RPR MAC Transit Path Design

>>

Sanjay K. Agrawal, Luminous

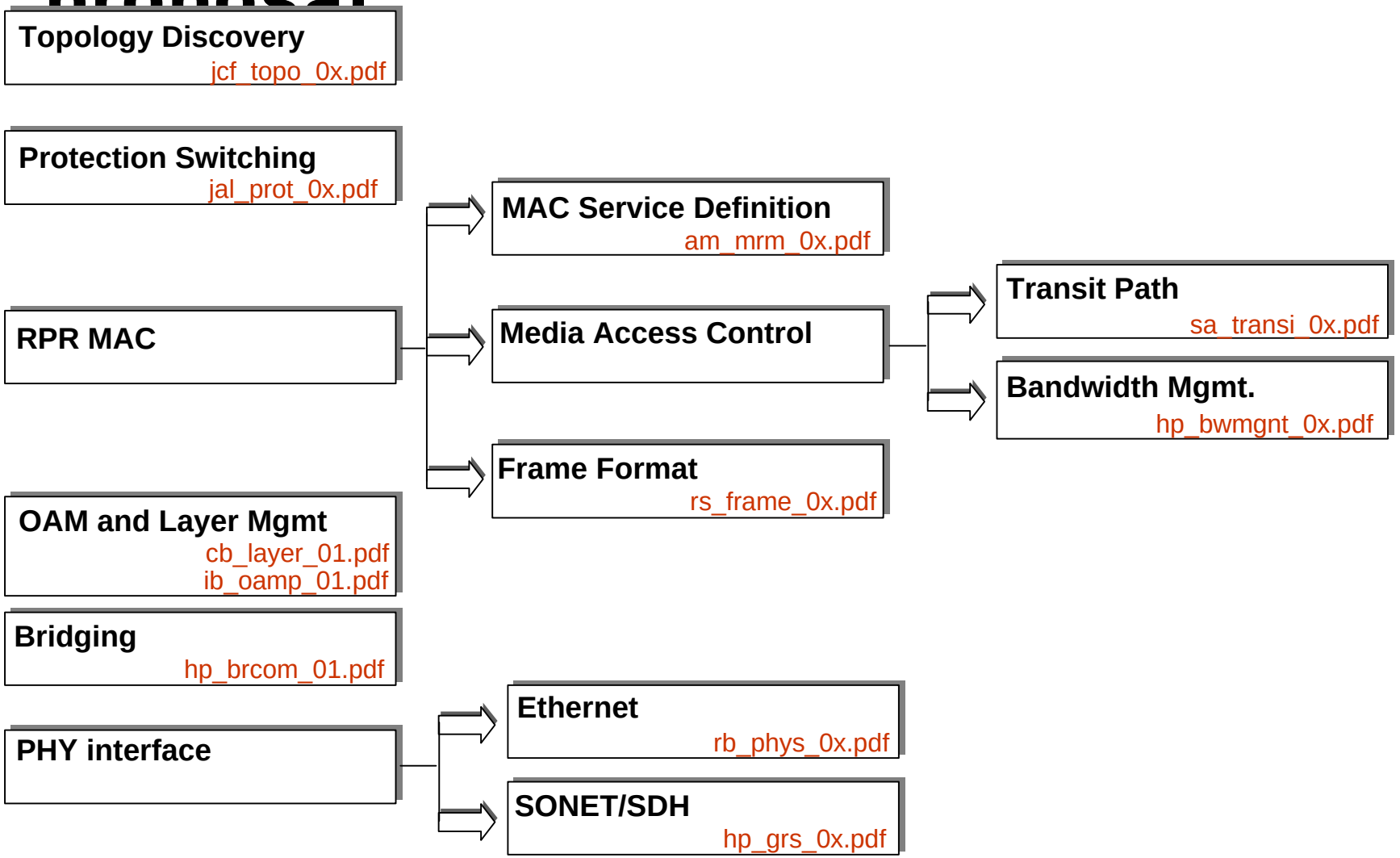
Jean De_Jaegher, Alcatel

Adisak Mekkittikul, Lantern

Harry Peng, Nortel

Fredric Thepot, Dynarc

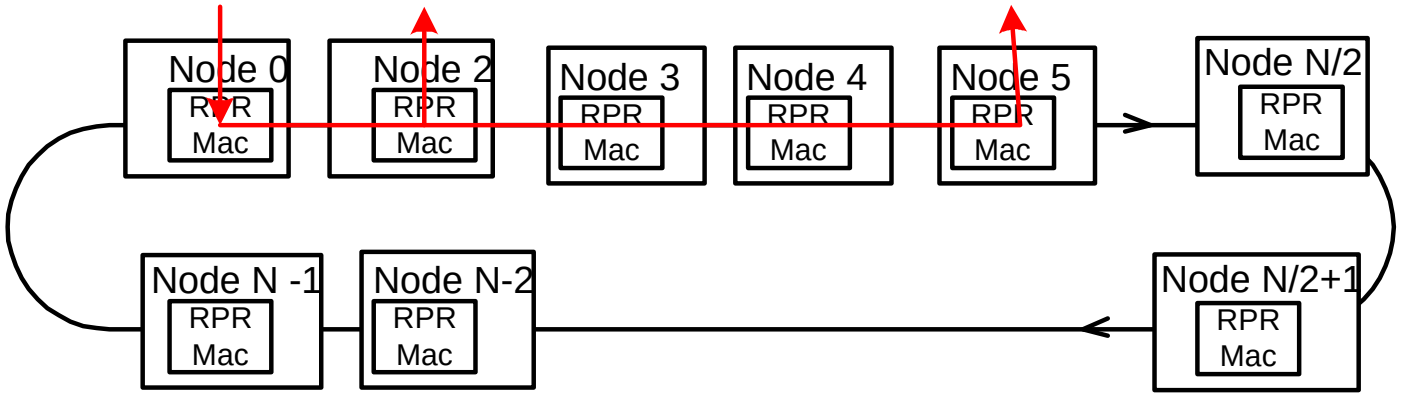
Components of a complete RPR proposal



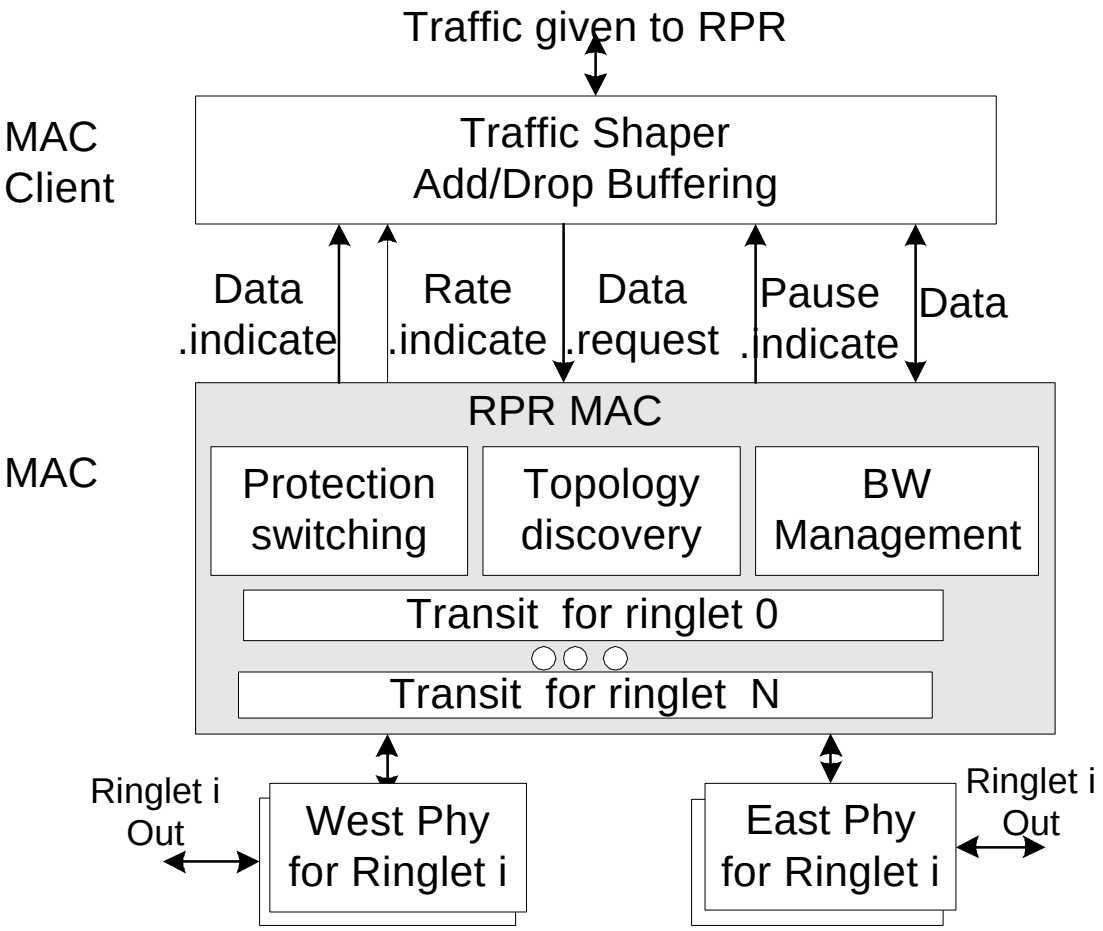
Outline

- RPR System Architecture
- RPR MAC Requirements and Objectives
- RPR MAC Framework Proposal
- RPR MAC Transit Path Design
- Optional Modes of Operation
- RPR MAC Client Add/Drop Path Design
- Conclusion

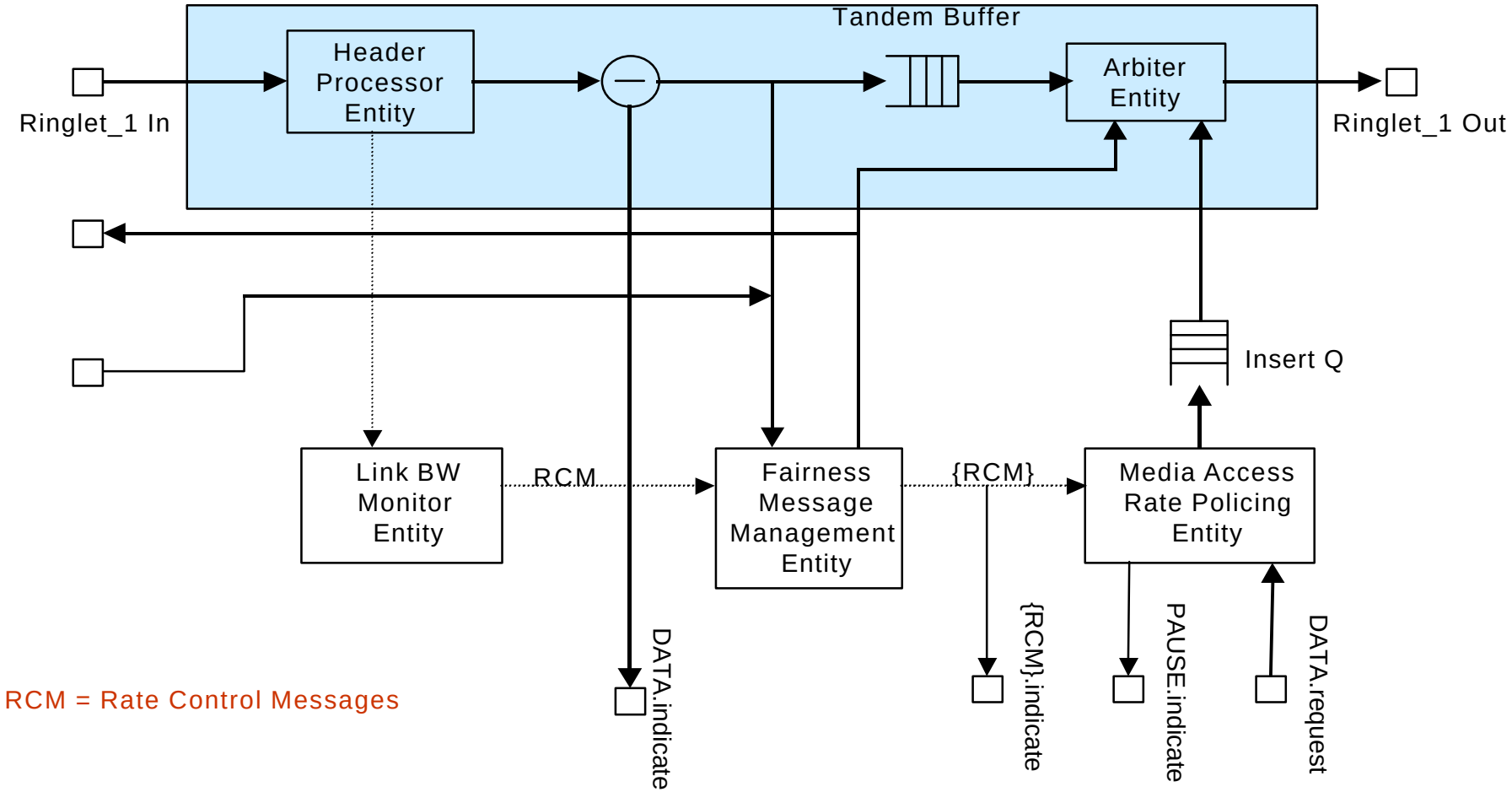
RPR Network


Node 0
RPR Mac
Node 2
RPR Mac
Node 3
RPR Mac
Node 4
RPR Mac
Node 5
RPR Mac
Node N/2
RPR Mac
Node N -1
RPR Mac
Node N-2
RPR Mac
Node N/2+1
RPR Mac

RPR System Architecture



Signaled Mode: MAC Control Path Architecture

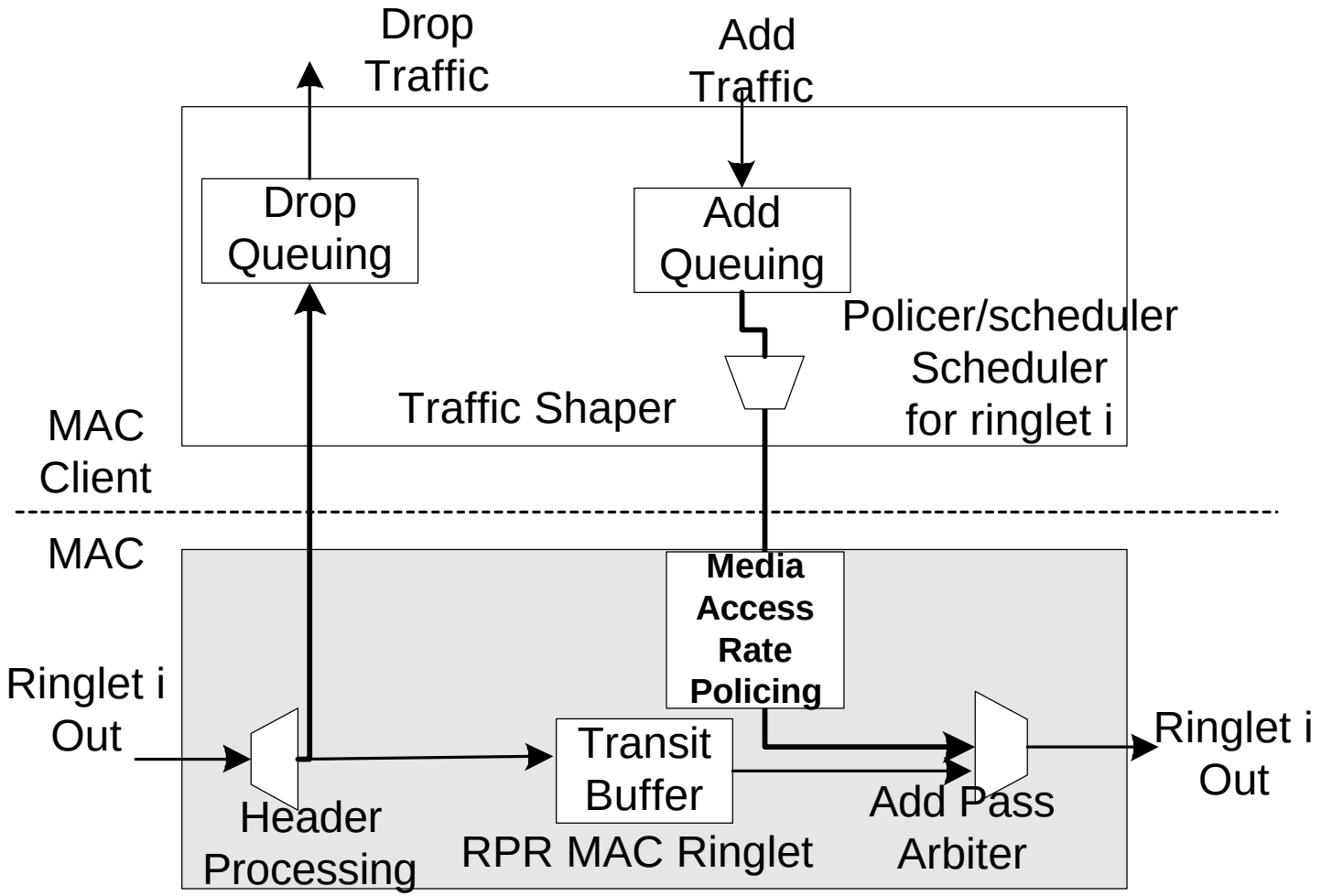


RCM = Rate Control Messages

Objectives and Requirements for the Transit Path

- The transit path is part of the shared medium
- The transit path is lossless.
- The transit path implements destination and source stripping.
- The transit path implements broadcasting and multicasting: drop and pass mode
- Minimal buffering in the transit path
  - Minimize the cost of the standard RPR MAC chip saving memory cost
  - Minimize delay in the transit path
  - Maximize scalability as RPR MAC chip scales at higher-speed and multiple ringlets.

RPR MAC: Transit Data Path



RPR MAC Reception rules

- When a frame arrives at RPR MAC, the DA MAC address is matched with the RPR database in the header-processing block.
- The decision to strip or bypass the frame:
  - If the frame DA matches in the RPR database
    - The Frame is stripped from the ring.
  - If the frame DA is a broadcast, multicast
    - If TTL > 1, the frame is both stripped and copied.
    - If TTL=1, the frame is stripped
  - If the frame SA matches the RPR MAC database,
    - The frame is stripped, and discarded.
  - If the frame has a bad HEC on the RPR MAC header,
    - The frame is stripped and discarded. A bad HEC counter is incremented.
  - If the DA MAC address of the incoming frame does not match the RPR database and TTL <=1
    - The frame is stripped and discarded.

RPR MAC Reception rules (cont…)

- Else, the frame is passed through.
  - The TTL field in the RPR MAC header is decremented by one.
- Reception of only frame header needed for forwarding decision.
- Promiscuous Mode:
  - RPR MAC allows all the transit traffic to be received to the MAC client.
  - Promiscuous mode of operation is necessary for debugging operations.

RPR MAC Transit Rules

- Transit frames are sent to the transit buffer.

The scheduling algorithm:

Step 1: Choose a frame to be transmitted

If a transit buffer has a frame ready

Choose a frame from the transit buffer

Else if an insert buffer has a frame ready

choose a frame from the insert buffer

Step 2: transmit the chosen frame with no pre-emption

Step 3: complete the transmission, repeat step 1

- In the store and forward mode of operation transit frames are received entirely before they are sent out.
- Minimum buffering needed in the transit buffer for the transit frame is single MTU for contention resolution between add and pass frames.

RPR MAC Discard Rules

- HEC is incorrect
  - The frame is discarded.
- Source MAC address matches the RPR MAC database in Header Processing block,
  - The frame is discarded.
- TTL expired
  - The frame is discarded

RPR MAC Add Rules

- Add frames are sent to the RPR MAC through data.request primitive.
- RPR MAC inserts the frame into the medium:
  - There is no packet under transmission.
  - Transit buffer is empty.
  - Media access rate control has not asserted PAUSE.
- Media Access Rate Control
  - PAUSE shall be asserted to prevent the MAC client from exceeding the allocated bandwidth on any segment downstream.

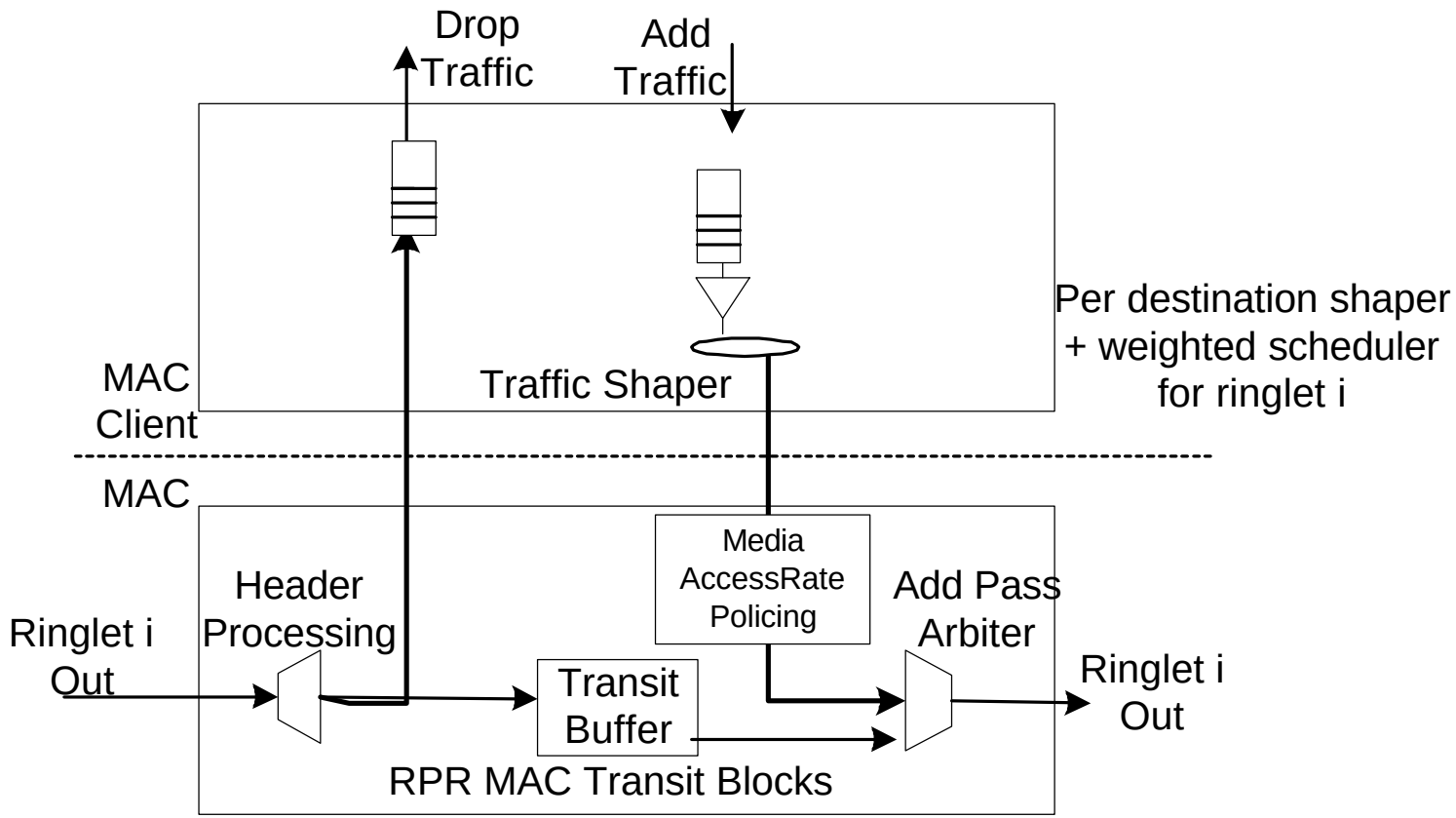
Support for Virtual Output Queuing

- Objective: maximize the spatial reuse and the link utilization for frame flows with arbitrary (source, destination) pairs.
- Problems if MAC proposal has no VOQ support:
  - MAC sets the access rate low to satisfy the bandwidth allocated by one congested destination
    - Severely limits the access rates to other uncongested destinations.
  - (HoL) blocking problem occurs in a single queue access.
    - Frame destined to uncongested destination waits behind an frame congested destinations.
- Proposed Solution
  - Signaling messages propagates independent media access rate control for each ring segment in the RPR MAC.
  - Virtual Output Queuing (VoQ) in the MAC client
    - Dedicated output queue for each destination.
    - Dedicated output queue for unknown destination.

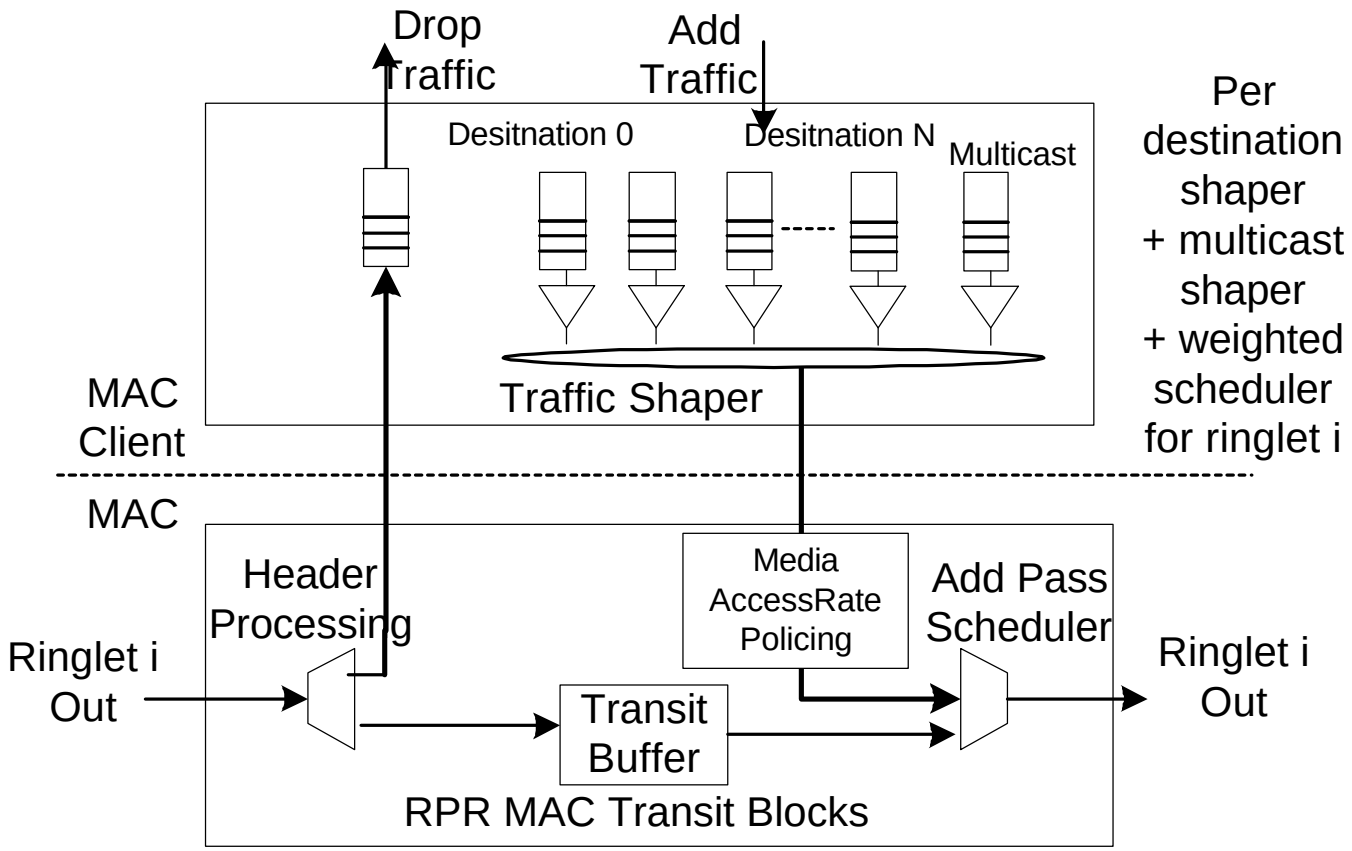
Optional modes of operation and transit buffer considerations

- **Cut-through Mode:**
  - Frame transmission can begin before it is entirely received.
  - RPR header should be received entirely before beginning transmission out of the outgoing ringlet, since the header has to be processed.
    - Reduces the delay that frames experience in the transit path.
- Store & forward mode:
  - Frame is entirely received before it is considered for transmission.
  - This mode of operation allows FCS errored frames to be stripped and transit error counter incremented.
    - Eliminates degraded frames in the transit path at instance of FCS error.

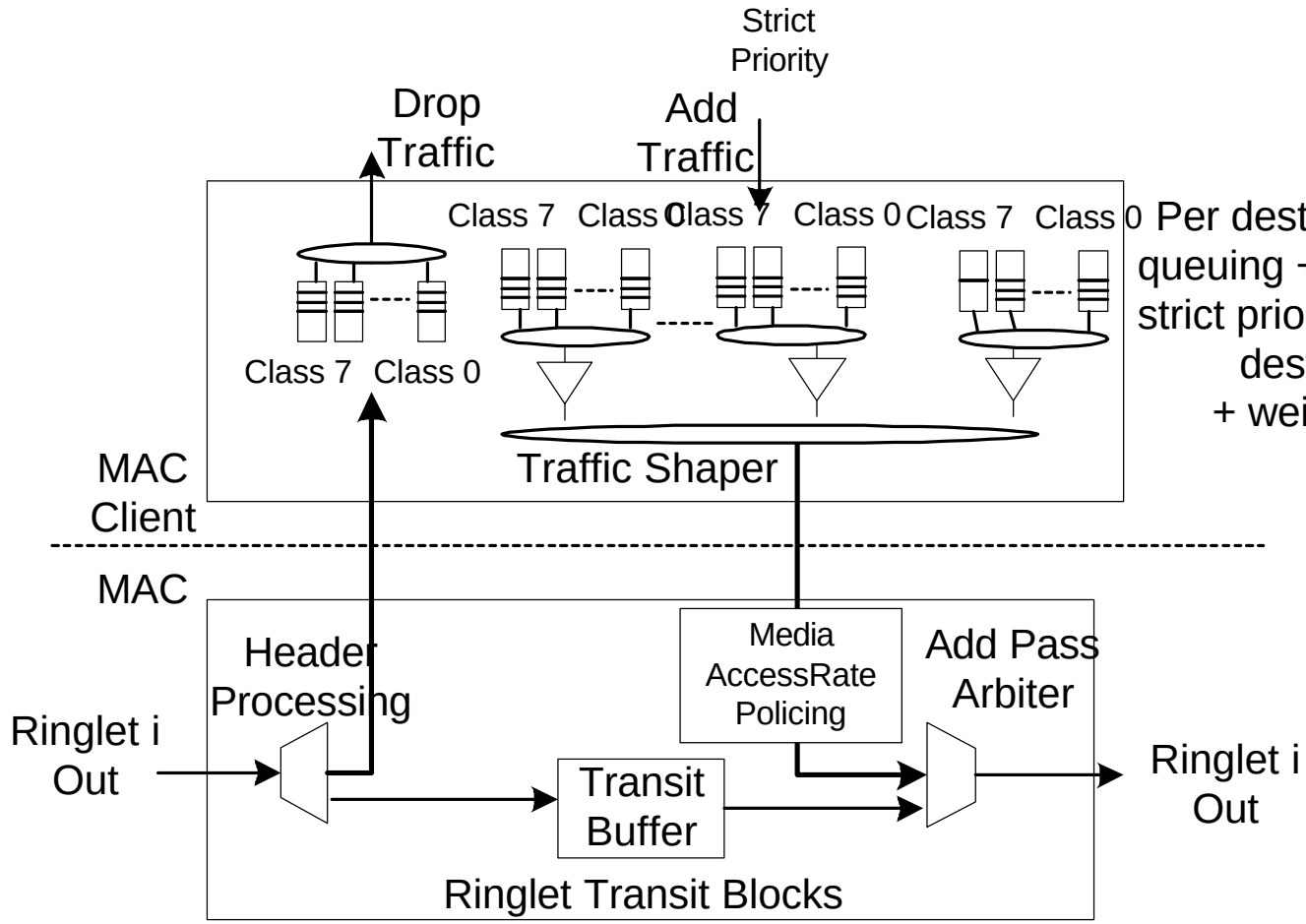
MAC Client Add/Drop Path Options



MAC Client Add/Drop Path Options: Virtual Output Queuing

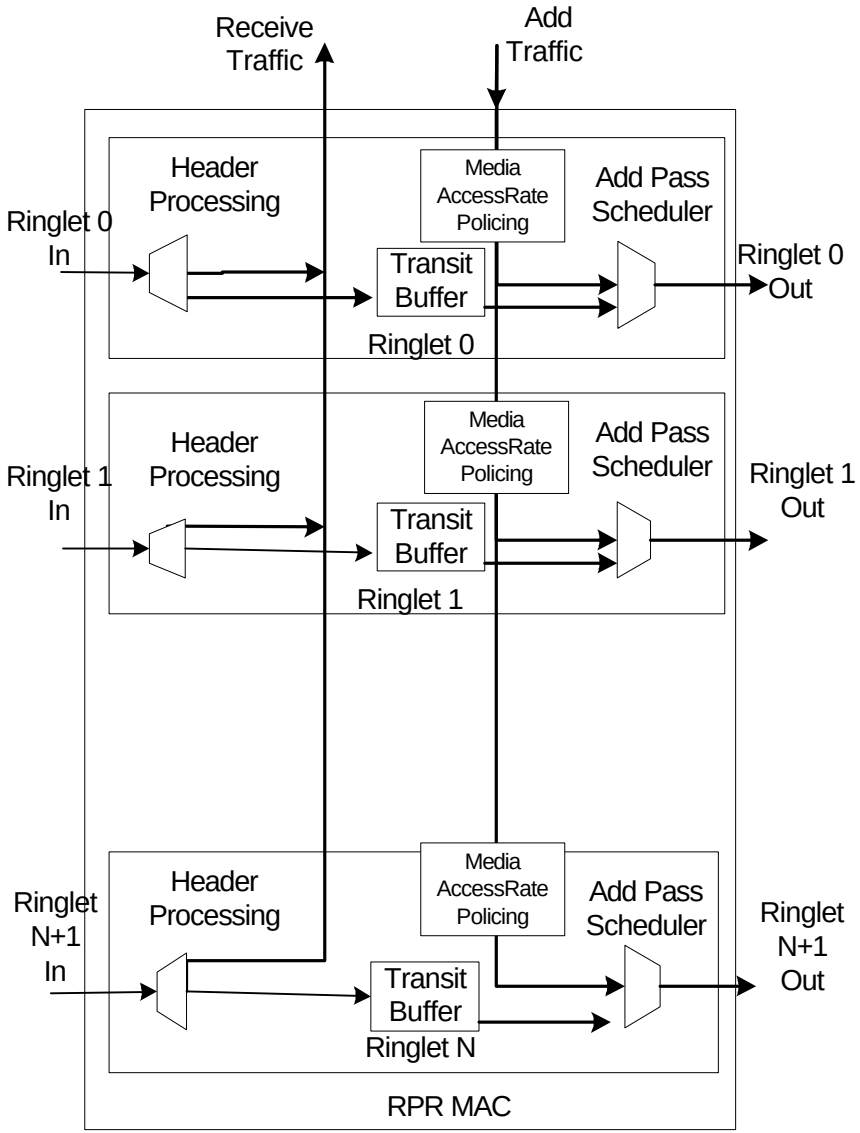


MAC Client Add/Drop Path Options: Class of service based queuing per destination



Per destination class based queuing +multicast queueing, strict priority scheduling + per destination shaper + weighted scheduler for ringlet i

Transit Path: Multiple Ringlets Option



Conclusion

- Conforms to the 802 shared MAC medium.
- Scalable for high-link speeds.
- Cost effective solution that minimizes the cost of silicon implementation.

Media Access Rate Control

At each 10usec interval

for each link segment

calculate the node (for this MAC) allowed BW, fj.

*fj = rj + wj*RCF*

give credit for each segment

if (segment_credit) < 15,000,000

segment_credit += fj

if (segment_credit) < 0 // client BW exceeds limit

assert PAUSE.indicate

end FOR

At each DATA.request

if no PAUSE.indicate asserted, accept DATA.request

for each segment between this and the dest nodes

deduct segment credit

*segment_credit -= frame_length*10,000*

end FOR