CS/CoE 536

Reconfigurable System On Chip Design

Lecture 8 : Class-based and Flow-based Queuing Systems

Washington University
Fall 2002

http://www.arl.wustl.edu/~lockwood/class/cs536/

John Lockwood

Lockwood@arl.wustl.edu

Outline

- Buffering of Data Enables
  - Quality of Service in Routers
  - Guards against Denial of Service in firewalls
- Types of Queuing
  - Class-based
    - Priority-based Service Disciplines
    - Differentiated Services
  - Per-Flow
    - Fluid Fair Queuing / Weighted Fair Queuing
    - Constant-Time Queue Service Disciplines
- Hardware Implementation of Flow Queues
  - Running time limitations in Gigabit networks

Quality of Service

- Gives some flows better service than others
  - Zero-sum game :
    - The amount of bandwidth on a link is fixed.
  - Design the system to maximize reward
    - Not all Internet users are created equal.
    - Will users pay for something that is not well defined?
  - Some flows get worse service!
- Applications
  - Determine which flows to drop under heavy load
  - Try to provide virtual link with end-to-end guarantee on bandwidth and delay

Fluid Fair Queue Model

- Fluid Traffic Model
  - Imagine a pipe that carries N flows, f_1 .. f_N
- C : Capacity of the pipe
  - Measured in (bits/second)
- ϕ_I : Service Share (weight) of flow i
  - With N flows : ϕ_1, ϕ_2, ϕ_3 .. ϕ_N
  - Can be equal or different
  - Unitless
- τ : A Short time interval
- $S_i(t,t+\tau)$: Amount of service given to f_i
  - During interval t to t+τ
  - Measure in (bits)
- B(t) : Backlogged channels at time T

$$S_i = C \cdot \tau \cdot \frac{\phi i}{\sum_{j \in B(t)} \phi j}$$

Fairness of Generalized Processor Sharing

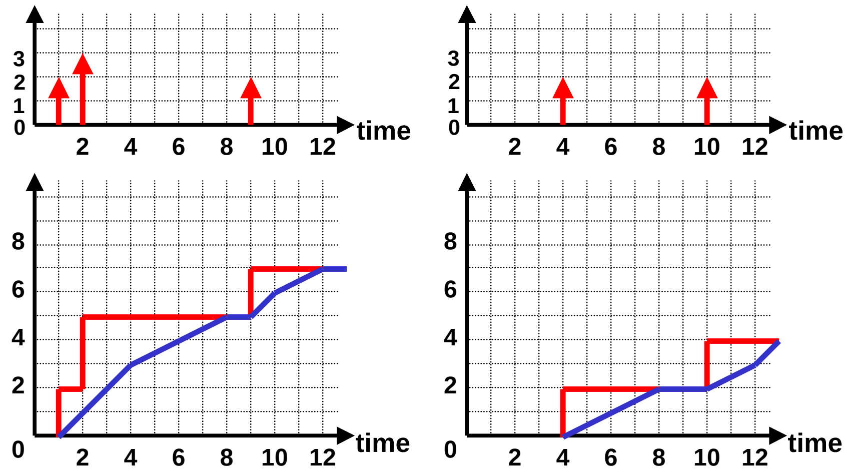
- For any backlogged flow, i, and another flow j (that may or may not be backlogged), the amount of service given to the flows will satisfy:

$$\frac{S_i(t,t+\tau)}{S_j(t,t+\tau)} \geq \frac{\phi_i}{\phi_j},\ j \in \{1..N\}$$

PGPS: Packetized GPS = WFQ

- Orders transmission time by F_p
  - F_p: Finish time of packet p
  - Approximates GPS, except that large packets with distant finish times will start and complete while short packets arrive.
- All backlogged queues receive service in proportion to share
  - Packetized GPS (PGPS)
  - Weighted Fair Queueing (WFQ)

Packetized GPS (PGPS)



Weighted Fair Queueing

- Parekh
  - MIT PhD Thesis, 1992
- v(t)=Virtual time
    - 0 for idle flows
  - B_j backlogged flows
  - τ time interval
  - ϕ weight

$$V(t_j + \tau) = V(t_j) + \frac{\tau}{\sum_{i \in B_{j-1}} \phi_i}$$

- F_i^k=Finish time
  - i_{th} packet
  - k_{th} queue
  - a_i^k arrival time
  - L_i^k length

$$F_i^k = \max\{F_i^{k-1}, V(a_i^k)\} + \frac{L_i^k}{\varphi_i}$$

Virtual Time (2) - Start & Finish Times

- S_i^k = Start time
- F_i^k = Finish time
- $S_i^k = \max \{ F_i^{k-1} , V(a_i^k)$
  – Later of last packet's completion time or this packets arrival time
- $F_i^k = S_i^k + L_i^k/\phi_I$
  – Arrival time plus length of packet divided by flow weight

Issues with Fair Queuing

- Core routers process a large number of flows
  – Currently hundreds of thousands of active flows
  – Soon Millions of flows
  – Billions of flows not far off considering population of the world and fraction of Internet usage.
- Finish Times require sorting
  – O(n lg(n)) Computational time
- Constant Time Solutions are preferred
  – O(1) Solutions scale for large N

Priority Queuing

- Maintain multiple queues of different priority
- Service highest priority queue before lower priority queues
- Upside
  – Inexpensive
- Downside
  – Starvation
  – How to classify packets as a class

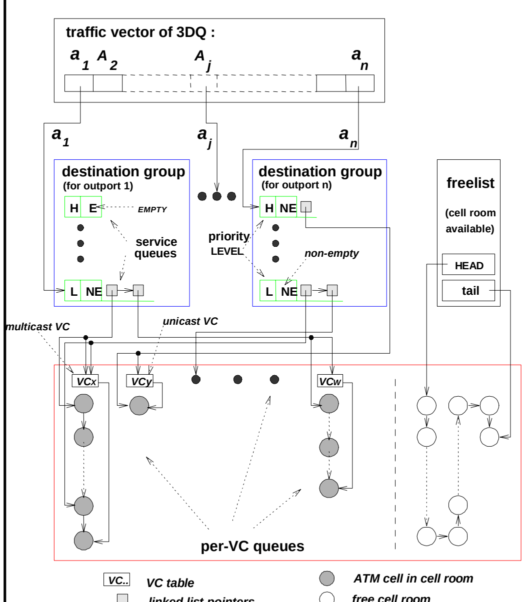
Round-Robin Queuing

- Maintain Separate Queue for each flow
- Service each flow, one at a time
  – Only service the backlogged flows
- Upside
  – Every flow gets a fraction of the bandwidth
  – Bursts of packets are distributed into even flow
- Issues
  – How to classify packets as flow

Deficit Round Robin (DRR)

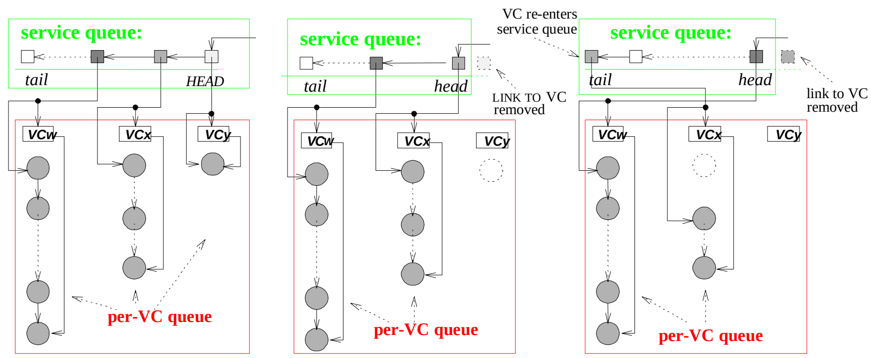
- DRR Transmits packets in constant time
  - Variation of Round-Robin through Flows
- f_i = Share to flow i = $Quantum_i$/Quantum
  - $Quantum_i$ = Number of allocated bits/round
  - Quantum = $Min_i(Quantum_i)$
  - Round=One iteration over backlogged queues
  - $Bytes_i(k)$ = # of bytes sent for flow i in round k
- Deficit Rule:
  - If ($bytes_i(1) < Quantum_i$) TX Packet
  - else DeficitCounter= $Quantum_i$ - $bytes_i(1)$
- Consider issues when weights are not equal.

3DQ Architecture



- Traffic groups
  - Can avoids HOL blocking
- Prioritized Service Queues
  - N Priority levels (4)
  - Highest priority first
- Per-flow Queues
  - Linked list
  - Multicast

Queue Re-entry Mechanism



(a) a service queue (b) an non-SQR case (c) a SQR case

– VC queue re-enters service queue when the VC queue is not empty

– SQR Statistically and Fairly Allocates Bandwidth to VCs That Belong to the Same Service Queue.

Additional References

– WFQ / PGPS
- A. K. Parekh, R. G. Gallager, *A Generalized Processor Sharing Approach to Flow Control in Integrated Services Networks: The Single-Node Case,* IEEE/ACM Transactions on Networking, number 3, volume 1, jun 1993, pages 344--357.

– Constant-Time Hardware Queuing Design
- H. Duan, J. W. Lockwood, S. M. Kang, J.D. Will, *A High-performance OC-12/OC-48 Queue Design Prototype for Input-buffered ATM Switches*, IEEE Infocom '97, Kobe, Japan, April 7-11, 1997, pp 20-28.