98-0151: A Definition of Generalized Fairness and its Support in Switch Algorithms

Bobby Vandalore, Sonia Fahmy, Raj Jain, Rohit Goyal and Mukul Goyal

Depa[illegible] y

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

Overview

- General Fairness: Definition
- Relationship to Pricing/Charging Policies
- Achieving General Fairness
- Example modification to a Switch Algorithm
- Simulation: Configuration and Parameters
- Simulation: Results
- Conclusion

Notation

- Define following [Notation from TM4.0]:
  - A = Total available bandwidth
  - U = Sum of bandwidth of underloaded connections
  - B = A - U, excess bandwidth
  - N_a = Number of active connections
  - N_u = Number of active connections bottlenecked elsewhere
  - n = N_a - N_u,
    number of active connections bottlenecked on this link

Notation (Cont)

- M = Sum of MCRs of active connections
- B(i) = Generalized Fair allocation for connection i
- MCR(i) = MCR of connection i
- w(i) = pre-assigned weight associated with i

TM4.0 Definitions

1. B(i)=B/n
2. B(i)=MCR(i)+(B-M)/n
3. B(i) = Max{MCR(i), Max-Min Share}
4. B(i) = B*(MCR(i)/M)
5. B(i) = w(i)*B/Sum(w(j))

- Definition 5 does not always guarantee MCR
- Definition 3 may result in total of fairshare being more than the capacity

General Definition

- FairShare

$$B(i) = MCR(i) + \frac{w(i)\,(B - M)}{\Sigma_{j=1,n}\, w(j)}$$

- This definition is a superset of 1, 2, 4 in TM4.0
- Always ensures MCR

Mapping to TM 4.0

- $w(i) = w$, $MCR(i)=0$: $B(i) = B/n$
  This is Definition 1 (Max-min Fair).
- $w(i) = w$: $B(i) = MCR(i) + (B - M)/n$
  This is Definition 2 (MCR plus equal share)
- $w(i) = MCR(i)$:
  $B(i) = MCR(i) + (B-M)\, MCR(i) / M$
  $= B^* (MCR(i)/M)$
  This is Definition 4 (Proportional to MCR)

Pricing Function

- T = Small time interval, W = Number of bits
  R = Average rate W/T
- Cost $C = f(W,R)$. If C is restricted to continuous differentiable functions: $C = \Sigma_{ij}\, a_{ij}\, W^i R^j$
- For <u>all</u> values of W and R:
  - $C \geq 0 \quad \partial C/\partial W \geq 0 \quad \partial C/\partial R \geq 0$
  - $\partial(C/W)/\partial W \leq 0$ [Economy of Scale]
  - $\partial(C/R)/\partial R \leq 0$ [Economy of Scale]
- The <u>only</u> function that satisfies all 5 conditions is:

$$C = a_{00} + a_{10}W + a_{01}R + a_{11}\, WR$$

A Simple Pricing Fn

- f() is non-decreasing w.r.t to W
  f() is non-increasing w.r.t to T $\Rightarrow$ non-decreasing w R
- A simple function satisfying these requirements is:
  $C = c + w\, W + r\, R$
  Here, c = Fixed cost per connection
  w = Cost per bit (How much)
  r = Cost per Mbps (How fast)

Pricing With MCR

- Let L = MCR
- Cost C = c + w W + r (R-L) + m L
  Here, m = dollars per Mbps of MCR
  r = dollars per Mbps of extra bandwidth.
- Consider two users with MCRs L_1, L_2. Rates R_1, R_2 and bits transmitted W_1, W_2 (assume $W_1 \geq W_2$)
$$C_1 = c + w\,W_1 + r\,(R_1 - L_1) + m\,L_1$$
$$C_2 = c + w\,W_2 + r\,(R_2 - L_2) + m\,L_2$$
- Economy of Scale: C/W is a decreasing function of W
$$C_1/W_1 \leq C_2/W_2$$

Pricing (cont.)

- $c/W_1 + w + r\,(R_1 - L_1)/W_1 + mL_1/W_1 \leq$
  $c\,/W_2+w+r(R_2-L_2)/W_2+mL_2/W_2$
- Using $R_i = W_i/T$
- $c/(R_1T) + w + r(R_1-L_1)/(R_1T)+ mL_1/(R_1T) \leq$
  $c\,/(R_2T)+w+ r(R_2-L_2)/(R_2T)+mL_2/(R_2T)$
  $c/R_1 -rL_1/R_1+mL_1/R_1 \leq c/R_2 -r\,L_2/R_2+mL_2/R_2$
  $(c+(m-r)L_1)/(c+(m-r)L_2) \leq R_1/R_2$
  $(R_1-L_1)/(R_2-L_2) \geq (a+L_1)/(a+L_2)$
- Here, $a = c/(m-r)$
  $\Rightarrow$ Weight should be a linear function of MCR.
  This is the policy used in this contribution.

Achieving Gen. Fairness

- $B(i) = MCR(i) + w(i)\,(B - M)/\Sigma_{j=1,n}\, w(j)$
- Switch allocates MCR and a weighted share of the excess bandwidth
- $ACR(i) = MCR(i) + ExcessFairshare(i)$
- $ExcessFairshare(i) = w(i)\,(B\text{-}M)/\Sigma_{j=1,n}\, w(j)$
- ACR(i) - MCR(i) should converge to ExcessFairshare(i)

Activity Level

- The allocation should also consider activity level of a source.
  There is no point in giving extra bandwidth to sources not using it.
- Activity level AL(i)
  = min{1, (SrcRate(i)-MCR(i))/ExcessFairshare(i)}
- ExcessFairshare(i) = $w(i)AL(i)(B-M)/\Sigma_{j=1,n} w(j)\ AL(j)$
- Recursive definition. Converges in just a few iterations.

ERICA+

End of Averaging Interval:

- Total ABR Capacity= Link Capacity - VBR Capacity
- Target ABR Capacity = F(Q) x Total ABR Capacity
  F(Q) is a function of queue length.
  1-F(Q) of the capacity is used to drain the queues
- Overload z = ABR Input Rate/(Target ABR Capacity)
- Effective # of active sources = $\Sigma_{j=1,n}$ AL(j)
- Fairshare
  = Target ABR Capacity /Eff. # of active sources

ERICA+ (cont.)

When a BRM is received:

- FairShare(i) = AL(i)Fairshare*
- For Efficiency: VCShare(i) = (SrcRate(i))/z*
- ER(i) = max (FairShare(i), VCShare(i))*
- $ER_{in\_RM\_Cell}$
  $= \min\{ER_{in\_RM\_Cell}, ER(i), TargetABRCap\}$
- Near steady state the VCShare(i) term converges to Fairshare(i), achieving max-min fairness and efficiency.

*Done only on first BRM

Modified ERICA+

End of Averaging Interval:

- Total ABR Cap= Link Cap - VBR Cap
  - $\Sigma_{j=1,n}$ min{SrcRate(i), MCR(i)}
- Target ABR Cap = F(Q) x Total ABR Cap
- Input Rate
  = ABR Input Rate - $\Sigma_{j=1,n}$ min{SrcRate(i), MCR(i)}
- Overload z = Input Rate/(Target ABR Capacity)
- Effective weight of active sources = $\Sigma_{j=1,n}$ w(j)AL(j)
- ExcessFairshare
  = Target ABR Cap /Eff. weight of active sources

Mod. ERICA+ (cont.)

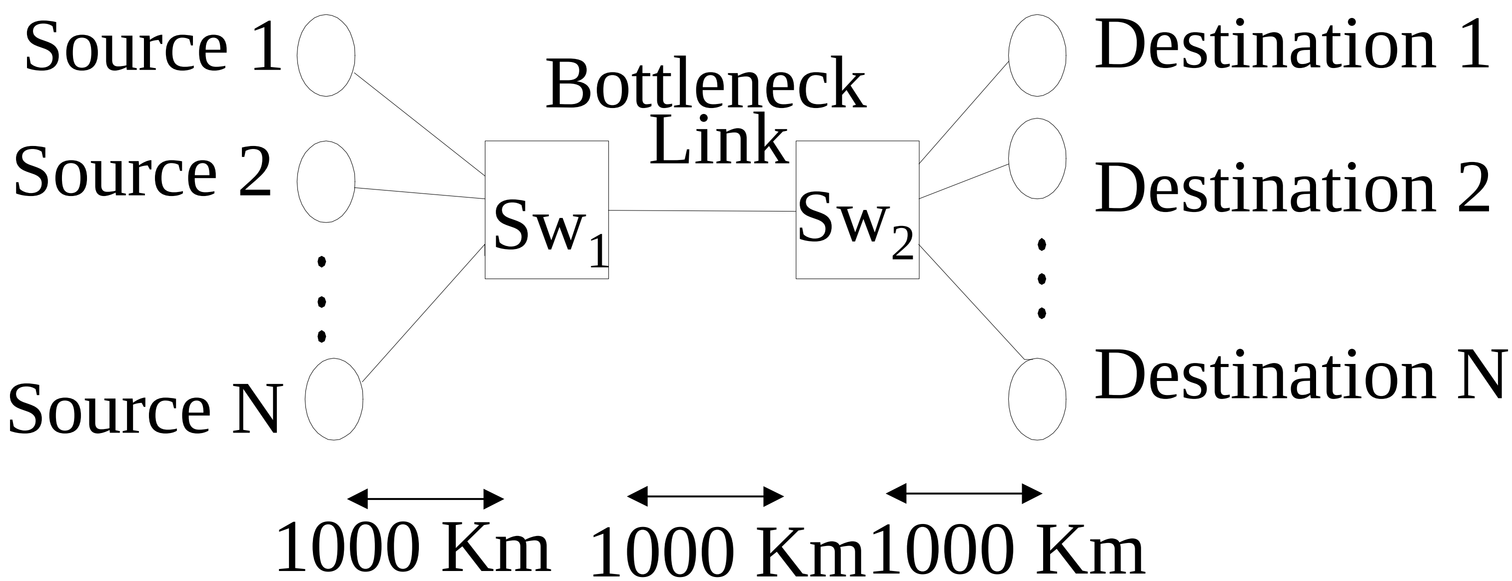
When a BRM is received:

- ExcessFairShare(i) = w(i)AL(i)ExcessFairshare
- For Efficiency: VCShare(i) = (SrcRate(i) - MCR(i))/z
- ER(i)
  = MCR(i) + max {ExcessFairshare(i), VCShare(i)}
- $ER_{in\_RM\_Cell}$
  = min{$ER_{in\_RM\_Cell}$, ER(i), TargetABRCap}
- Near steady state the VCShare(i) term converges to ExcessFairshare(i), achieving generalized fairness and efficiency.

Configuration 1

Simple configuration

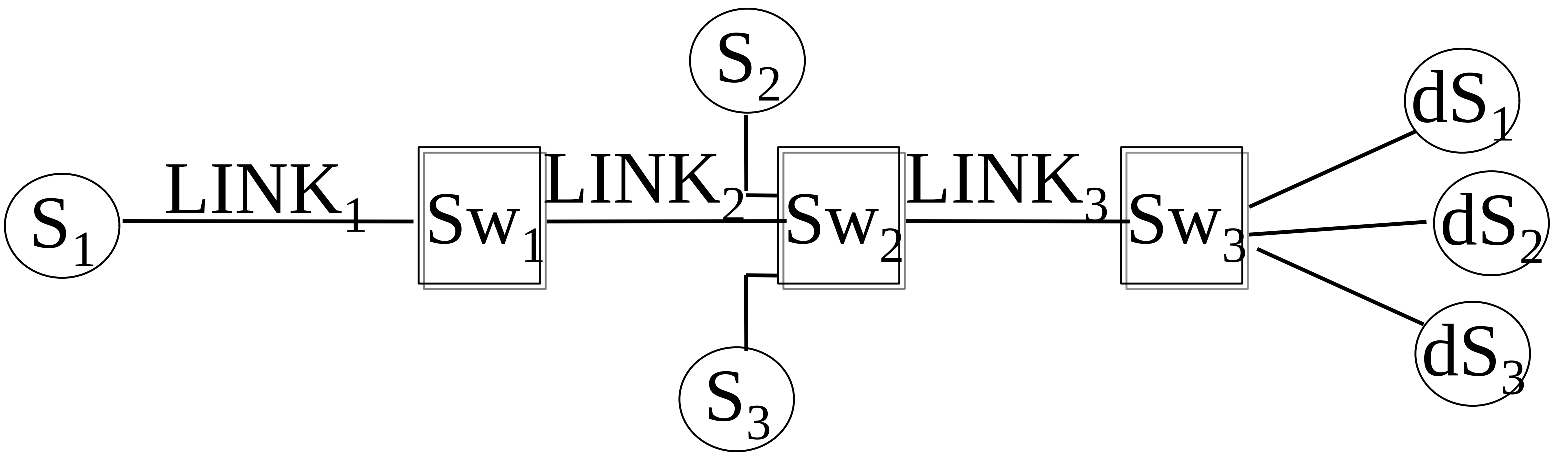
- N infinite ABR source,
  N ABR destinations (N = 3 in simulations)
- One way traffic. From sources to destination



Configuration 2

Source Bottleneck configuration

- Source S1 is bottlenecked at 10 Mbps
  (i.e., it always sends data at a rate of upto 10 Mbps, irrespective of its ACR)



Configuration 3

Generic Fairness Configuration (GFC-2)

- D - distance of links = 1000 Km
- All links between switches are bottleneck links

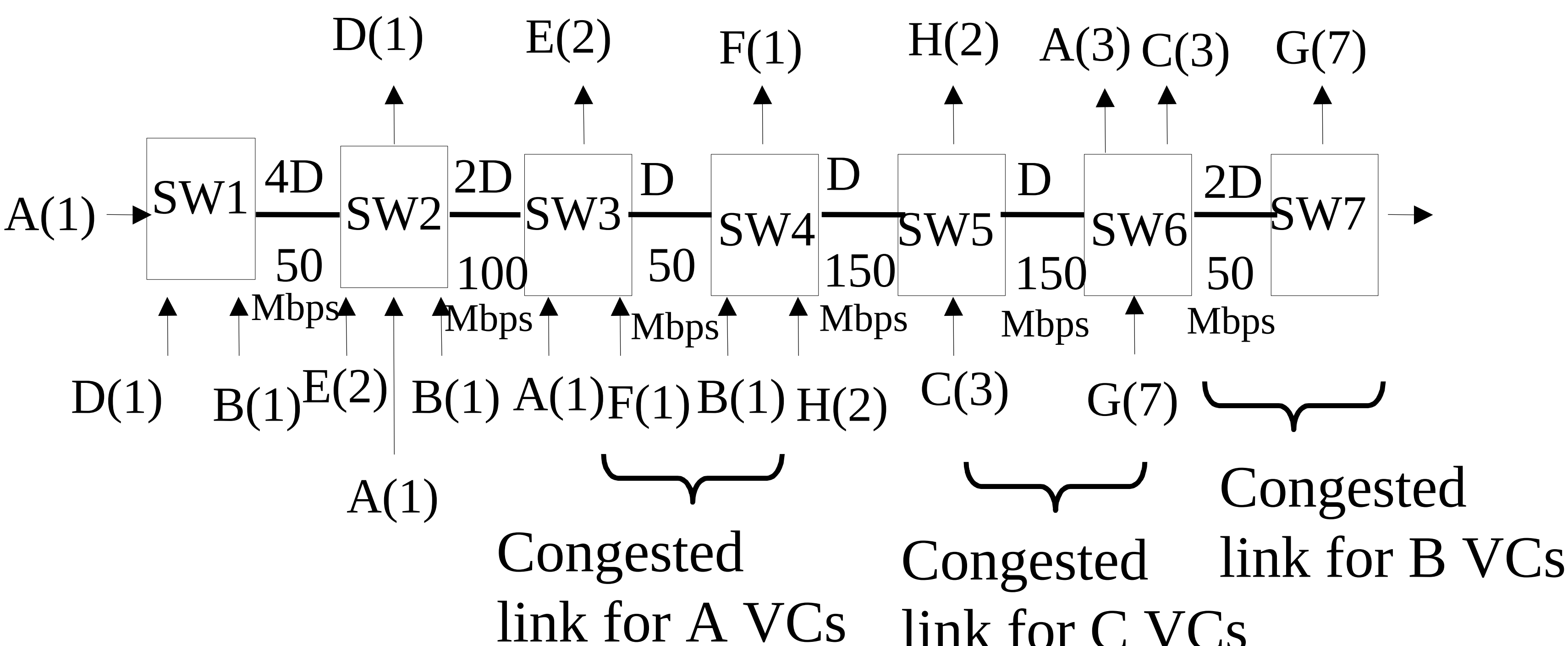


Table 1: Simulation Parameters

Configuration Name	Link Distance	Averaging Interval	Target Delay
Three Sources	1000 Km	5 ms	1.5 ms
Source Bottleneck	1000 Km	5 ms	1.5 ms
GFC-2	1000 Km	15 ms	1.5 ms

Table 1: 3-Src Results

Case Number	Src Num	MCR	a	Weight Function	Expected Fair Share	Actual Share
1	1	0	∞	1	49.92	49.92
	2	0	∞	1	49.92	49.92
	3	0	∞	1	49.92	49.92
2	1	10	∞	1	29.92	29.92
	2	30	∞	1	49.92	49.92
	3	50	∞	1	69.92	69.92
3	1	10	5	15	18.53	16.64
	2	30	5	35	49.92	49.92
	3	50	5	55	81.30	81.30

For all 3 cases, the algorithm achieves desired allocation

3-Src ACRs

- Case 1: a = ∞, MCRs = 0. All weights are equal. Allocation is 149.76/3

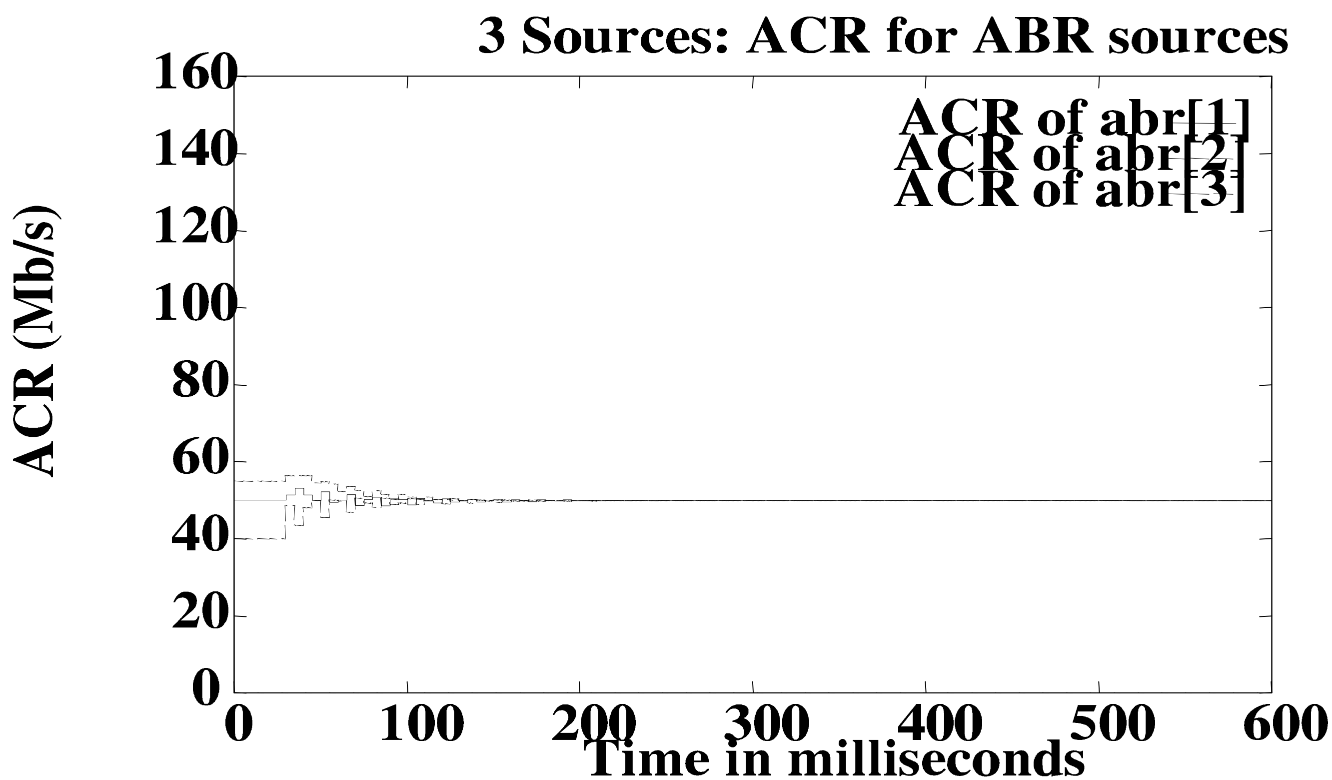


Table 3: 3-Src Transient

Case Num.	Src Num	MCR	a	weight func.	Expected Frshare (non-trans.)	Actual (non-trans) share	Expted Frshare (trans.)	Actual (trans.) share
1	1	0	∞	1	74.88	74.83	49.92	49.92
	2	0	∞	1	-	-	49.92	49.92
	3	0	∞	1	74.88	74.83	49.92	49.92
2	1	10	∞	1	54.88	54.88	29.92	29.83
	2	30	∞	1	-	-	49.92	49.92
	3	50	∞	1	94.88	95.81	69.92	70.93
3	1	10	5	15	29.92	29.23	18.53	18.53
	2	30	5	35	-	-	49.92	49.92
	3	50	5	55	119.84	120.71	81.30	81.94

- Source 2 (transient) is active only between 400-800 ms. Expected allocation achieved.

3-Src Transient ACRs

- Case 2: a = ∞, MCRs ≠ 0. All weights are equal. Allocation is (29.92,39.92,69.92)

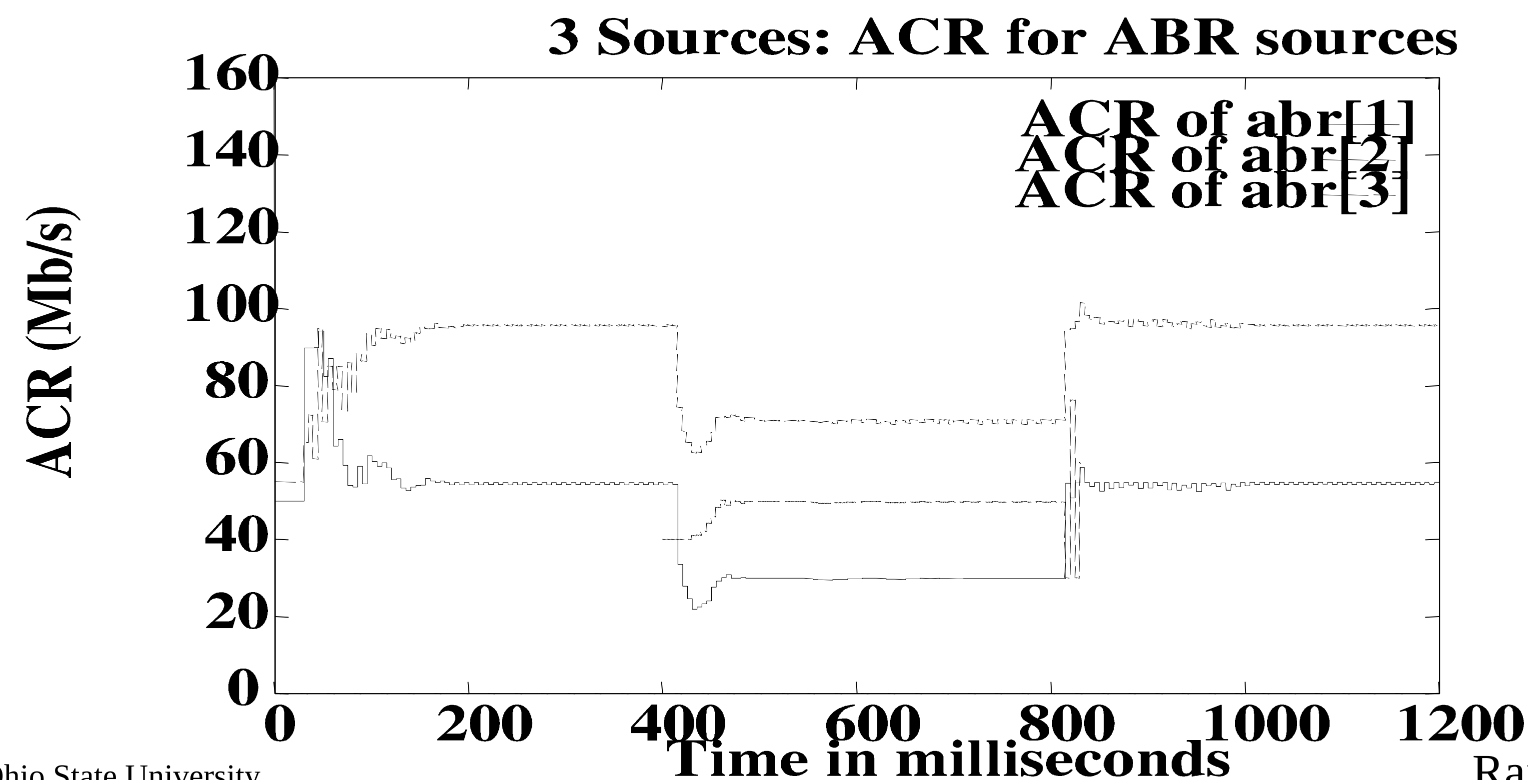


Table 4: Source-Bottleneck

Case Num	Src Num	MCR	a	Wt. Func.	Expectd Fairshre	Using CCR in RMcell	Using Measurd CCR
1	1	0	∞	1	49.92	49.85	49.92
	2	0	∞	1	49.92	49.92	49.92
	3	0	∞	1	49.92	49.92	49.92
2	1	10	∞	1	29.92	-	29.62
	2	30	∞	1	49.92	-	49.60
	3	50	∞	1	69.92	-	71.03
3	1	10	5	15	18.53	-	18.42
	2	30	5	35	49.92	-	49.92
	3	50	5	35	81.30	-	81.93

- Rates converge only if measured source rate is used

Source Bottleneck ACRs

- Case 1: a = 5, MCRs ≠ 0. W(i)=5 + MCR
  Allocation is (16.64, 49.92, 83.2)

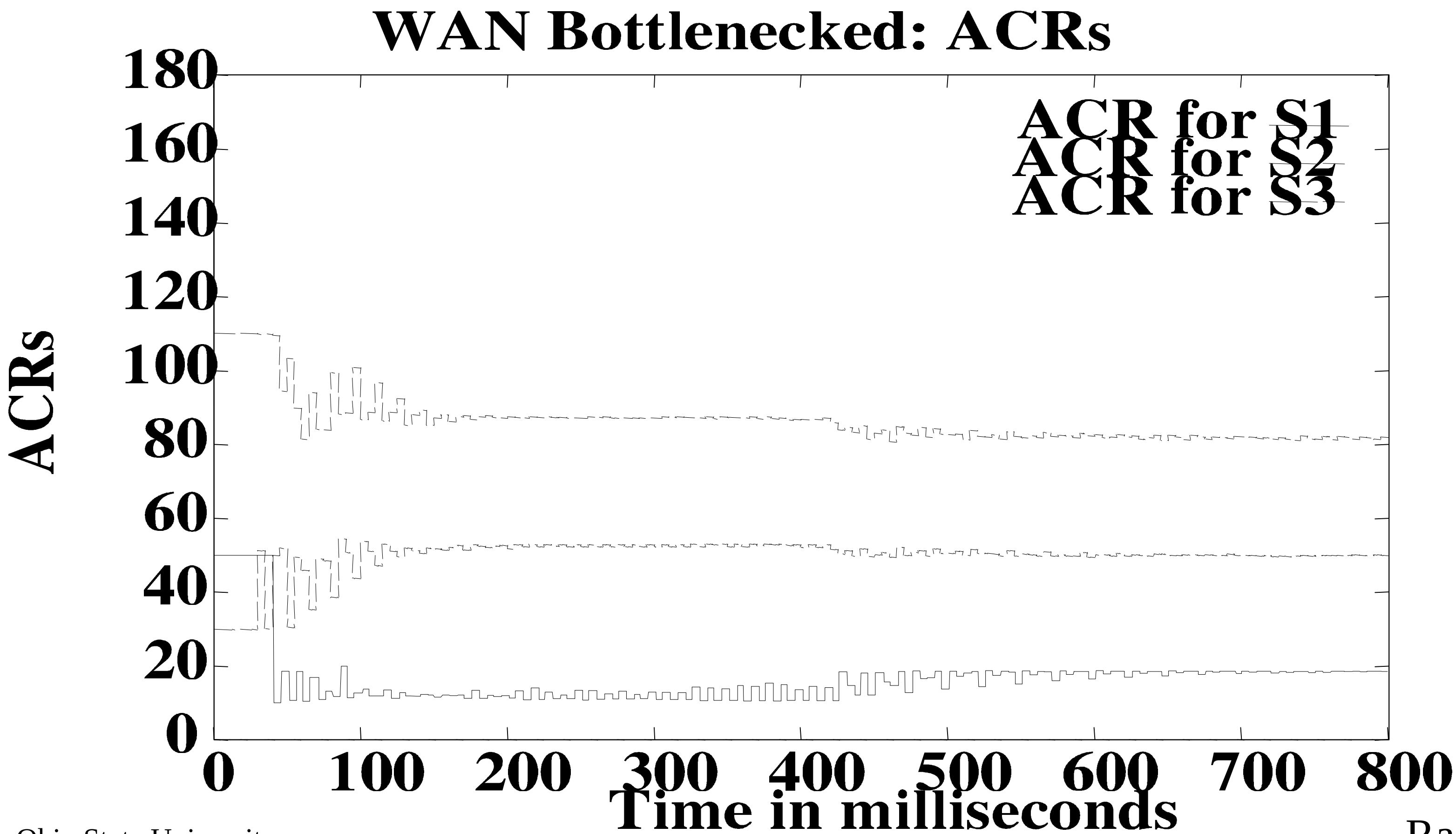


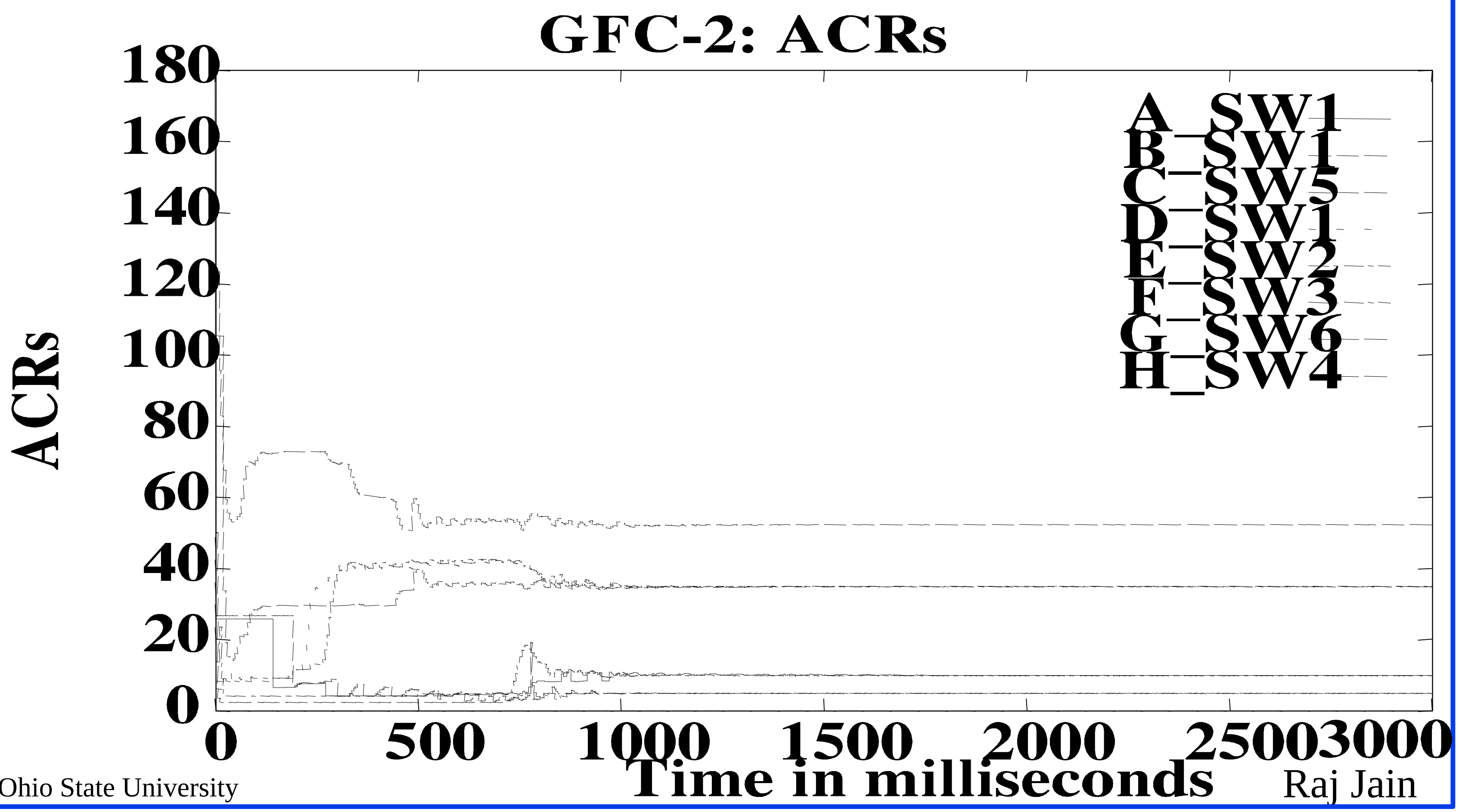
Table 5: GFC-2

VC type	Expected allocation	Actual Allocation
A	10	9.85
B	5	4.97
C	35	35.56
D	35	35.71
E	35	35.34
F	10	10.75
G	5	5.00
H	52.5	51.95

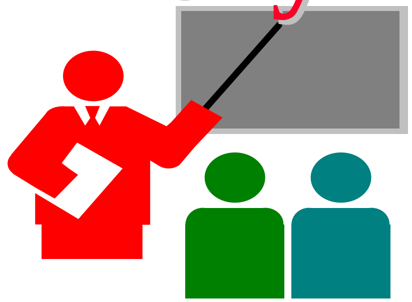
- For all VCs, a = ∞ and MCR=0 (Max-min share). Fairness is achieved in presence of link bottleneck

GFC-2 ACRs

- a = ∞, MCRs = 0. All weights are equal. Allocation as in table 5



Summary



- Fair Allocation = MCR(i)
  + Weighted Share of Excess Bandwidth
- Different TM4.0 definitions map to general fairness
- Effective weight = Weight × Activity level of VCs
- Modified ERICA+ achieves general fairness
- Source bottleneck configuration need per VC accounting to correctly measure the source rate

Motion

Add the following to Section I.3 Example

Fairness Criteria in TM4.0

6. MCR plus weighted share:
   The bandwidth allocation for a connection is its MCR plus a weighted share of the bandwidth B with used MCRs removed.

$$B(i) = MCR(i) + (B-M)\times(w(i)/\text{sum } w(j))$$

Comments: Max-Min, MCR plus equal share, and Allocation proportional to MCR are special cases. The weights may be defined independent of MCR or dependent on MCR.