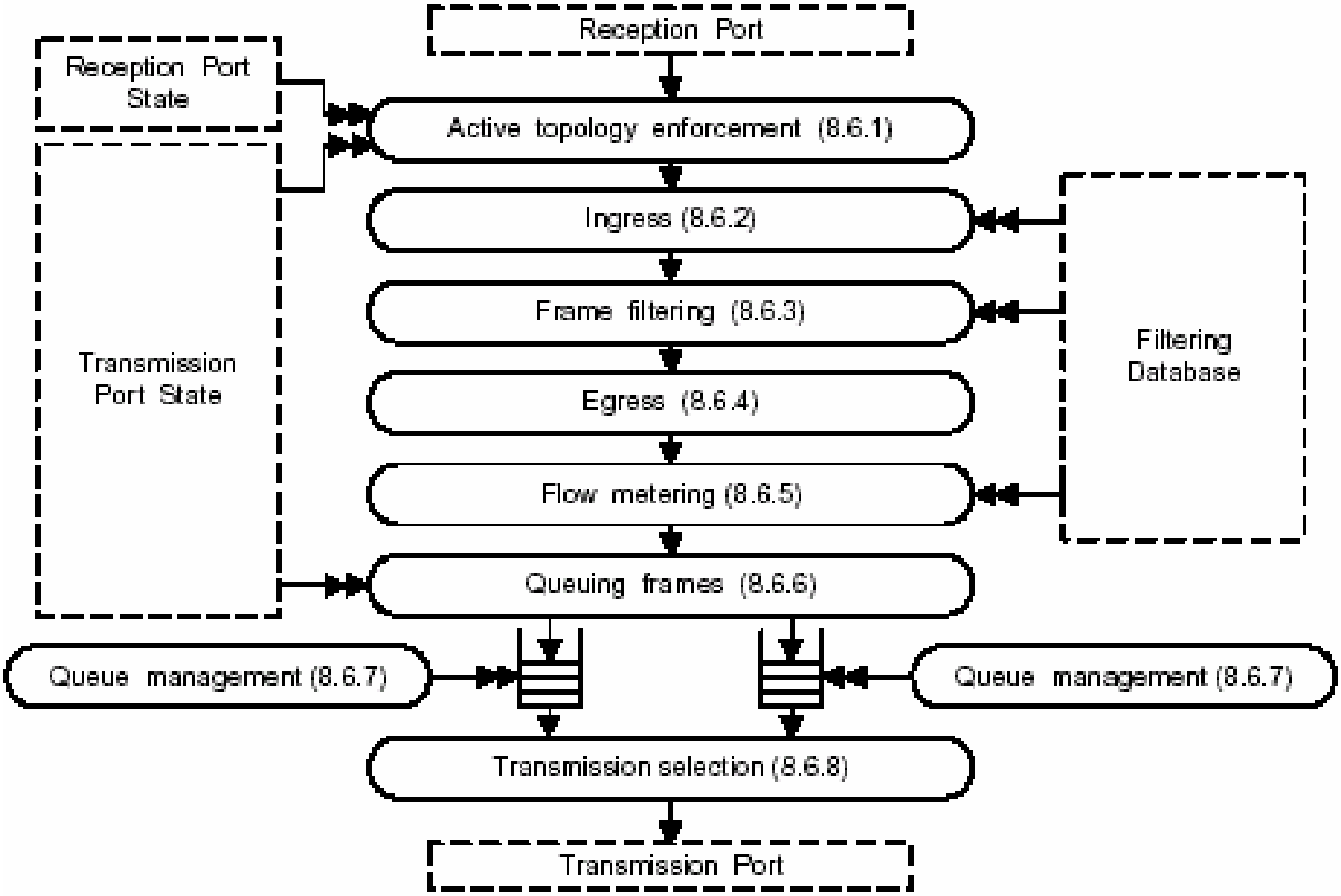
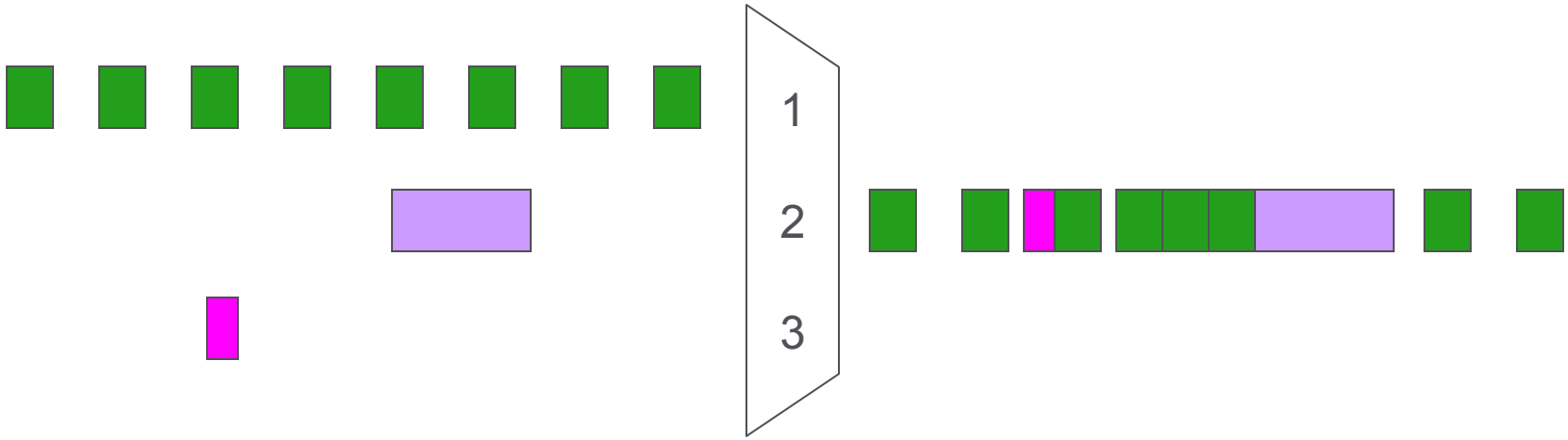


Figure 8-9—Forwarding Process functions

Strict Priority Queuing



From 8.6.8

frames are selected from the corresponding queue for transmission only if all queues corresponding to numerically higher values of traffic class supported by the Port are empty at the time of selection.

Pros and Cons to Strict PQ

ProCurve Networking
HP Innovation

Pros

- Simple to implement
- Provides differentiated treatment for well behaved traffic streams
- Delay bounds may be approximated with admission control and well behaved applications

Cons

- Without policing or admission control of higher priorities, lower priorities have no guarantees
- No delay bounds are provided

Using 802.1p with Reservations

ProCurve Networking
HP Innovation

RFC 2814 and 2815 define Integrated Service Mapping on IEEE 802 networks

Performs admission control by mapping application reservation requests to 802 priorities

Can’t support Guaranteed Service because of in ability to enforce delay bounds and network control traffic mapping

Can get very close approximation of Controlled Load and Guaranteed Service with well behaved applications.

Bandwidth Guarantees without Admission Control

ProCurve Networking
HP Innovation

To avoid starvation, use an algorithm that assigns classes a percentage of output port bandwidth (e.g. RC-PQ, DWRR)

RC-PQ = Rate Controlled Priority Queuing

- Frames in a high priority queue are scheduled before frames in a lower priority queue only if the amount of traffic in the high-priority queue remains below a user defined threshold

DWRR = Deficit Weighted Round Robin

- Frames from a non-empty queue are scheduled based upon the accrual of credits that are proportional to the queues configured bandwidth percentage. Credits are decremented when frames are transmitted. Credits are incremented when frames are queued.

Current 802.1 Metering Observations

ProCurve Networking
HP Innovation

Metering with discard is a way to perform ingress rate limiting for:

- DA MAC
- VLAN
- Traffic Class

No specification of how to quantify a meter

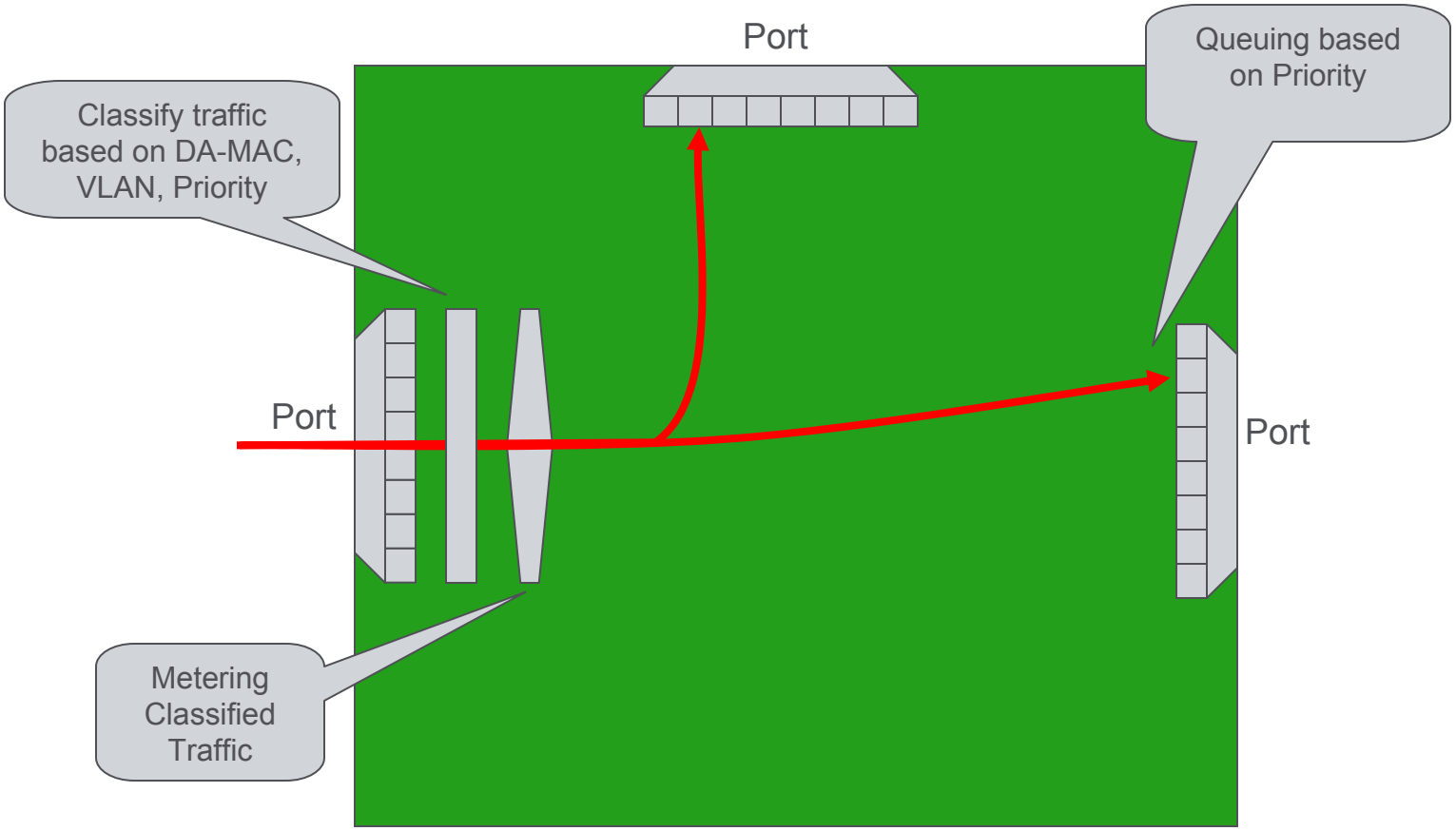
- How is it measured? Kbps, Mbps, Gbps or Percentage Bandwidth?
- Over what time quantum?

No way to assign a meter to classified traffic

No way to know how many meters a bridge implements

Meters should operate on all received frames, so where does it fit in the forwarding process?

Current Metering and Queuing



DiffServ Models

ProCurve Networking
HP Innovation

RFC 2475

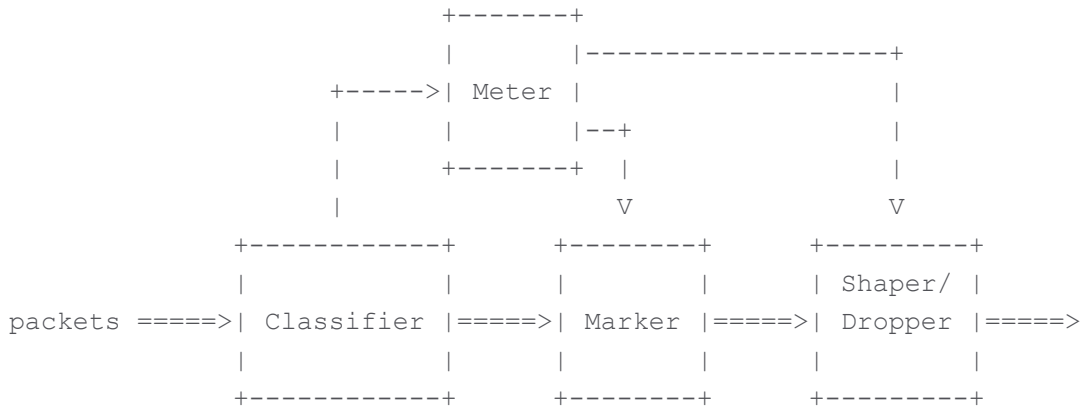


Fig. 1: Logical View of a Packet Classifier and Traffic Conditioner

RFC 3290

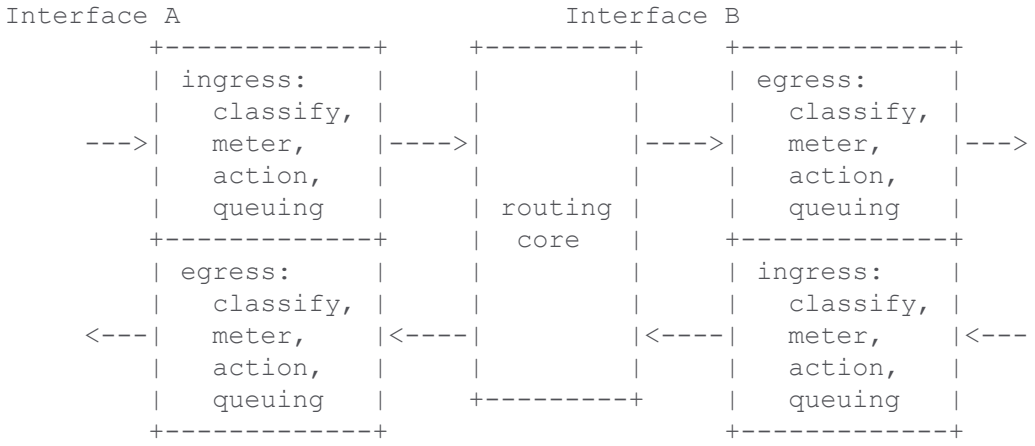


Figure 2. Traffic Conditioning and Queuing Elements

So where do we go from here?

ProCurve Networking
HP Innovation

We could… (but not limited to)

- Do nothing, everything is fine.
- Consider alternative scheduling algorithms that…
  - Provide minimum bandwidth guarantees for egress scheduling
  - Provide maximum rate limits on egress scheduling
- Schedule transmission based on classifications other than priority
  - Match metering classifications
- Define metering configuration values
- Adopt a forwarding model that aligns with the DiffServ model and MIB
- Define signals from bridge queuing to trigger forward congestion notifications (e.g. enable the setting TCP/IP ECN)
- Combinations of the above
- Others…

Recommendations

ProCurve Networking
HP Innovation

Time to do something

Need to keep it simple (don't define algorithm variables, just results)

Work in support of current internet models

Acknowledge shipping solutions and avoid obsolescence

Evaluate scheduling algorithm work done in MEF, IETF, DSL Forum TR-059, ITU

Consider the following three modifications to clause 8.6.7 and 8.6.8:

1. Minimum bandwidth guarantees for Egress classes
2. Maximum bandwidth limits for Egress classes
3. Configuration of Ingress meters

Recommended Modifications

ProCurve Networking
HP Innovation

1. Egress minimum bandwidth guarantees
   - Define 8 parameters per port as minimum bandwidth percentages
2. Egress rate limits
   - Define 8 rate parameters per port that match metering rate definitions
3. Metering rate definitions
   - Define way to represent rate
   - Define quantum over which rate is measured