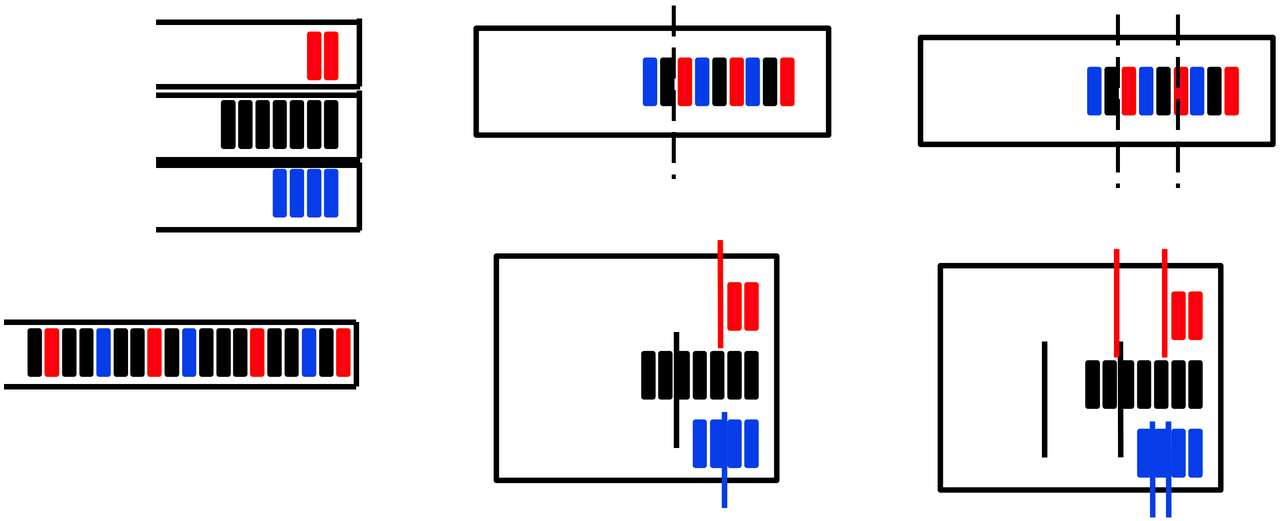
98-0405
Buffer Management for the GFR Service

Rohit Goyal, Raj Jain, Sonia Fahmy, Bobby Vandalore

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

- Buffer Management for GFR
- DFBA Description
- DFBA Simulation Results

GFR Options



Queuing	Per-VC	FIFO
Buffer Management	Per-VC Thresholds	Global Threshold
Tag-sensitive Buffer Mgmt	2 Thresholds	1 Threshold

Options (Cont)

- FIFO queuing versus per-VC queuing
  - Per-VC queuing is too expensive.
  - FIFO queuing should work by setting thresholds based on bandwidth allocations.
- Buffer management policies
  - Per-VC accounting policies need to be studied
- 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

- Accounting: Per-VC, Global
  Multiple or Single
- Threshold: Single or Multiple
- Four Types:
  - Single Accounting, Single threshold (SAST)
  - Single Accounting, Multiple threshold (SAMT)
  - Multiple Accounting, Single threshold (MAST)
  - Multiple Accounting, Multiple threshold (MAMT)

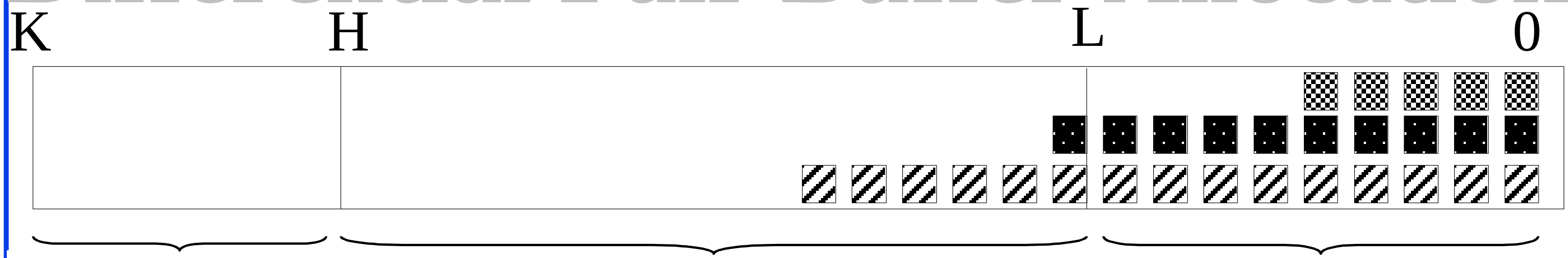
Buffer Mgmt Schemes

Group	Examples	Threshold	Drop Type	Tag Sensitive
SA ST	EPD, PPD	Static	Deterministic	No
	RED	Static	Probabilistic	No
MA ST	FRED Selective Drop,FBA	Dynamic	Probabilistic	No
	VQ+DEPD	Dynamic	Deterministic	No
MA MT	PME+ ERED	Static	Probabilistic	Yes
	DFBA	Dynamic	Probabilistic	Yes
SAMT	Priority Drop	Static	Deterministic	Yes

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



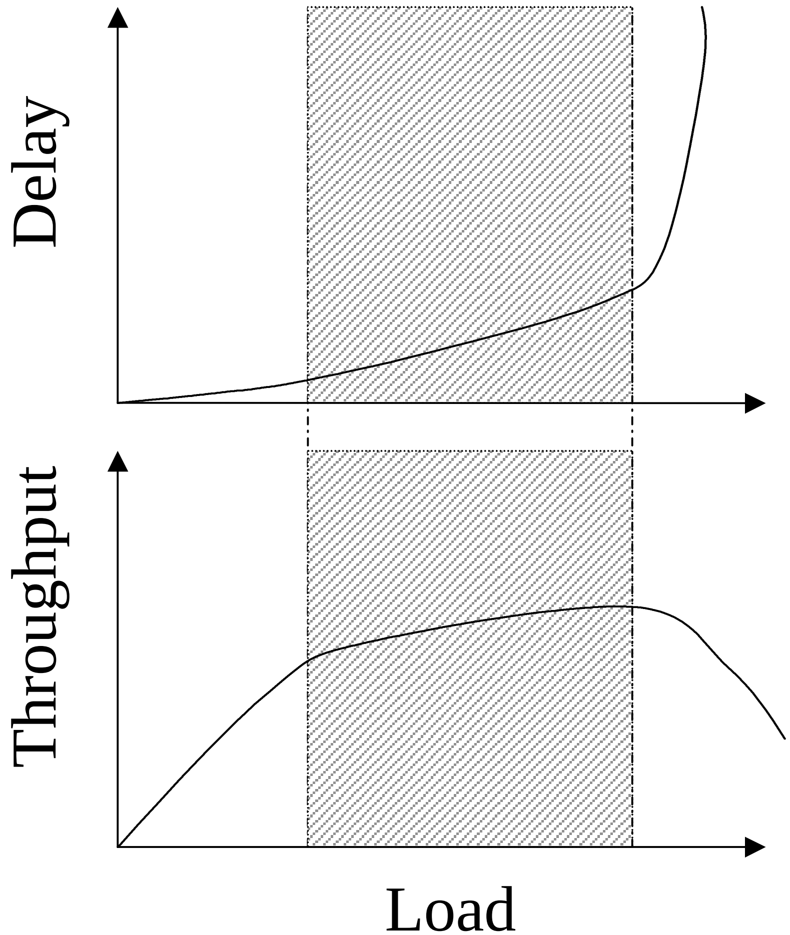
$X > H$
$\Rightarrow$ EPD

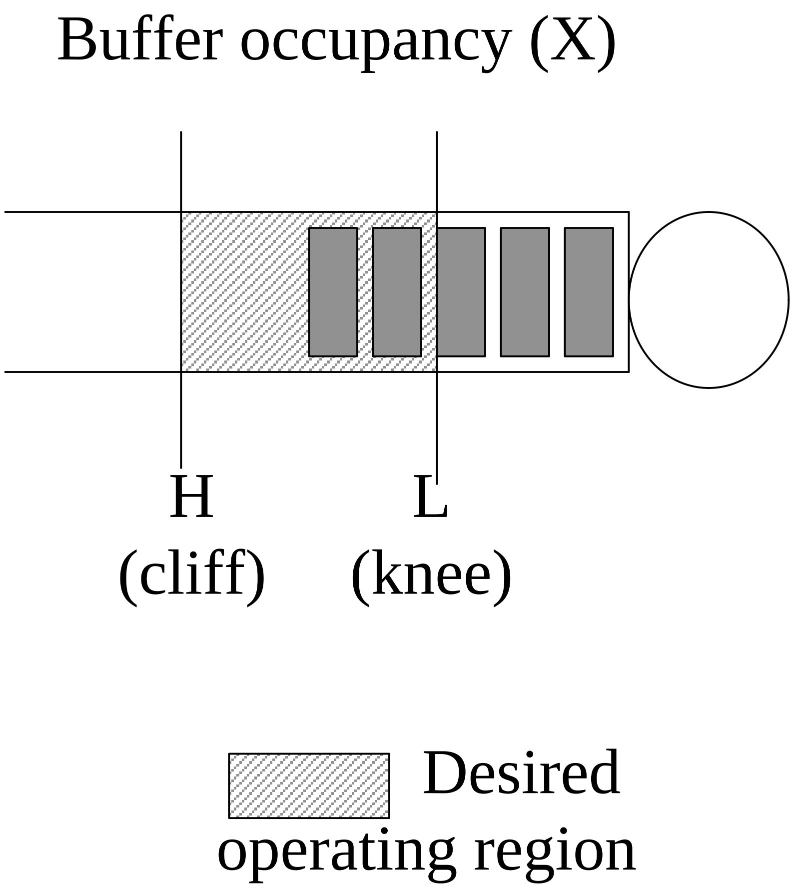
$X > L \Rightarrow$ Drop all CLP1.
$X > L$ and $X_i > X * W_i/W \Rightarrow$
Probabilistic Loss of CLP0

$X \leq L$
$\Rightarrow$ No Loss

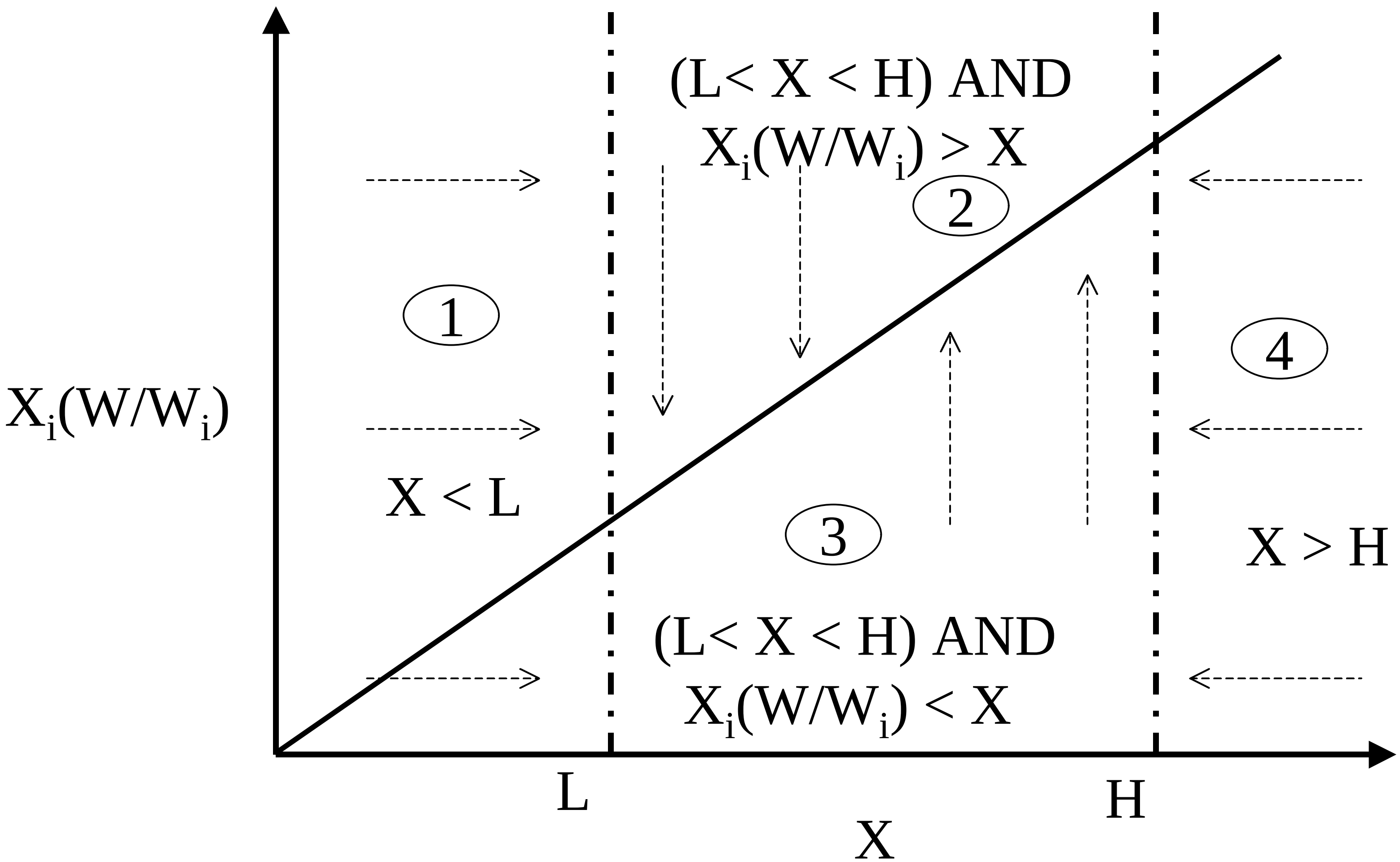
- W_i = Weight of VC_i = MCR_i/(GFR Capacity)
- $W = \Sigma W_i$
- L = Low Threshold. H = High Threshold
- X_i = Per-VC buffer occupancy. ($X = \Sigma X_i$)
- Z_i = Parameter ($0 \leq Z \leq 1$)

DFBA Operating Region





DFBA (contd.)



DFBA (contd.)

Region	Condition	Action
1	Underload	Improve efficiency
2	Mild congestion, more than fair share	Drop low priority packets, bring down to fair share
3	Mild congestion, less than fair share	Drop low priority packets, bring up to fair share
4	Severe congestion	Reduce load

DFBA Algorithm

- When first cell of frame arrives:
- IF $(X < L)$ THEN
  - Accept frame
- ELSE IF $(X > H)$ THEN
  - Drop frame
- ELSE IF ($(L < X < H)$ AND $(X_i \leq X \times W_i/W)$)
  - Drop CLP1 frame
- ELSE IF ($(L < X < H)$ AND $(X_i > X \times W_i/W)$)
  - Drop CLP1 frame
  - Drop CLP0 frame with

$$P\{Drop\} = Z_i \left(\alpha \times \frac{X_i - X \times W_i/W}{X(1 - W_i/W)} + (1-\alpha) \times \frac{X - L}{H - L} \right)$$

Drop Probability

- Fairness Component
  (VC_i's fair share = $X \times W_i/W$)

$$\frac{X_i - X \times W_i/W}{X \times (1 - W_i/W)}$$

  Increases linearly as X_i increases from $X \times W_i/W$ to X

- Efficiency Component

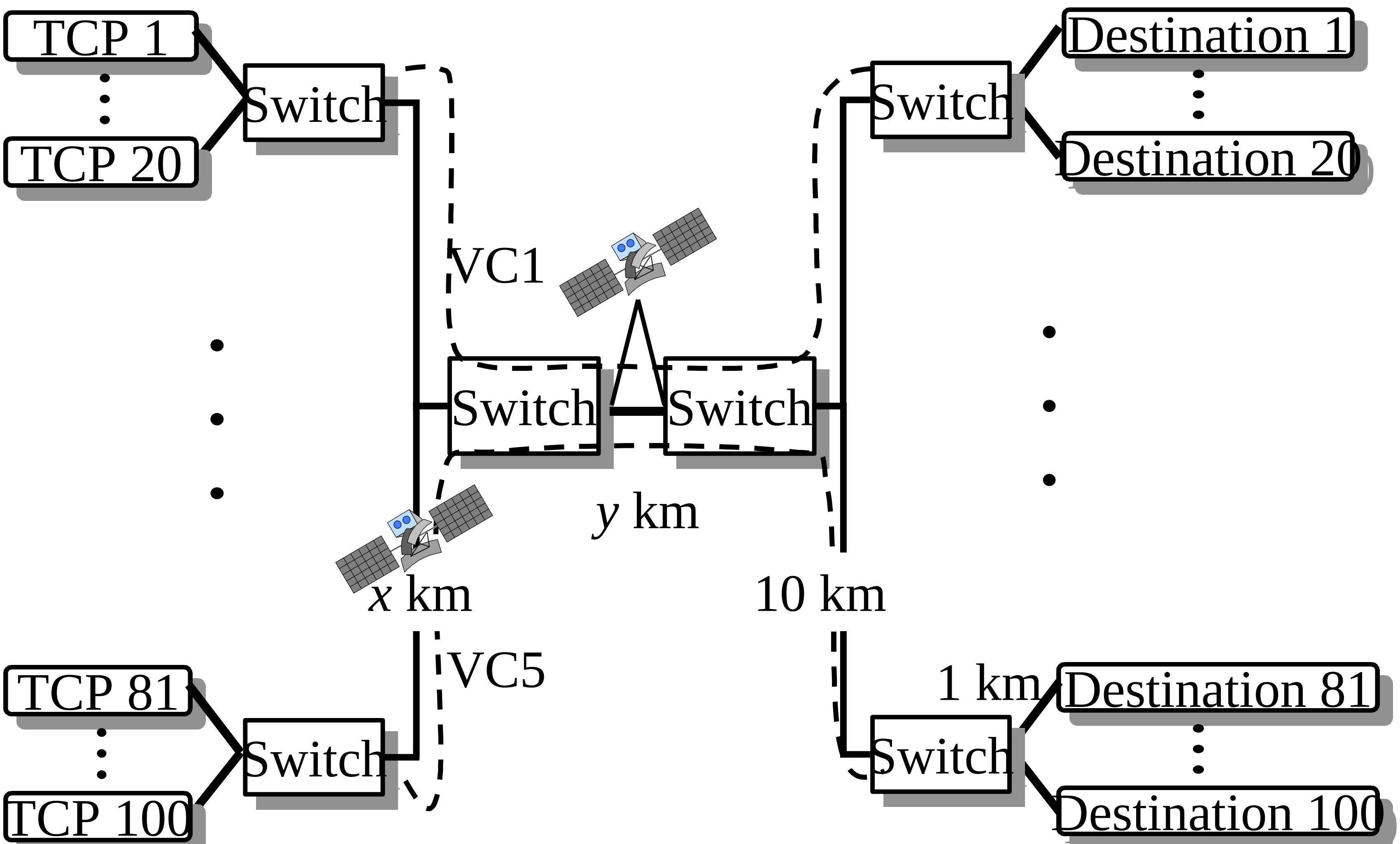
$$\frac{X - L}{H - L}$$

  Increases linearly as X increases from L to H

Drop Probability (contd.)

- Z_i allows scaling of total probability function
  - Higher drop probability results in lower TCP windows
  - TCP window size $W \propto 1/\sqrt{P\{Drop\}}$ for random packet loss [Mathis]
    TCP data rate $D \propto \frac{MSS}{RTT \times \sqrt{P(drop)}}$
  - To maintain high TCP data rate for large RTT:
    - Small P(Drop)
    - Large MSS
- Choose small Z_i for satellite VCs.
- Choose small Z_i for VCs with larger MCRs.

DFBA Simulation Configuration



DFBA Simulation Configuration

- SACK TCP, 50 and 100 TCP sources
- 5 VCs through backbone link.
- Local switches merge TCP sources.
- x = Access hop = 50 μs (Campus), or 250 ms GEO
- y = Backbone hop = 5 ms (WAN or LEO) or 250 ms (GEO)
- GFR capacity = 353.207 kcells/sec (≈155.52 Mbps)
- $\alpha = 0.5$

Simulation Configuration (contd)

- 50 TCPs with 5 VCs (50% MCR allocation)
  - MCR_i = 12, 24, 36, 48, 60 kcells/sec, i=1, 2, 3, 4, 5
  - W_i = 0.034, 0.068, 0.102, 0.136, 0.170
  - Σ (MCR_i /GFR capacity) = $\Sigma W_i = W \approx 0.5$

Simulation Configuration (contd)

- 50 and 100 TCPs with 5 VCs (85% MCR allocation)
  - MCR_i = 20, 40, 60, 80, 100 kcells/sec, $i=1, 2, 3, 4, 5$
  - W_i = 0.0566, 0.1132, 0.1698, 0.2264, 0.283
  - $\Sigma (MCR_i / \text{GFR capacity}) = \Sigma W_i = W \approx 0.85$

Simulation Results

MCR	Achieved Throughput	Excess	Excess / MCR
4.61	11.86	7.25	1.57
9.22	18.63	9.42	1.02
13.82	24.80	10.98	0.79
18.43	32.99	14.56	0.79
23.04	38.60	15.56	0.68
69.12	126.88	57.77	

- 50 TCPs with 5VCs (50% MCR allocation)
- Switch buffer size = 25 kcells
- Z_i=1, for all i
- MCR guaranteed. Lower MCRs get higher excess.

Effect of MCR Allocation

MCR	Achieved Throughput	Excess	Excess/MCR
7.68	12.52	4.84	0.63
15.36	18.29	2.93	0.19
23.04	25.57	2.53	0.11
30.72	31.78	1.06	0.03
38.40	38.72	0.32	0.01
115.2	126.88	11.68	

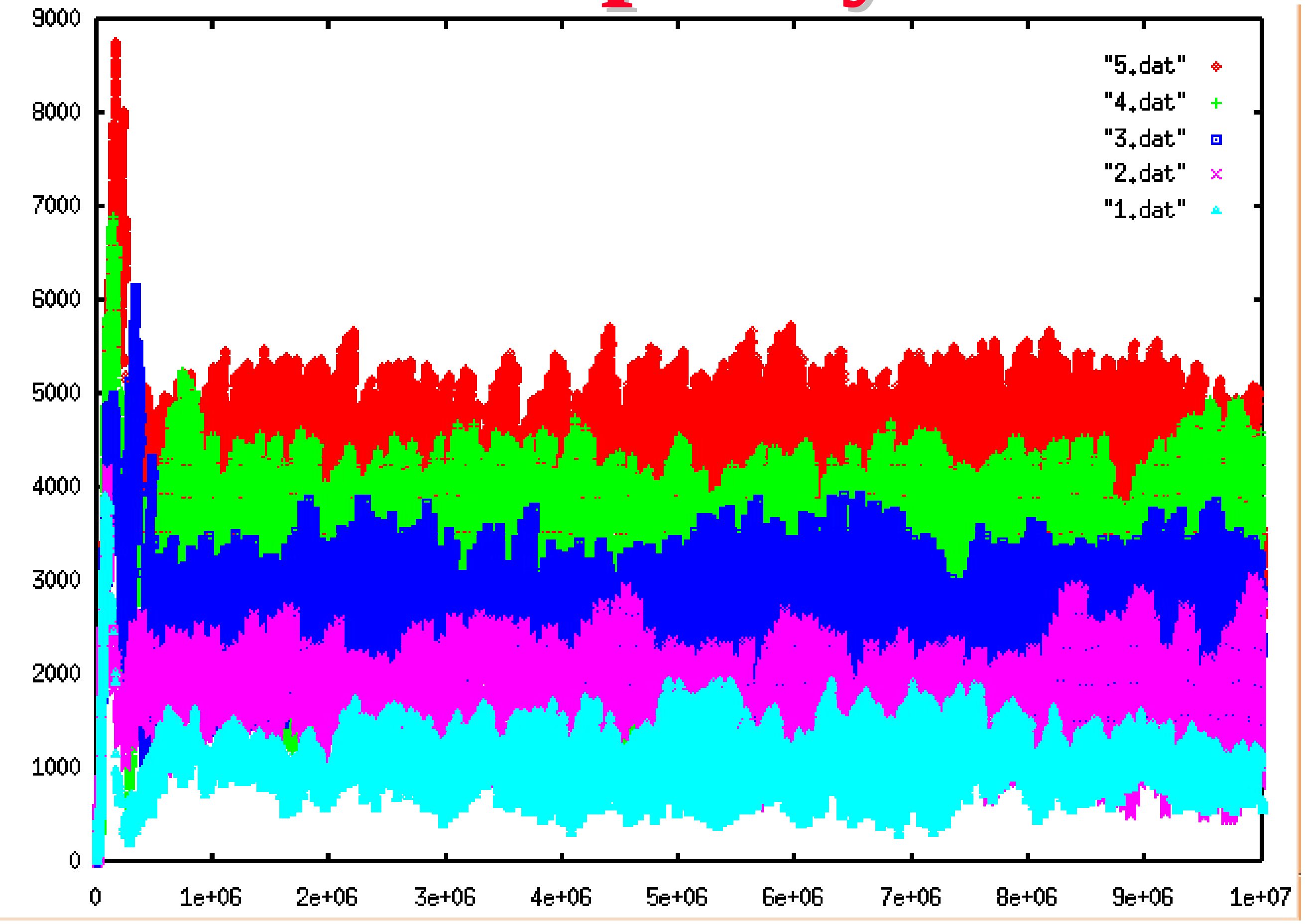
- 50 TCPs with 5 VCs (85% MCR allocation)
- Switch buffer size = 25 kcells
- Z_i=1, for all I
- MCR guaranteed. Lower MCRs get higher excess

Effect of Number of TCPs

MCR	Achieved Throughput	Excess	Excess/MCR
7.68	11.29	3.61	0.47
15.36	18.19	2.83	0.18
23.04	26.00	2.96	0.13
30.72	32.35	1.63	0.05
38.40	39.09	0.69	0.02
115.2	126.92	11.72	

- 100 TCPs with 5 VCs (85 % MCR allocation)
- Switch buffer size = 25 kcells
- Z_i=1, for all i
- Independent of the number of sources

Buffer Occupancy



- 100 TCPs with 5 VCs (85 % MCR allocation)
- Switch buffer size = 25 kcells

Effect of Buffer Size

MCR	Achieved Throughput	Excess	Excess/MCR
7.68	11.79	4.11	0.54
15.36	18.55	3.19	0.21
23.04	25.13	2.09	0.09
30.72	32.23	1.51	0.05
38.40	38.97	0.57	0.01
115.2	126.67	11.47	

- 100 TCPs with 5 VCs (85 % MCR allocation)
- Switch buffer size = 6 kcells
- Z_i=1, for all I
- MCR guarantees for small buffer size

Effect of Buffer Size

MCR	Achieved Throughput	Excess	Excess/MCR
7.68	10.02	2.34	0.30
15.36	19.31	3.95	0.26
23.04	25.78	2.74	0.12
30.72	32.96	2.24	0.07
38.40	38.56	0.16	0.00
115.2	126.63	11.43	

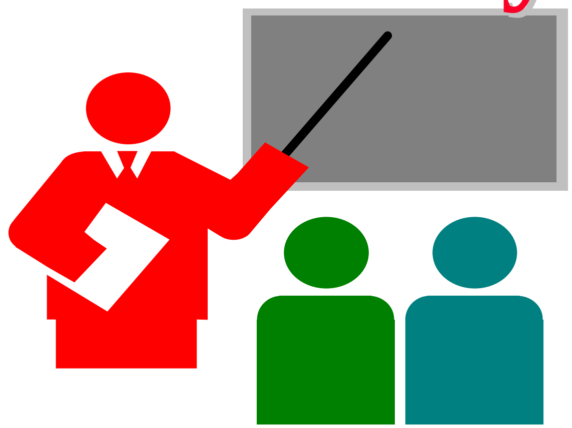
- 100 TCPs with 5 VCs (85 % MCR allocation)
- Switch buffer size = 3 kcells
- Z_i=1, for all I
- MCR guarantees for small buffer size

Effect of Z_i

$Z_i = 1-W_i/W$		$Z_i = (1-W_i/W)^2$	
Excess	Excess/MCR	Excess	Excess/MCR
3.84	0.50	0.53	0.07
2.90	0.19	2.97	0.19
2.27	0.10	2.77	0.12
2.56	0.08	2.39	0.08
0.02	0.02	3.14	0.08

- 100 TCPs with 5 VCs (85 % MCR allocation)
- Switch buffer size = 6 kcells
- Small Z_i for large MCR enables MCR proportional sharing of excess capacity

Summary



- Buffer Management Policy: DFBA for GFR
  - Allocates MCR proportional buffers.
  - Guarantees throughput and provides fairness
- Survey and classification of buffer management schemes.