Merits of Open Loop

Siamack Ayandeh
sayandeh@onexco.com

Onex Communications Corp

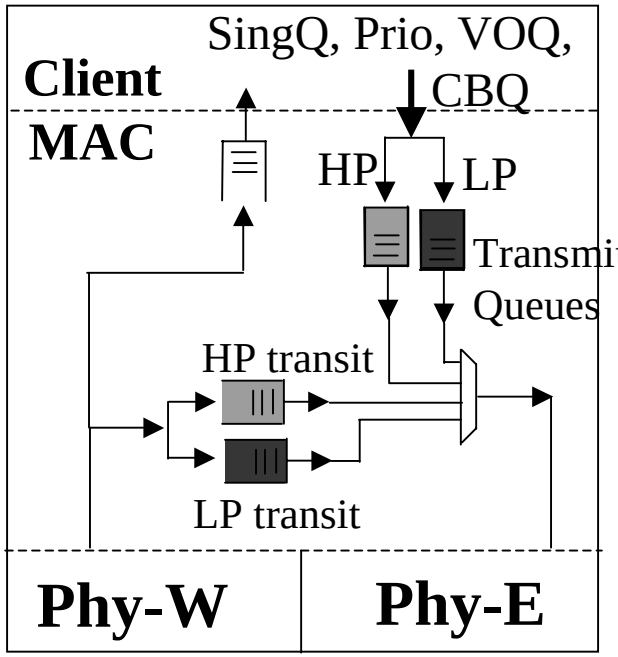
a subsidiary of

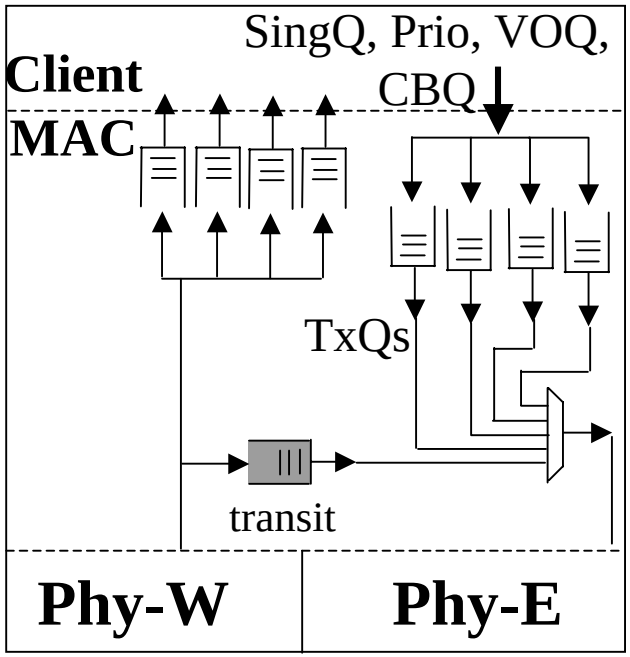
TranSwitch Corp.

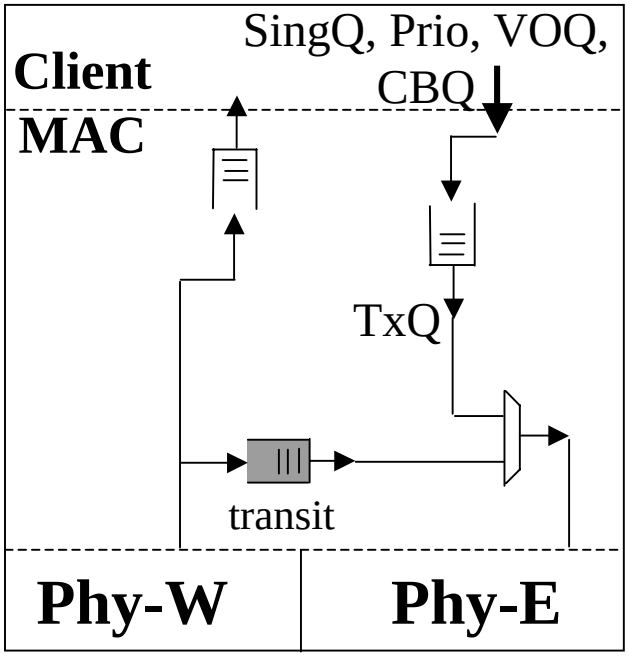
Outline

- Allows for dynamic partitioning between the High and Low priority traffic
- No HOL blocking issues
- Relatively low configuration and operational complexity
- Comparable performance to CA
- Not prone to link aggregation issues

3 Flavors of CoS







CoS capable MAC
HP Txm or transit
has priority over LP class

CoS access to MAC
Classless ring
Transit priority over Transmit

Classless MAC
Transit priority over Txm

HP MAC e2e delay is reduced by CoS capable transit & txm queues
Question is by how much?

Static Partitioning of High & Low Priority Traffic

Example of Static Partitioning i.e. no stat-muxing

High Priority Class

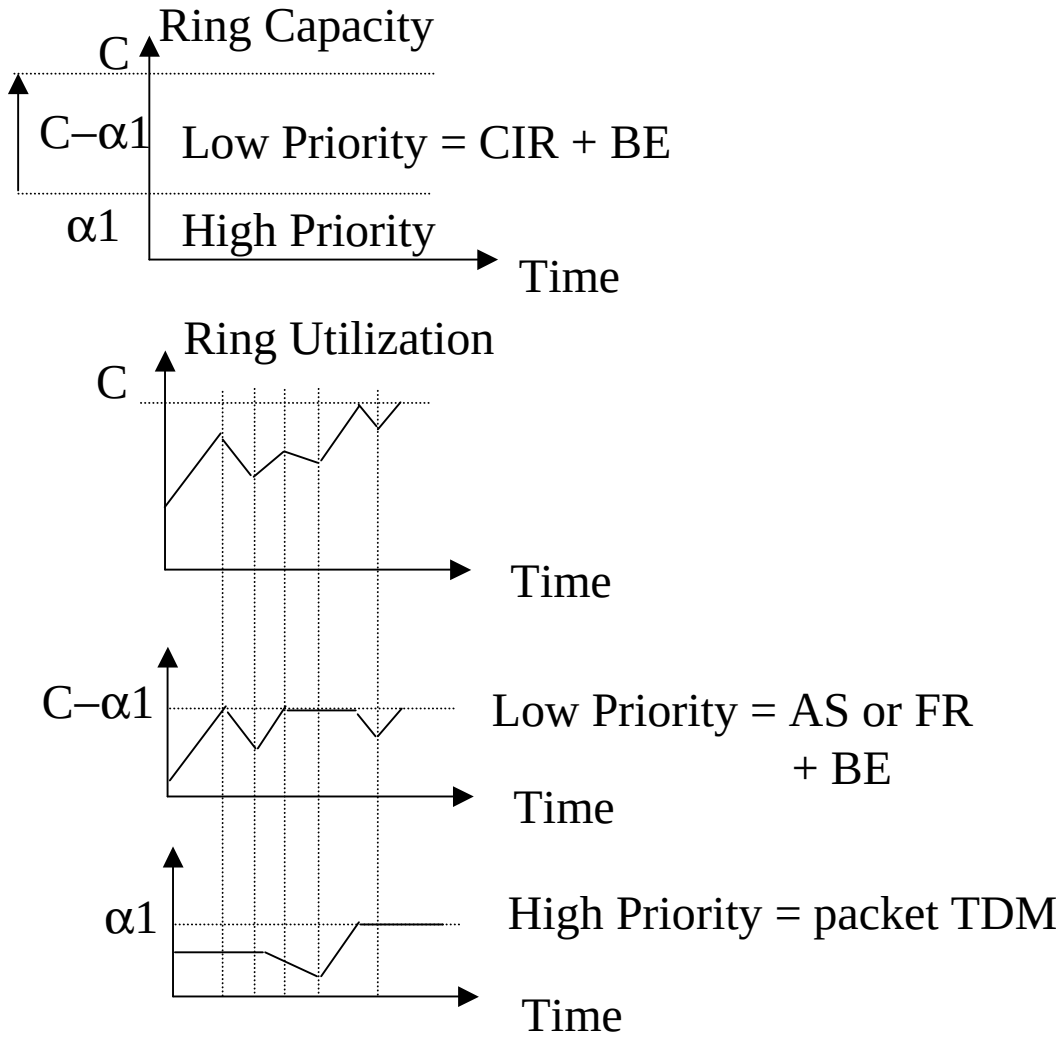
e.g. PacketTDM

- Bandwidth reservation (peak rate is enforced)
- Delay guarantee

Low Priority Class

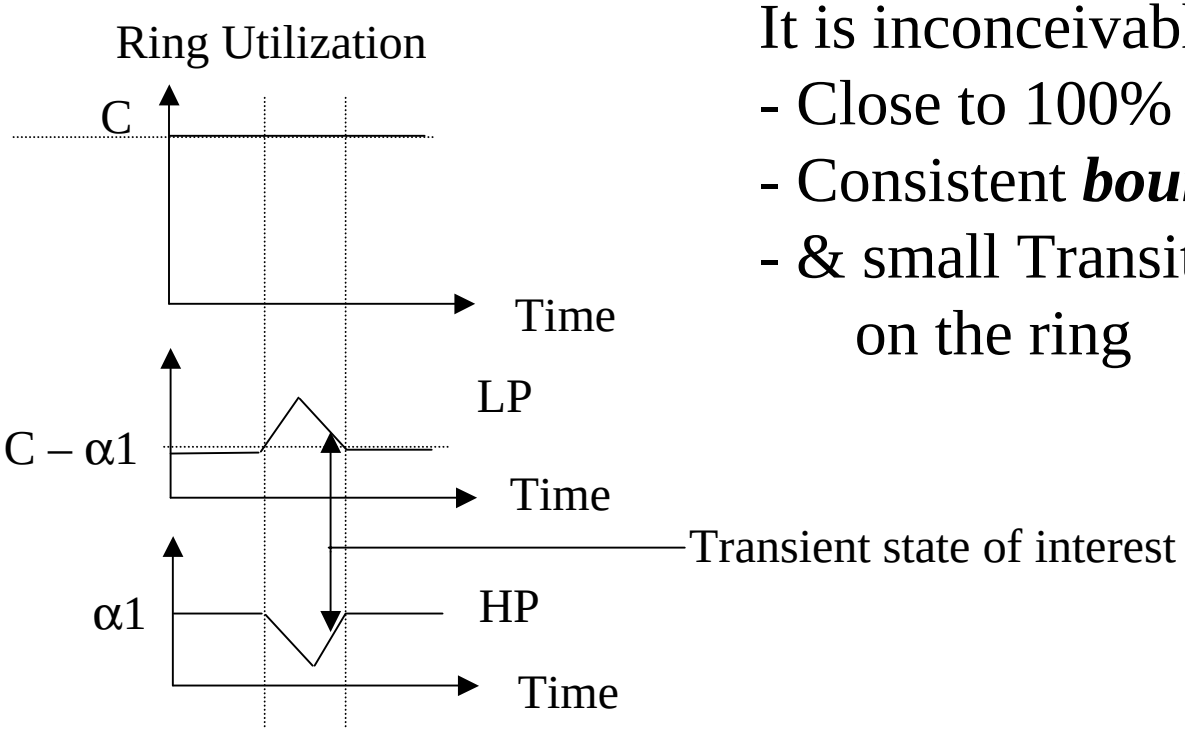
e.g. Frame Relay

- Bandwidth reservation (CIR)
- Ability to burst in excess of CIR (carried as best effort)
- No delay guarantee



Dynamic Partitioning

Dynamic Partitioning allows stat-muxing



It is inconceivable to have:
- Close to 100% ring utilization
- Consistent ***bounded delay*** for HP
- & small Transit Buffers with no loss on the ring

Open loop caters to dynamic partitioning, CA may not

What's meant by bounded delay

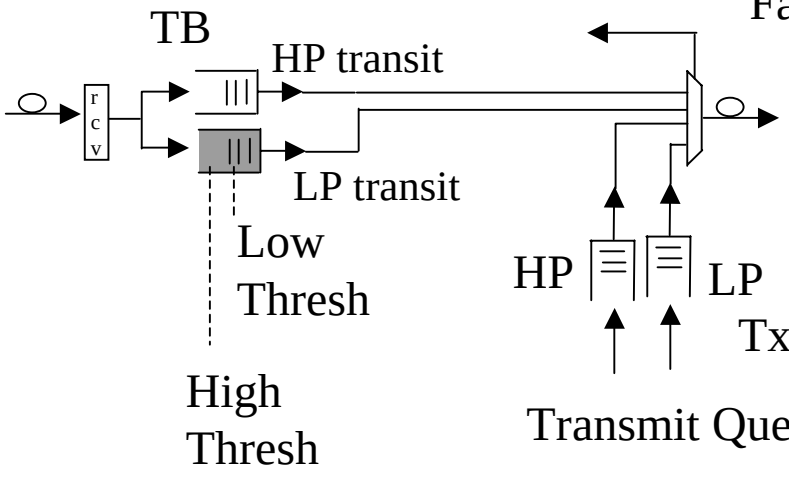
- There is an upper limit on MAC e2e delay of High Priority packets
- This upper bound can be controlled by resources provisioned for HP class only
- Min and Max-plus algebra e.g. can be applied to derive analytic bounds
- Results of this analysis can be used by service providers to control HP class performance

3 examples of congestion avoidance

References

- SRP-fa, Spatial Re-use Protocol
  - rfc2892
  - Conexant SRP MAC overview
  - SRP-fa performance evaluation 3/14/01
- iPT-CAP, Inter WAN Packet Transfer
  - iPT
  - iPT-CAP 07/11/00
  - iPT fairness CAP simulation report
- VOQ-aware MAC
  - Proposed VOQ-aware MAC 05/01
  - Simulation Results 03/12/01

SRP-fa



Fairness signal = If (congested)

{ rev_usage = MIN (lp_my_usage, rcvd_usage) }

// fair_usage = min(my_usage_on_ring)

Table of SRP Scheduling Order **?**

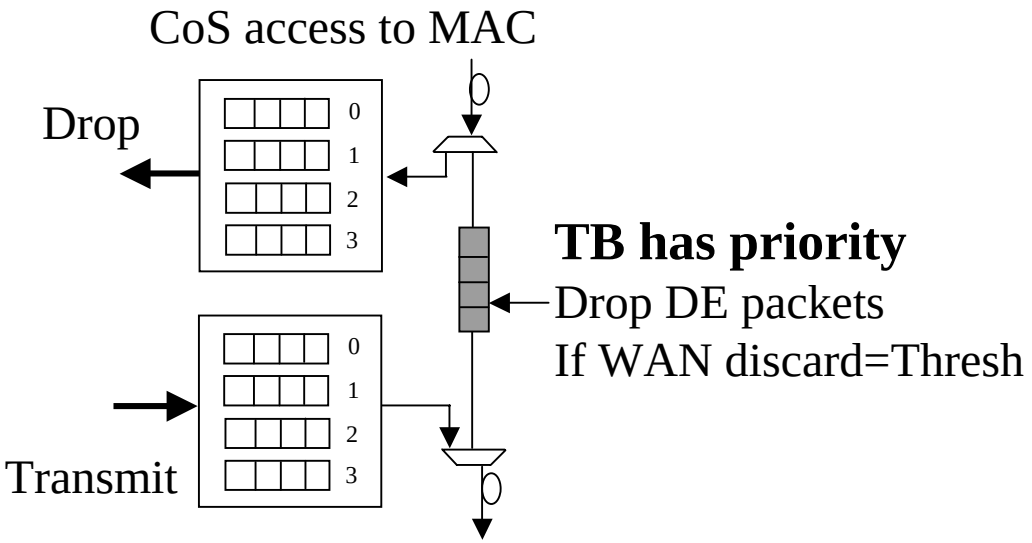
! congested && (my_usage < allow_usage)	congested || {(lo_tb_depth>0) && (my_usage > fwd_rate)}	(lo_tb_depth> TB_HI_THRESH)
HP transit HP host	HP transit HP host	HP transit
LP host LP transit	LP transit	LP transit

SRP-fa Engineering Parameters

2 options

- (LP_HI_thresh – LP_Low_thresh/2) >= bytes in transit (i.e. large enough TB allows dynamic partitioning)
- (LP_HI_thresh – LP_Low_thresh/2) < bytes in transit (i.e. Host HP MAC access delay for HP class is ***un-bounded***)
- Un-bounded means that HP class delay depends on traffic from other classes

iPT-CAP

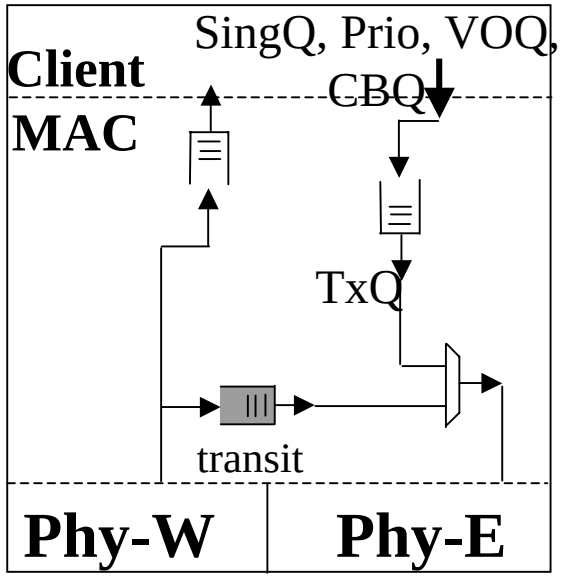



```
If (!congested)
  { target_rate = C';
    advertise (target_rate);
  }
else
  { detect (#active_stations);
    target_rate = C'/#active_stations;
    advertise (target_rate);
  };
```

- C' = C – α1 leads to static partitioning
- Seems to be the only way to bound high priority delay

VOQ-aware MAC

- MAC is classless,
- $$f_i = r_i + w_i \frac{\left(C - \sum_{active} r_i\right)}{\sum_{active} w_i}$$



- Where f_i is the BW share of station-i on a segment & is sum of it's committed access rate (r_i) + its share of excess ring bandwidth
- It seems to be a case of un-bounded delay

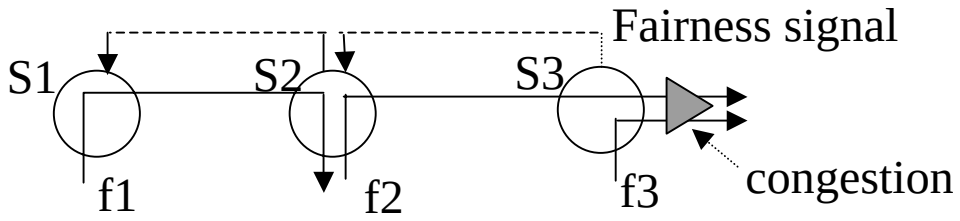
Conclusion 1

- High priority ring access delay may not be bounded when using congestion avoidance
  - Low priority transit gets through first
- Avoidance algorithms/weighted fairness if applied to low priority traffic only
  - Lead to static partitioning of ring bandwidth between high and low priority traffic

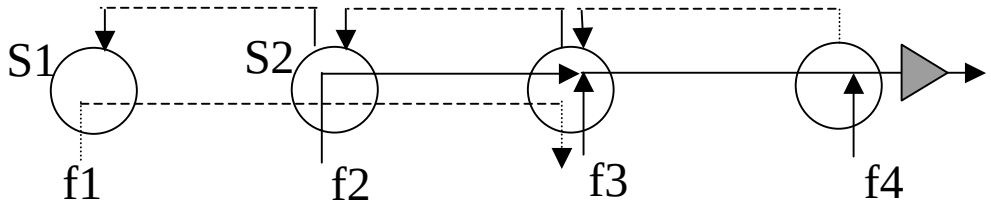
No HOL Blocking

- With open loop only connections which cross the congested link are throttled
- Congestion avoidance on the other hand exhibits HOL blocking in one or two flavors
  - Un-intended throttling of stations
  - Un-intended throttling of add/host traffic (Adisak's quiz 05/01)

HOLB: Station Throttling

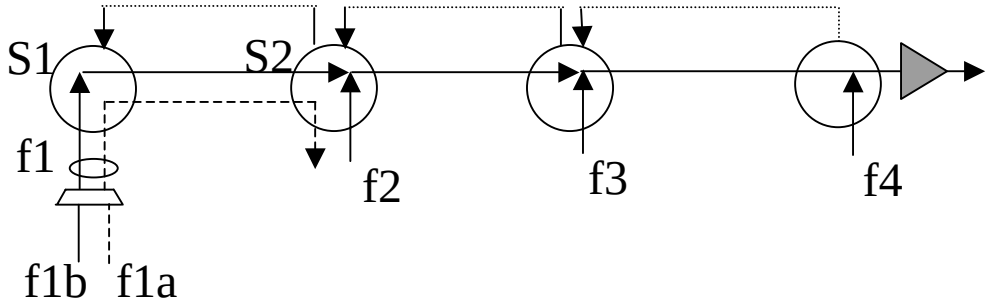


- SRP has mechanism to allow for spatial re-use i.e. if at S2 (allow_usage>fwd_rate); f1 is not throttled



- f1 however is throttled to bottleneck rate (1/3 vs. ½) as (fwd_rate>allow_usage) at station-2
- Solutions based on global state require per segment monitoring and state, and dissemination of all this info to VOQ clients which may not know the ring segment topology after all

HOLB: Host Throttling

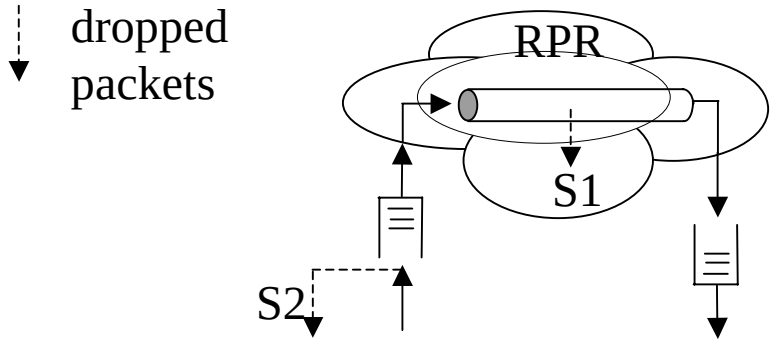


- f1 host is rate shape limited based on bottleneck rate which is due to f1b + f2 + f3 + f4 + …
- f1a is therefore denied full access to the ring while resources are available on S1-S2 span

Lower Configuration & Operational Complexity

- Weighted fairness by definition requires global knowledge of two parameters per station
  - committed bandwidth per station (r_i)
  - weight of station (w_i)
- Global knowledge requires identical copies of two tables at every station $\{r_0 \ldots r_n\}$ & $\{w_i \ldots w_n\}$
- A change in r or w has to be communicated to all stations

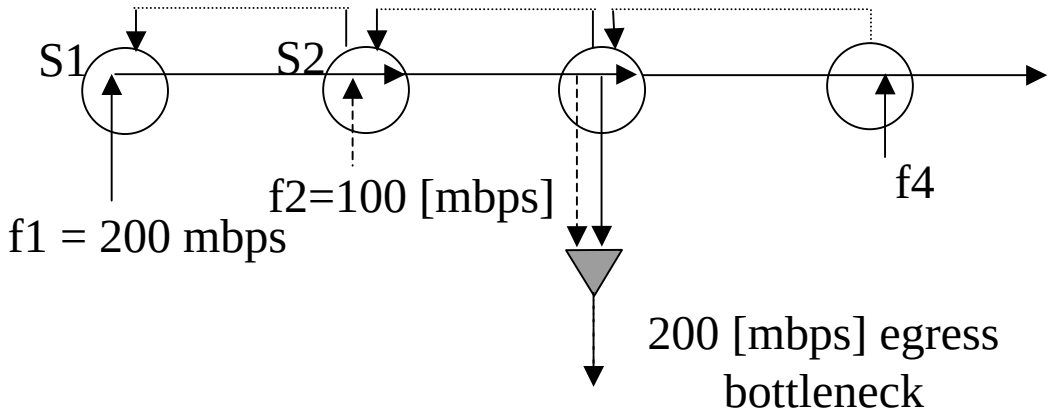
Comparable Performance



Performance not likely to be a differentiator

- Suitable metric for comparing open loop & CA is client good-put
- TCP drops 6-8% of it's traffic irrespective
  - Open loop drops at the congested link S1
  - CA drops at the RPR MAC client layer S2
- Rings are overbooked by factor of 4, 20, or more
  - Therefore there may be little or no excess bandwidth to allocate by fairness schemes any way
  - Provisioned traffic at each station is what gets through
  - Excess bandwidth is dynamic, so getting less of it is equivalent to quiet stations reclaiming their committed share

TCP fairness at congested egress



- Consider the congested egress scenario, where TCP is the only mechanism at work
- Depending on the number of TCP flows constituting f1 & f2 bundles, the egress rate of each flow would vary
  - And is not controlled by any MAC fairness schemes
- Simulation studies should include TCP clients & compare avoidance schemes being on or off

Open Loop Fairness

- Open loop offers “fairness” as controlled by TCP in the face of congestion
- If it’s good enough for the rest of the network, it’s good enough for RPR
- No need for global knowledge of weights or rates
- Provisioning is weighted, while allocation of excess bandwidth is on a best effort basis
  - suffers from station location advantage, hence is fair with dynamic traffic patterns i.e. premise behind spatial re-use
  - is impacted by the number of contending TCP connections
  - IMHO weighted best effort offered by CA is contradiction in terms

Merits of Open Loop

Open loop

- Offers dynamic bw partitioning & CoS capable MAC
- No HOL blocking
- Low configuration complexity
- Best effort access to excess bandwidth
- Works with link aggregation

Congestion Avoidance

- Choice is between dynamic partitioning, un-bounded delay, or small transit buffers
- 2 flavors of HOL blocking creates congestion
- Needs global state and topology aware client
- Weighted access to excess bandwidth. Is not activated when there is no excess bandwidth e.g. with overbooked rings, or when congestion is at egress
- May need design modification to deal with link aggregation, etc…