97-0831: GFR -- Providing Rate Guarantees with FIFO Buffers to TCP Traffic

Rohit Goyal, Raj Jain, Sonia Fahmy, Bobby Vandalore, Shivkumar Kalyanaraman

Raj Jain is now at Washington University in Saint Louis, jain@cse.wustl.edu http://www.cse.wustl.edu/~jain/

Sastri Kota, Lockheed Martin Telecommunications

Pradeep Samudra, Samsung Telecom America, Inc.

Contact: jain@cis.ohio-state.edu

http://www.cis.ohio-state.edu/~jain/

Overview

- Guaranteed frame rate
- Goals of this study
- Controlling TCP windows
- Differential Fair Buffer Allocation
- Simulation results

Guaranteed Frame Rate (GFR)

- GFR guarantees:
  - Low loss ratio to conforming frames
  - Best effort to all frames
  - Fair share of unused capacity
    (Not well defined. May be removed.)
- User specifies an MCR and a maximum frame size
- Conforming Frames = Frames which are untagged by the end system and pass the GCRA like policing mechanism.

Motivation

- GFR VCs could be used by routers separated by an ATM cloud.
- Users could also set up GFR VCs for traffic that could benefit from rate guarantees.
- Higher layers would expect some guarantees at that level.
- Higher layer traffic management may interact with GFR traffic management and achieve unfair throughput.
- A good GFR implementation should “work with” most common traffic types.

GFR Implementation Issues

- FIFO queuing versus per-VC queuing
  - Per-VC queuing is too expensive.
  - FIFO queuing should work by setting thresholds based on bandwidth allocations.
- Network tagging and end-system tagging
  - End system tagging can prioritize certain cells or cell streams.
  - Network tagging used for policing -- must be requested by the end system. [??]
- Buffer management policies
  - Per-VC accounting policies need to be studied

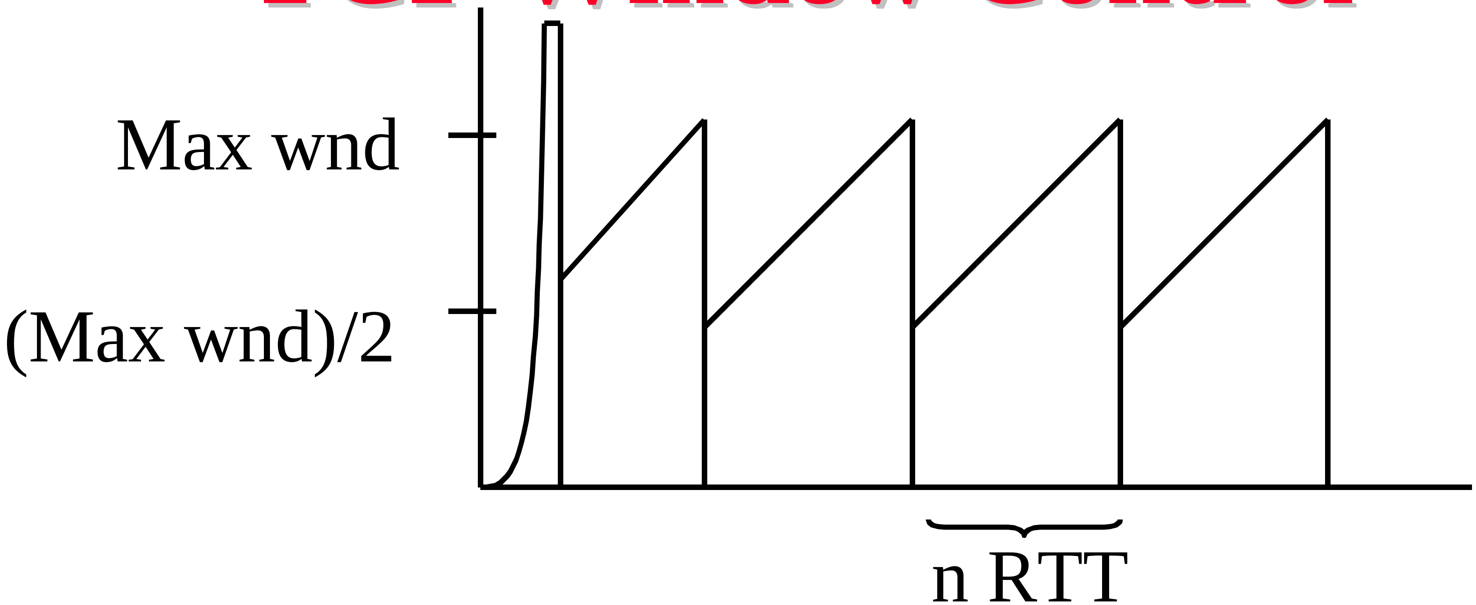
Summary of Past Results

- In the July meeting it was shown
  - Difficult to guarantee TCP throughput with FIFO queuing.
  - Can do so with per-VC queuing.
- All FIFO queuing cases were studied with high target network load, i.e., most of the network bandwidth was allocated as GFR.
- Need to study cases with lower percentage of network capacity allocated to GFR VCs.

Goals

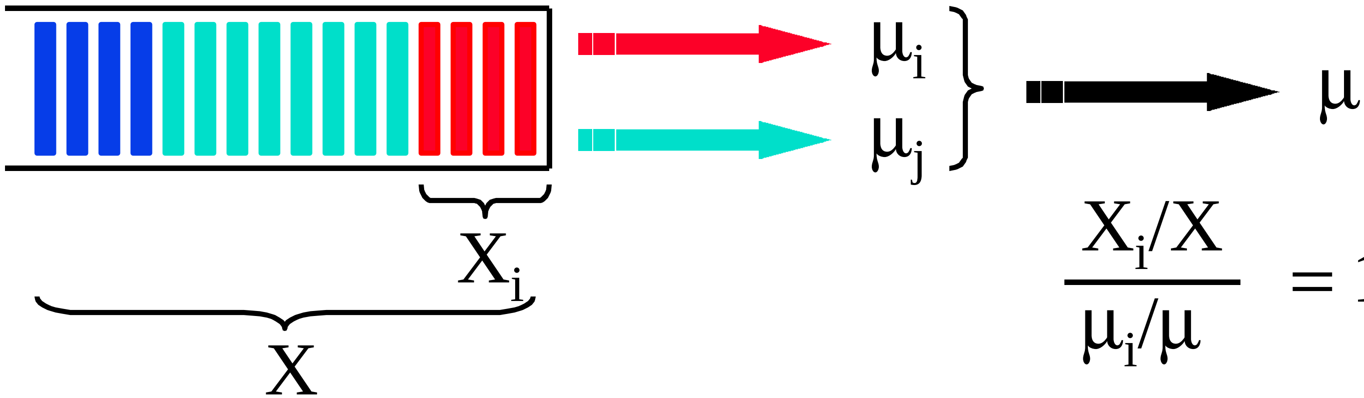
- Provide minimum rate guarantees with FIFO buffer for TCP/IP traffic.
- Guarantees in the form of TCP throughput.
- How much network capacity can be allocated before guarantees can no longer be met?
- Study rate allocations for VCs with aggregate TCP flows.

TCP Window Control



- For TCP window based flow control (in linear phase)
  - Throughput = (Avg wnd) / (Round trip time)
- With Selective Ack (SACK), window decreases by 1/2 during packet loss, and then increases linearly.
  - Avg wnd = $[\Sigma_{i=1,\ldots,n} (\text{max wnd}/2 + \text{mss}*i)] /n$

FIFO Buffer Management



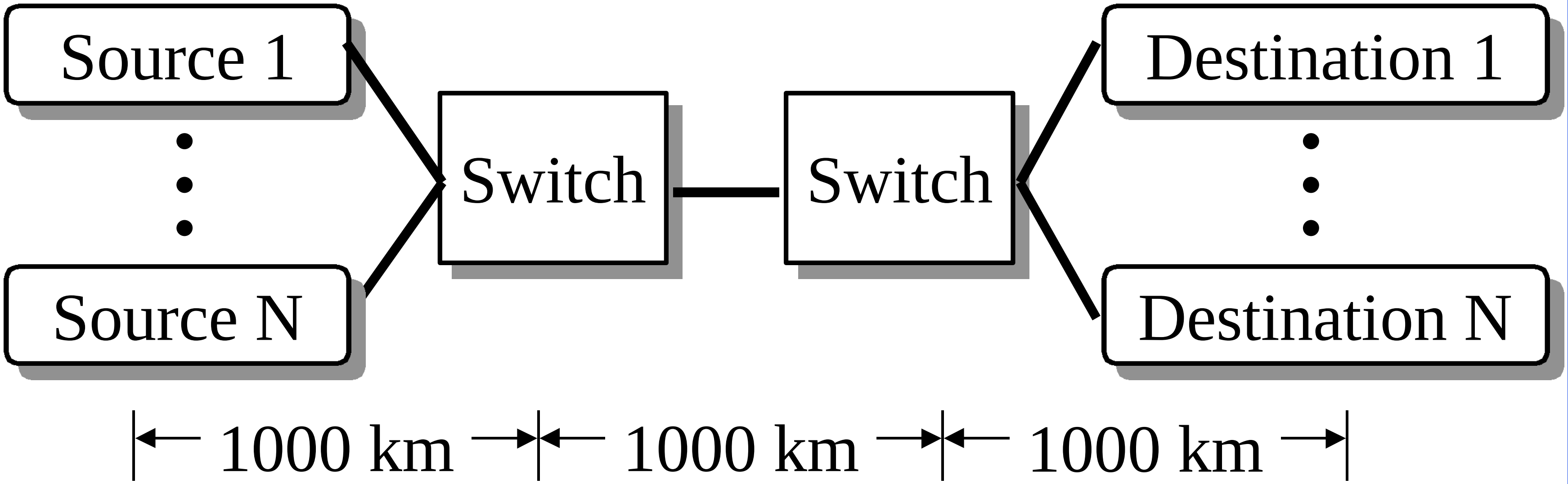
$$\frac{X_i/X}{\mu_i/\mu} = 1$$

- Fraction of buffer occupancy (X_i/X) determines the fraction of output rate (μ_i/μ) for VCi.
- Maintaining average per-VC buffer occupancy enables control of per-VC output rates.
- Set a threshold (R_i) for each VC.
- When X_i exceeds R_i, then control the VC's buffer occupancy.

Buffer Management for TCP

- TCP responds to packet loss by reducing CWND by one-half.
  - When *i*th flow's buffer occupancy exceeds R_i, drop a single packet.
  - Allow buffer occupancy to decrease below R_i, and then repeat above step if necessary.
- K = Total buffer capacity.
- Target utilization = $\Sigma R_i / K$.
- Guaranteed TCP throughput = Capacity $* R_i/K$
- Expected throughput, $\mu_i = \mu * R_i / \Sigma R_i$. $(\mu = \Sigma \mu_i)$

Simulation Configuration



- SACK TCP.
- 15 TCP sources (N = 15).
- Buffer Size = K = 48000 cells.
- 5 thresholds ($R_1,\ldots,R_5$).

Simulation Config (contd.)

Sources	Expt 1	Expt 2	Expt 3	Expt 4	Expected Throughput
1-3 (R_1)	305	458	611	764	2.8 Mbps
4-6 (R_2)	611	917	1223	1528	5.6 Mbps
7-9 (R_3)	917	1375	1834	2293	8.4 Mbps
10-24 (R_4)	1223	1834	2446	3057	11.2 Mbps
13-15 (R_5)	1528	2293	3057	3822	14.0 Mbps
$\Sigma R_i/K$	29%	43%	57%	71%	

- Threshold $R_{ij} \propto \lfloor K*MCR_i/PCR \rfloor$
- Total throughput μ = 126 Mbps. MSS =1024B.
- Expected throughput = $\mu * R_i / \Sigma R_i$

Simulation Results

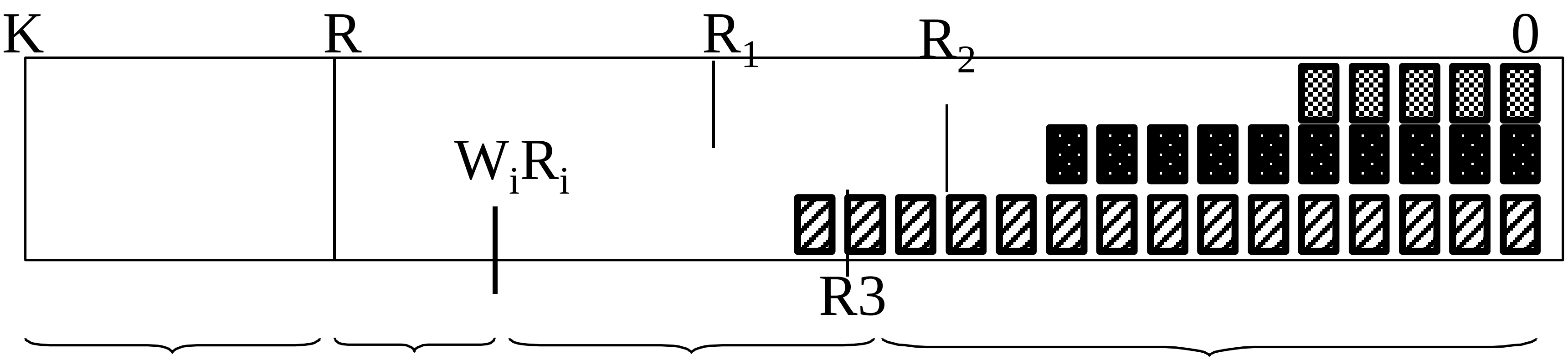
TCP Number	Throughput ratio (observed / expected)			
1-3	1.0	1.03	1.02	1.08
4-6	0.98	1.01	1.03	1.04
7-9	0.98	1.00	1.00	1.02
10-12	0.98	0.99	0.98	0.88
13-15	1.02	0.98	0.97	1.01

- All ratios close to 1.
  Variations increases with utilization.
- All sources experience similar queuing delays

TCP Window Control

- TCP throughput can be controlled by controlling window.
- FIFO buffer ⇒ Relative throughput per connection is proportional to fraction of buffer occupancy.
- Controlling TCP buffer occupancy
  ⇒ May control throughput.
- High buffer utilization ⇒Harder to control throughput.
- Formula does not hold for very low buffer utilization
  Very small TCP windows
  ⇒ SACK TCP times out if half the window is lost

Differential Fair Buffer Allocation



$X > R$ $\Rightarrow$ EPD	Drop All tagged	$X_i > R_i \Rightarrow$ Probabilistic Loss, $Xi > Z*R_i \Rightarrow$ EPD	$X_i \leq R_i$ $\Rightarrow$ No Loss

- W_i = Weight of VCi.
- R_i = per-VC threshold (R_i depends on W_i).
- X_i = per-VC buffer occupancy. ($X= \Sigma X_i$)
- $Z > 1$. $Z*R_i$ = per-VC high threshold.

Differential Fair Buffer Allocation

When first cell of frame arrives:

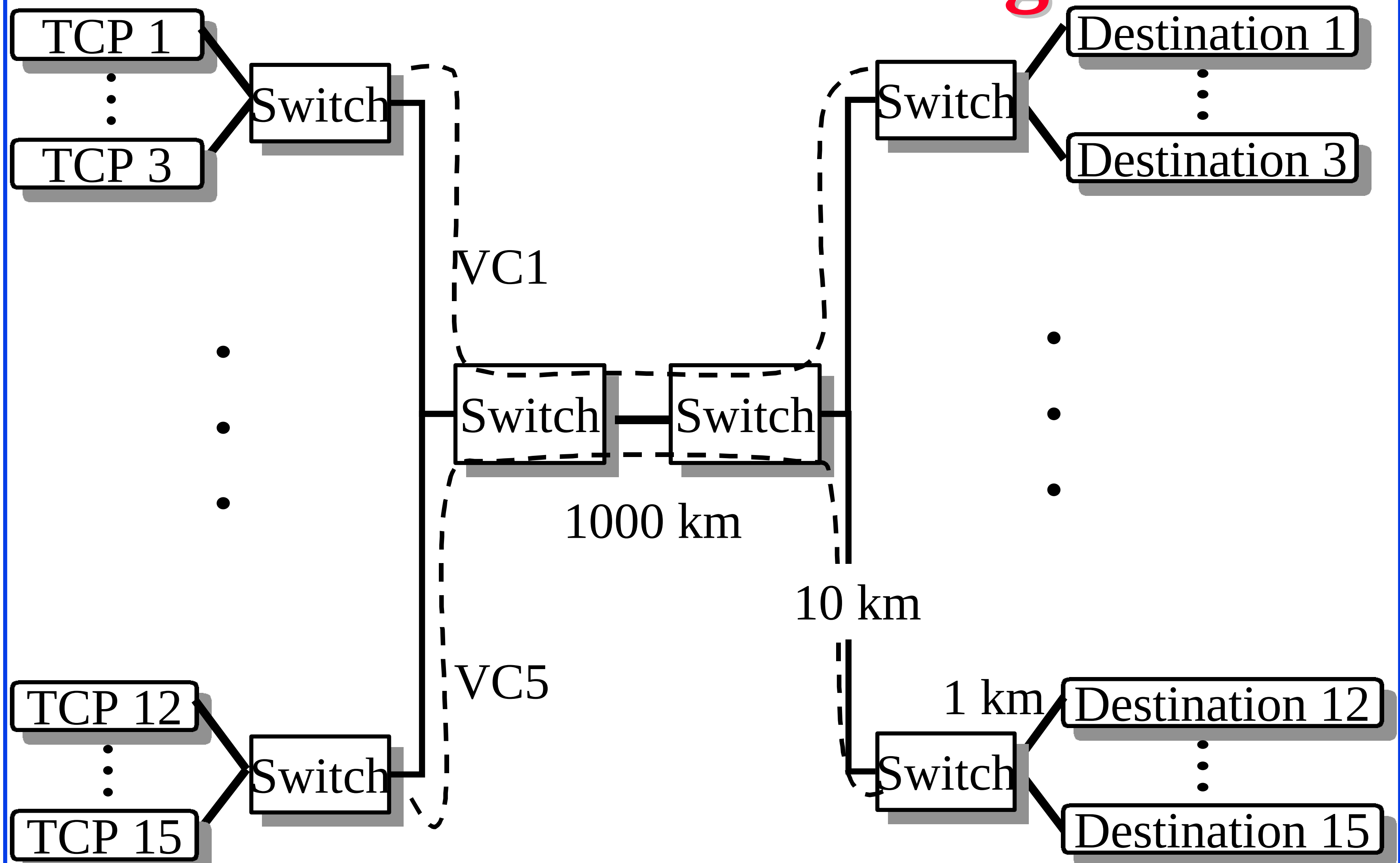
- IF ($X_i < R_i$) THEN
  - Accept frame
- ELSE IF (X > R) OR (Xi > Z*Ri) THEN
  - Drop frame
- ELSE IF (X < R) THEN
  - Drop cell and frame with

$$P\{drop\} = W_i * \frac{X_i - R_i}{R_i*(Z-1)}$$

Drop Probability

- Increases as X_i increases above R_i
  - Indicates higher levels of congestion.
- Proportional to W_i
  - With larger window, more packets can be dropped without timing out.
- $X_i > Z*R_i \Rightarrow$ EPD is performed.

DFBA Simulation Configuration



DFBA Simulation Configuration

- SACK TCP, 15 TCP sources.
- 5 VCs through backbone link. 3 TCP's per VC.
- Local switches merge TCP sources.

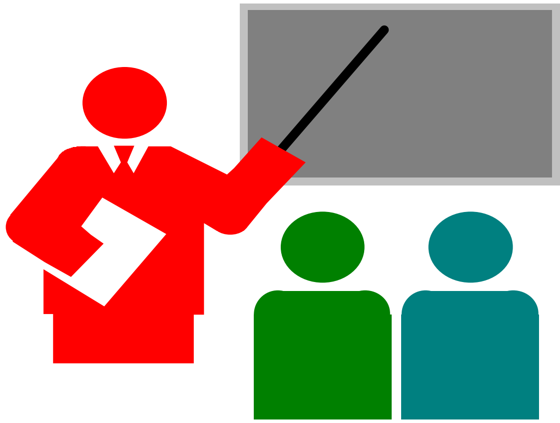
VC Number	Thresholds for backbone switch		
1	152	305	611
2	305	611	1223
3	458	917	1834
4	611	1223	2446
5	764	1528	3057

Simulation Results

VC Number	Throughput Ratios		
1	1.04	1.01	1.16
2	1.05	1.02	1.06
3	0.97	1.03	1.05
4	0.93	1.00	1.13
5	1.03	0.99	0.80

- Achieved throughput per-VC proportional to fraction of threshold allocated to the VC.
- Higher variation with increase in buffer allocation.

Summary



- SACK TCP throughput may be controlled with FIFO queuing under certain circumstances:
  - TCP, SACK (?)
  - Σ MCRs < Uncommitted bandwidth
  - Same RTT (?), Same frame size (?)
  - No other non-TCP or higher priority traffic (?)

Future Work

- Other TCP versions.
- Effect to non-adaptive (UDP) traffic
- Effect of RTT
- Effect of tagging
- Effect of frame sizes
- Parameter study
- Buffer threshold setting formula?
- How much buffer can be utilized?