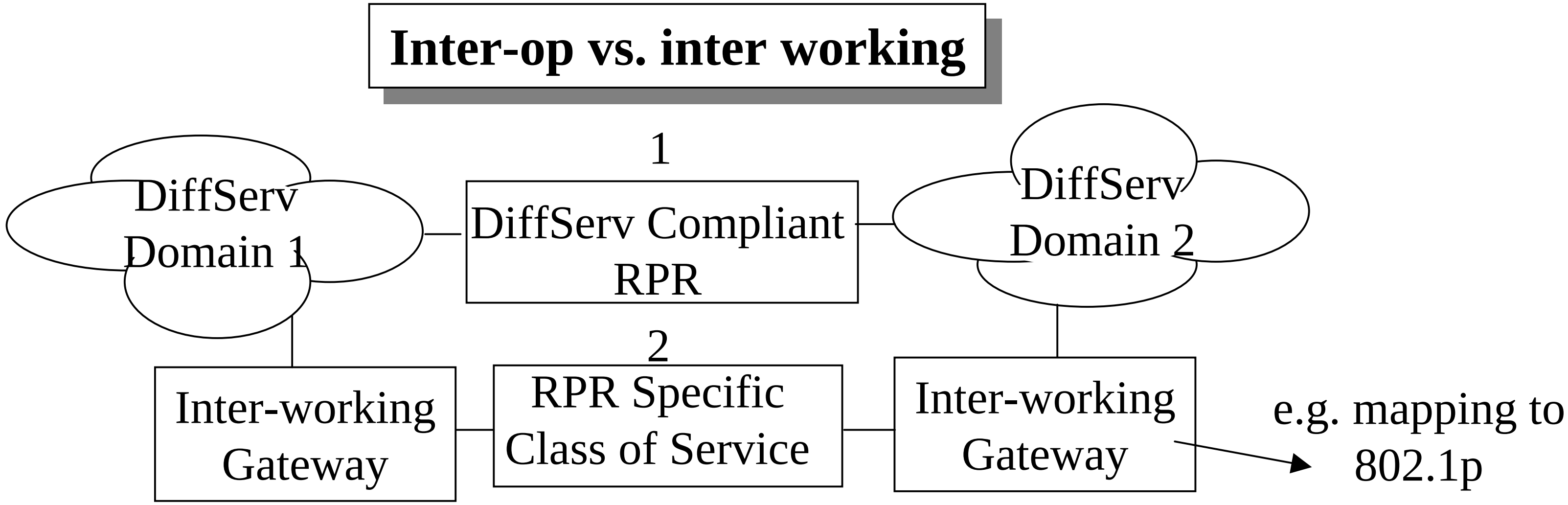
IP Differentiated Services Requirements for RPR

Siamack Ayandeh
sayandeh@onexco.com
Igor Zhovnirovsky
igorz@onexco.com
Onex Communications Corp

Outline

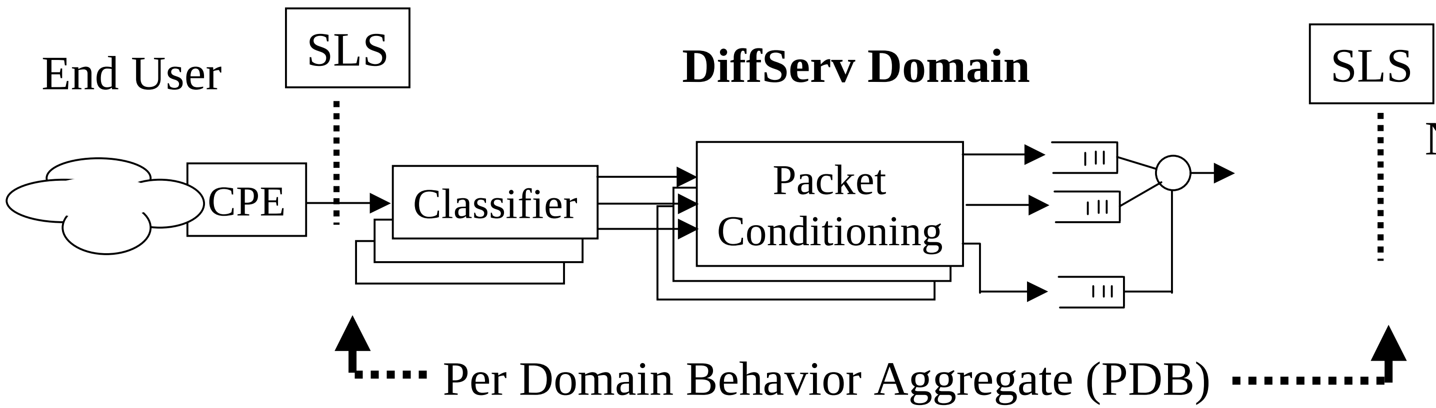
- Inter-operability or inter-working
- IP Differentiated Services background
  - Why DiffServ
- IP Differentiated Services requirements for RPR
  - Bandwidth allocation & scheduling
  - Buffer management & packet drop criteria
- RPR options in supporting Differentiated Services
- A phased approach

Reference Model for DS Compliant RPR



- First option is similar to RPR being the media within a DiffServ domain
- First option is the focus of this presentation
- First option leads to interoperability which is desirable vs. inter-working gateways

Differentiated Services Architecture



Classification	Packet Cond	Per Hop Behavior	Per Domain BA
- Behavior Aggregate	- Metering	- Expedited Forw	- Routing
- Multi-field classification	- Marking	- Assured Forw	- Network metrics e.g.
* IP version	- Dropping	- Class Selector	max hop count, edge to
* Src/Dst IP address	- Shaping	Code Pnt	edge delay/jitter
* Protocol type		- Default Forw	- How to measure metrics
* Src/Dst Port #			- Example use to create
* etc.			services

80/20 rule favors single DS domain deployments

7 Reasons to use DiffServ

- There is more to offering service differentiation than scheduling bandwidth and managing buffers
- Context for service differentiation includes:
  - Service definition and pricing models
  - Service sale, activation, and change
  - Provisioning & configuration of network elements
  - Service monitoring
  - Accounting and billing support
- IETF has spent over three years, several dozen drafts, and Gigabytes of email to move DiffServ forward

7 Reasons to use DiffServ

- All power to service providers
  - Control over allocation/partitioning of resources
  - Control over service definition
- Was designed for data centric networks
- Specifies enough to achieve inter operability while allowing vendor differentiation in implementations
- TTM, re-use, and simplicity

DiffServ Building Blocks needed in RPR

Per Domain Behavior Aggregates (PDB)	Per Hop behavior (PHB)	Characteristics & likely services
Virtual wire	Expedited Forwarding (EF)	**Low delay, jitter, loss** VLL, voice
Assured rate	Assured Forwarding (AF) {AF11, AF12, AF13} gold {AF21, AF22, AF23} silver {AF31, AF32, AF33} bronze {AF41, AF42, AF43}	**Controlled Overbooking,** Gold, silver, Bronze
Yet to come	Class Selector Code Pnts {CSC 1..7}	**Legacy TOS, Control & network Traffic**
•••	Best Effort Bkcg	etc.

DS Forwarding Requirements

	MAC	$\overline{\text{MAC}}$
• **Minimum bit rate**		
– Required by both EF & AF per hop behaviors	✓	✓
– EF also requires low delay, jitter, and loss	✓	✓
• **Active queue management**		
– Assured forwarding requires properties of a WRED like algorithm, avoiding tail drop	?	✓
– Some amount of buffer space	✓	✓

Map DiffServ to 802.1p

Per Domain Behavior Aggregates (PDB)	Per Hop Behavior (PHB)	802.1p Assuming 7 queues
Virtual wire	EF	Q6, voice
Assured rate	AF_{11}, …, AF_{33}	Q3, 4, 5 Exc effort, CL, Video
Yet to come	CSC 1..7	Q7, Network control
•	BE	Q2, best effort
•	Bkcg	Q1, background
•		

Desired mapping is up to the service provider
However it is not fully DS compliant

RPR DiffServ Requirements

- **Minimum of four Classes of Service**
  - Class 1: TDM look & feel — **$ Enforced peak rate, P2P circuit characteristic**
  - Class 2: Control & network traffic
  - Class 3: Assured service — **$ Controlled overbooking, Single ended SLAs, connectionless**
  - Class 4: Best Effort — **$ Overbooked**
- **Buffer Management**
  - Class 1: no drop due to congestion (provision $\alpha 1$ ~25%)
  - Class 2: no drop due to congestion (provision $\alpha 2$ ~ 5%)
  - Class 3: drop according to DiffServ rules (provision $\alpha 3$)
    - If these rules are kept outside of MAC layer, then MAC should not drop class 3 packets due to congestion
  - Class 4: may be tail dropped or use RED

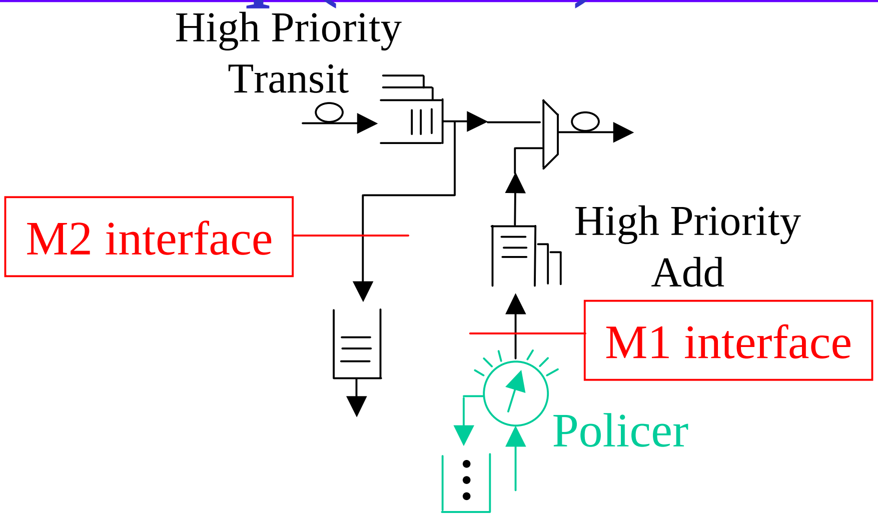
TDM RPR Requirements

- **TDM requires high (non-preemptive) scheduling priority**
  - Is peak rate limited at each station (outside of the MAC layer) and has p2p routes
  - Provisioned at (<= α1) small fraction of ring capacity, therefore no loss is expected or enforced
    - In practice expect less than 15% occupancy due to TDM traffic
  - “Fairness” = bounds on transfer delay & jitter for class-1 packets
    - irrespective of the station, port, or flow they belong to

TDM RPR Req (cont.)

Peak Rate [kbps] Per Station	Fraction of Ring
p1	α11
p2	α21
…	…
pn	αn1

$\sum \alpha_{i1} = \alpha 1$



- Note there is no mention of scheduler implementation
- Or how many queues there are
  - If transit and add are one FIFO or separate
  - Whether there are per flow queues
- Metrics are guaranteed for all the packets that pass policing (say between M1 ingress & M2 egress)

Assured Service RPR Req.

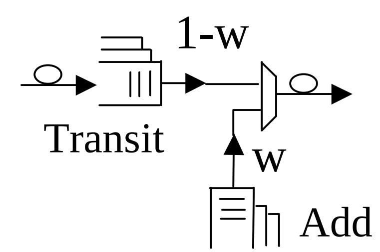
- **Assured service (AS) requires a minimum bit rate guarantee**
  - Should not starve best effort
  - Packets are marked outside of the MAC layer
    - Will the MAC or a shim header carry the markings? (need 6 bits)
  - Controlled over booking is driven by **single ended SLAs**, i.e.
    - The amount of AS traffic volume sourced per station is known
    - However the destinations may be one to one (video streaming), one to many (VPN), or one to any (Internet)
  - So volumes of traffic going to any destination are generally unknown
    - May not be able to explicitly reserve bandwidth along a given path

Single Ended SLA's

- Let's look at the sum of all the single ended SLA's per station for AS

Committed Rate [kbps]	Excess Rate [kbps]
Marked [AF_{x1}]	Marked [$AF_{x2,x3}$]
S1	S1'
S2	S2'
…	…
Sn	Sn'

Station's ring access weight
w = f (S)
w1
w2
…
wn



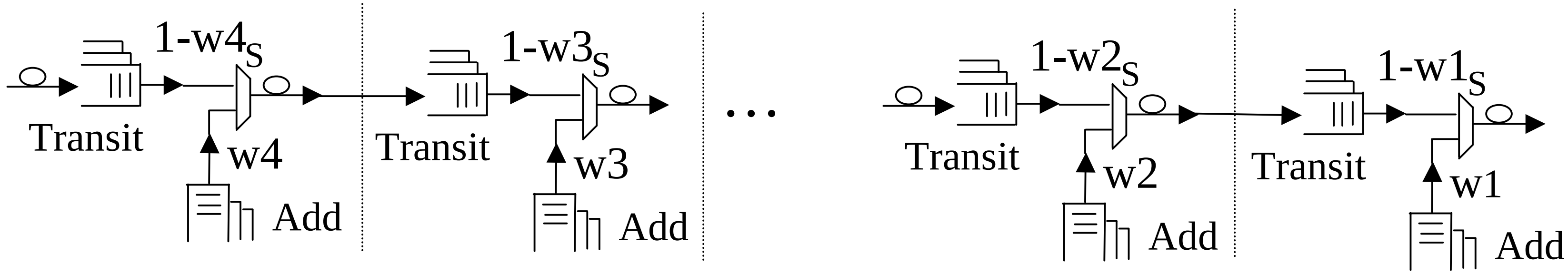
- Commitments are at the access only, often do not specify path, & may be
  - Hard with negligible probability of loss due to congestion or
  - Soft, with a given probability of loss specified in the SLA
- Excess rates are carried as best effort, and should be dropped first, using WRED like algorithm

All RPR standard needs to specify is a guarantee for a minimum rate S

Loss less Ring vs. WRED

- To support DiffServ's assured services (AS), RPR has two options:
  - Use some form of congestion avoidance and large transit buffers to avoid loss of AS packets on the ring, pushing congestion to add queues
    - Add queues would have to support WRED (presumably outside of MAC)
    - Many candidate algorithms (iPT, DPT, weighted fairness)+ are contending to achieve this goal over the next couple of years
  - Use open loop congestion control, i.e. enable transit & add queues with WRED
    - Leave congestion control to TCP
- For both options "fairness" has two components:
  - One, "weighted fairness" which is simply a minimum rate guarantee and is implemented through per station scheduling (slide #14)
  - Two, to remove station location advantage in accessing excess ring bandwidth i.e. weighted fair access to **EXCESS** ring bandwidth

Station Location Advantage



- If all the stations are accessing the ring, i.e. adding traffic, with or in excess of their allocated weight w, then fair access is ensured by per station scheduling
- If some of the stations are idle or below their weight, how should the excess bandwidth be scheduled?
- Upstream stations e.g. may have advantage in grabbing the excess bandwidth
- How this **EXCESS** bandwidth is allocated is purely a local matter and is not specified by DiffServ or any other standard
  - e.g if station-2 is idle, its share may be divided according to (w1, w3, w4) which happen to be currently active stations with traffic destined to outgoing fiber of station-1
- With **any to any traffic patterns** which is the basic assumption behind spatial re-use, different stations become upstream, and in the long run, the ring is fair
- For **hubbed traffic patterns**, the issue is persistent
- RPR WG may choose to deal with the specific hubbed scenario at a later phase

Issues with congestion avoidance

- It is difficult to tune these algorithms
  - Simulation of OC192 ring needs to mimic ~ 40 million packets/events per second and 10^5 to 10^6 simultaneous TCP connections
- Requires large transit buffers due to delay bandwidth product of the ring (ignoring nodal delay for now)
  - Buffers need to be engineered per ring configuration or for worst case

Rate	Distance [km]	Delay [ms]	Bytes in Transit
OC 3	300x2	3	64 [kbyte]
OC48	300x2	3	1 [mbyte]
OC48	600x2	6	2-4 [mbyte]
OC192	2000x2	20	32 [mbyte]

- Small transit buffer's, amongst other things increase jitter for add traffic
- Congestion notification traffic needs timely delivery and competes with class-1 service queues

CSC & Best Effort

- **Class Selector Code (CSC) Points {1…7}**
  - Forms a small amount of traffic $\alpha 2$ (approximately < 5%)
  - Should be carried with priority compared to assured services, yet has no tight requirement for bounded delay
  - Jitter is not an issue
  - Expects no loss
  - "Fairness" = Carry with priority compared to Assured Service & best effort, with no loss
- **Best effort** should not be starved ($\alpha 1 + \alpha 2 + \alpha 3 < 1$)
  - May be tail or RED dropped, matter of vendor differentiation
  - No particular fairness issue, however RPR may decide to resolve the "station location advantage"

RPR Requirements Summary

	Vendor Diff	Standard Inter-working issue	
		within RPR	with other layers
Service Differentiation frame work		✓	✓
Station location advantage		✓	
Priority for TDM/LF & min bit rate for assured		✓	✓
Packet drop precedence Or no loss		✓	✓
Scheduler implementation	✓		
Organization of buffers/queues (including per flow)	✓		
Best effort tail drop vs. RED	✓		

Roadmap to DiffServ Compliance

- Support a 3 level priority scheme (classes 1, 2, 4) to be 75% DiffServ compliant
  - Delivers TDM L&F, CSC {1…7}, and Best Effort
  - Transit queues may support RED or tail drop as a matter of vendor differentiation
- Support a minimum rate for add traffic of assured service class, and WRED in transit and add queues
  - Delivers full compliance with DiffServ

Today

- Resolve the "station location advantage" using w-fair congestion avoidance to deal with hubbed traffic patterns, (WRED would no longer be needed in transit queues)

Late 2002