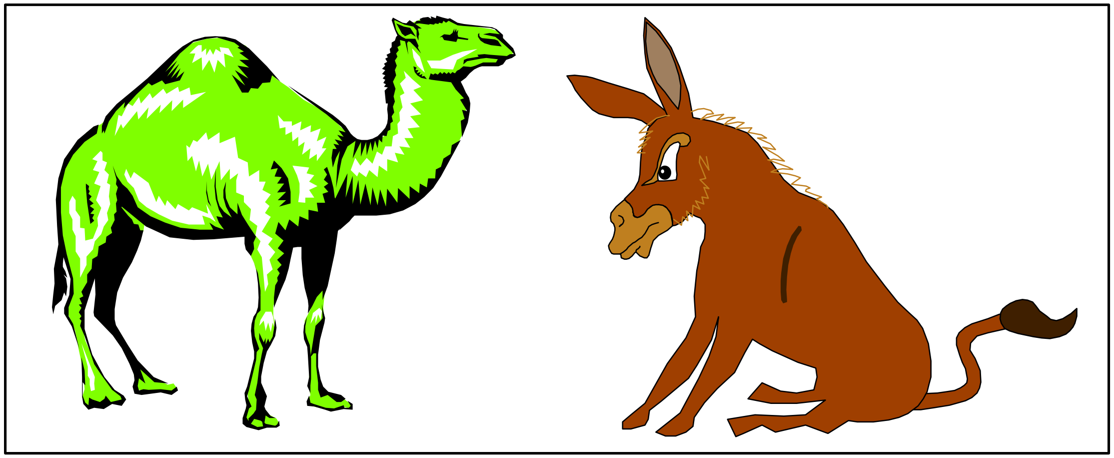
96-0517R1 Buffer Requirements for TCP over ABR



Raj Jain, Shiv Kalyanaraman, Rohit Goyal, Sonia Fahmy
The Ohio State University

Saragur M. Srinidhi
Sterling Software and NASA Lewis Research Center

Contact: Jain@CIS.Ohio-State.Edu

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

- Seven Facts about TCP
- Simulation Results
  ABR + Infinite buffers + 100 ms granularity
  + WAN and LAN
- Effect of RTT, Feedback delay, VBR, Switch scheme, parameters

Seven Facts about TCP

- TCP successfully avoids congestion collapse.
- TCP can automatically fill any available capacity.
- TCP performs best when there is NO packet loss.
  Even a single packet loss can reduce throughput considerably.
- Slow start limits the packet loss but loses considerable time.
  With TCP, you may not lose too many packets but you loose time.
- Bursty losses cause more throughput degradation than isolated losses.
- Fast retransmit/recovery helps in isolated losses but not in bursty losses.
- Timer granularity is the key parameter in determining time lost.

Three Facts about ATM

These apply to ABR as well as UBR:

- Cell loss rate (CLR) gives no indication of throughput loss.
  1% cell loss can cause 50% throughput loss.
  10% cell loss may result in only 10% throughput loss.
- Dropping all cells of a packet is better than dropping randomly (EPD).
- Never drop the EOM cell of a packet.
  It results in two packet losses.

Previous Results About ABR

- The buffers can not be allocated based on TBE
- Maximum queue length and TBE have little/no relationship

Are One RTT Buffers Sufficient?

- Answer 1: Yes. In Many cases.
- Example: Small number of sources. No VBR.
- Answer 2: No. In many cases.
- Example: Large number of sources. Even w/o VBR.

# of Sources	RTT	Feedback Delay	Maximum Queue	Total Throughput	Effici-ency	Fair-ness
5	30	10	10597=0.95*RTT	104.89	83.78	1.0000
10	30	10	14460=1.31*RTT	105.84	84.54	1.0000
15	30	10	15073=1.36*RTT	107.13	85.57	1.0000

Key Factors

- Switch Algorithm: Transient Response (settling) time
- Round Trip Time (RTT)
- Feedback Delay (bottleneck to source)
- Switch Algorithm *Parameters*:
  - Averaging Interval
  - Target Utilization
  - ERICA+ queue control
- Presence and characteristics of background VBR
- Number of VCs
- TCP Receiver window size

Observations About ABR

- ABR performance depends heavily upon the switch algorithm.
  Following statements are based on our *modified ERICA* switch algorithm.
  (For ERICA, see http://www.cis.ohio-state.edu/~jain/)
- No cell loss for *TCP* if switch has Buffers = 4 × RTT.
- No loss for any number of TCP sources w 4 × RTT buffers.
- No loss even with VBR. W/o VBR, 3×RTT buffers will do.
- Under many circumstances, 1× RTT buffers may do.
- Drop policies improve throughput but are not critical.
- In general:
  Qmax = a × RTT + b × Averaging Interval + c × Feedback delay + d × VBR

Modified ERICA

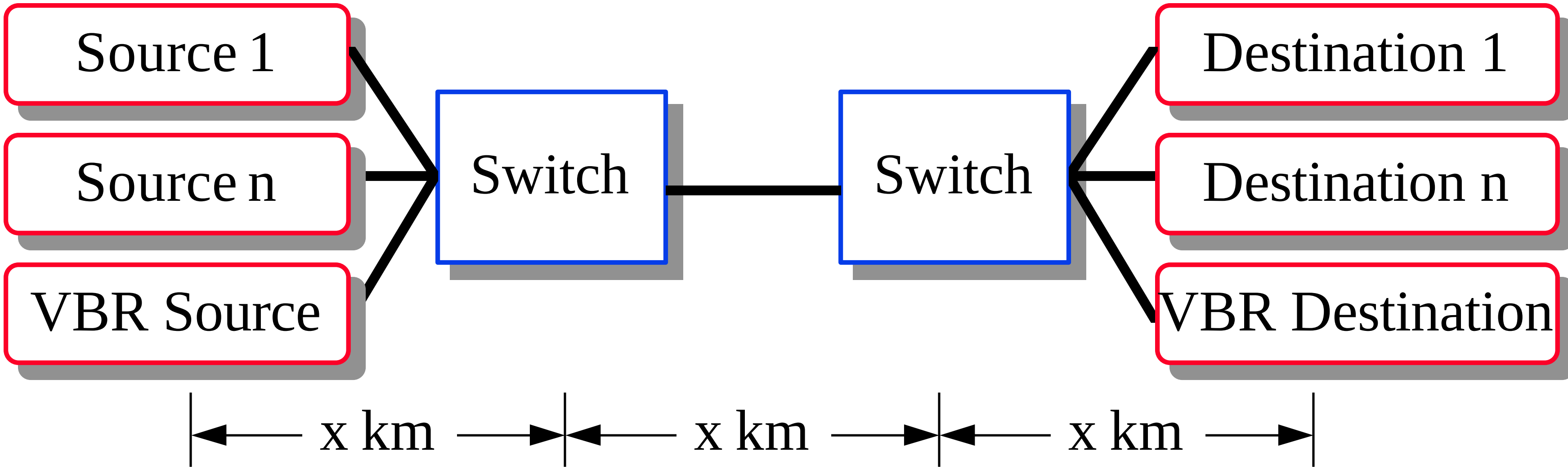
- Eliminates many short spikes
- Provides fast response even if the link is underutilized
- Correctly counts bursty sources
- Allows multiclass scheduling
- Achieves better fairness in many cases
  (Some flows bottlenecked earlier,
  Other flows with ACR≥FS, Overload=1)

Multiclass Scheduling

CBR		
rt-VBR		
nrt-VBR	ABR	UBR

- Ensures *no-starvation* for all classes even under overload.
- Each class has an *allocation* = Guaranteed under overload
- Some classes need minimum delay $\Rightarrow$ have *priority*.
- Some classes are greedy: They will send more than allocated and will want to use all left-over. *No left-over* capacity.
- Left-over capacity must be *fairly* allocated.
- ERICA scheduler achieves all these goals.

n Source + VBR Configuration



- All links 155 Mbps. Lengths x = 1000, 500, 200, 50, 1 km
- If VBR background , y ms on, y ms off, start at t = 2ms
  y=100, 30, 10, 1 ms.
- All traffic unidirectional, Large file transfer application
- Parameters: # sources={5, 10, 15}
  Infinite buffer size.

Simulation Parameters

- Source: Parameters selected to maximize ACR
  TBE = 512
  CDF (XDF) = 0.5
  ICR = 10 Mbps
  CRM (Xrm)= $\lceil$ TBE/Nrm $\rceil$
  ADTF = 0.5 sec
  PCR = 155.52 Mbps, MCR= 0, RIF (AIR) = 1,
  Nrm = 32, Mrm = 2, RDF = 1/512, Trm =100ms, TCR = 10 c/s
- Traffic: TCP/IP with Infinite source application
- Switch: ERICA modified, ERICA+
  Target Utilization = 90% and other values
  Averaging interval = min{100 cells, 1000 μs} and other values

TCP/IP Parameters

- Maximum Segment Size = 512 bytes
- Timer granularity = 100 ms
- No TCP processing time
- Max window = 16 × 64 kB,
  One-way delay = 15 ms = 291 kB
- No delay ack timer
- Fast retransmit/recovery or Early packet drop (EPD) have no impact on these results since there is no loss.

Performance Metrics

- Efficiency = Sum of throughputs/Maximum possible throughput
  - Maximum Segment Size = 512 data
    = 512 data + 20 TCP + 20 IP + 8 LLC + 8 AAL5
    = 12 cells = 12*53 bytes = 636 bytes in ATM Layer
  - Maximum possible throughput = 512/636 = 80.5%
    = 125.2 Mbps on a 155.52 Mbps link*
- Fairness = $$\frac{(\Sigma x_i)^2}{n \Sigma x_i^2}$$

  Where x_i = throughput of the ith TCP source

*ABR loses another 6% for RM cells.

Effect of RTT

# of Sources	RTT	Feedback Delay	Maximum Queue	Total Throughput	Effici-ency	Fair-Ness
15	15	5	12008=2.18*RTT	108.00	86.26	0.9995
15	6	2	6223=2.82*RTT	109.99	87.85	0.9999
15	1.5	0.5	1596=2.89*RTT	110.56	88.31	1.0000

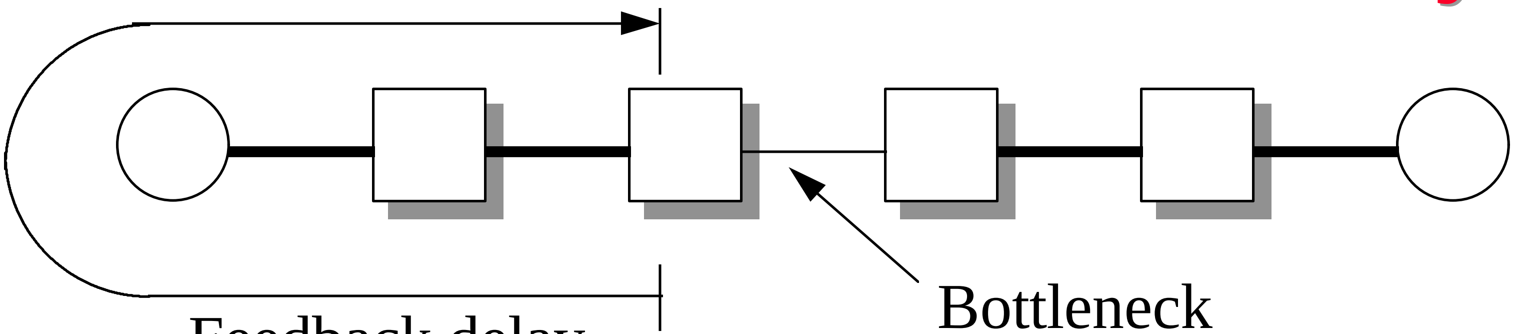
- Maximum queue length approaches 3*RTT, particularly if RTT is medium

TCP/IP over ABR in LANs

- Given a switch algorithm (modified ERICA):
  Qmax = a × RTT + b × Averaging Interval + c × Feedback delay + d × VBR
- In WANs: RTT is the dominant factor
- In LANs: RTT and Feedback delays are small, averaging interval dominates

Averaging Interval	RTT	F/b Delay	Maximum Queue	Total Thruput	Effici-ency	Fair-Ness
10ms,500cells	1.5	0.5	2511=3*RTT +1.71*AI	109.46	87.43	1.00
10ms,1000cells	1.5	0.5	2891=3*RTT +1.24*AI	109.23	87.24	1.00
10ms,500cells	0.03	0.01	2253=4.5*AI	109.34	87.33	1.00
10ms,1000cells	0.03	0.01	3597=3.6*AI	109.81	87.71	0.99

Effect of Feedback Delay



# of Sources	RTT	Feedback Delay	Maximum Queue	Total Throughput	Effici-ency	Fair-Ness
15	30.02	30	8701=0.79*RTT	106.78	85.29	1.0000
15	30.02	0.01	719=0.07*RTT	105.94	84.62	1.0000
15	10.02	10	5259=1.43*RTT	109.31	87.31	0.9999
15	10.02	0.01	709=0.19*RTT	108.69	86.81	1.0000

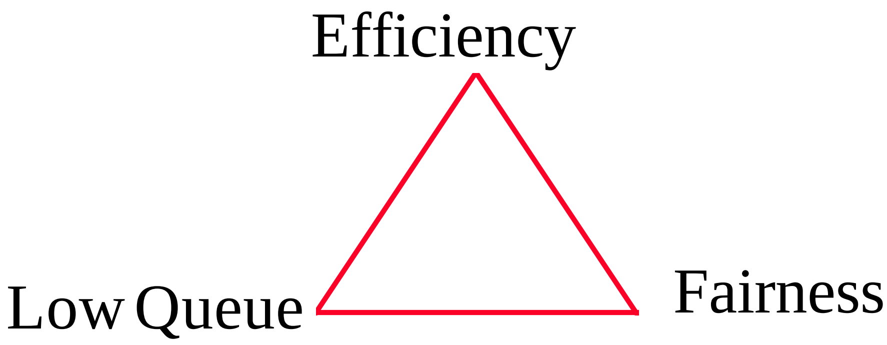
- Smaller feedback delay ⇒ Smaller queues

High Frequency VBR: Problem

- Limit of 1 × RTT due to VBR is good for large VBR cycle times.
  TCP and ABR get enough time to adjust.
- Faster VBR causes faster variations in available capacity. Neither TCP nor Switch algorithm may have time to adjust
  $\Rightarrow$ Can lead to instability at high utilization levels.

VBR On/Off	RTT	F/b Delay	Maximum Queue	Total Throughput	Effici-ency	Fair-Ness
30 ms	30	10	12359=1.12*RTT	69.60	92.65	0.9967
100 ms	30	10	13073=1.18*RTT	63.85	85.00	0.9987
10 ms	30	10	diverges			
1 ms	30	10	diverges			

Three Way Tradeoff



- Buffers vs Efficiency (Utilization) vs Fairness
- It is possible to have lower queues (lower buffer required) if the target utilization is kept low.

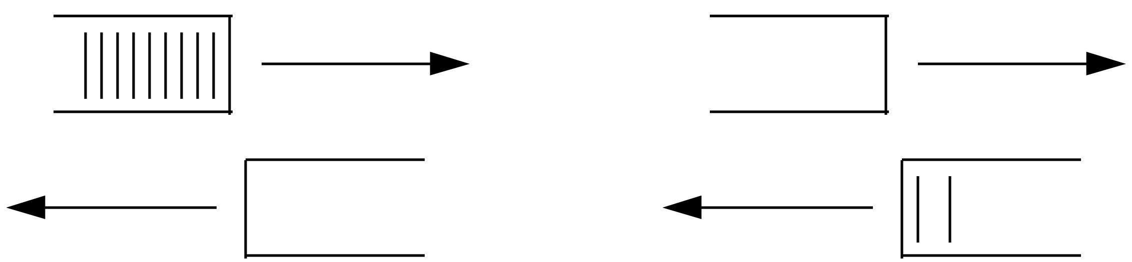
High Frequency VBR: Solution

- ERICA with target at 70%
- ERICA+ with queue delay of 0.5 ms
  ERICA+ gives high efficiency and stability
  Automatically compensates for measurement errors in input rate, available capacity, or number of active sources

Scheme	RTT	F/b Delay	Maximum Queue	Total Throughput	Efficiency	Fair-Ness
ERICA+	30	10	5435=0.49*RTT	69.22	92.15	0.9827
Target=70%	30	10	12359=1.12*RTT	50.52	67.25	0.9958

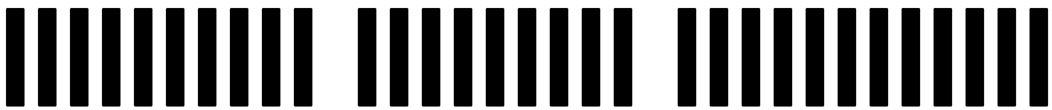
Out-Of Phase Effect

- Bursty load and backward RM cells are often out of phase.
- When there is load in the forward direction, there are no BRMs.
- By the time the switch sees BRMs, there is no load in the forward direction.
- The above effect disappears when the bursts become larger than RTT

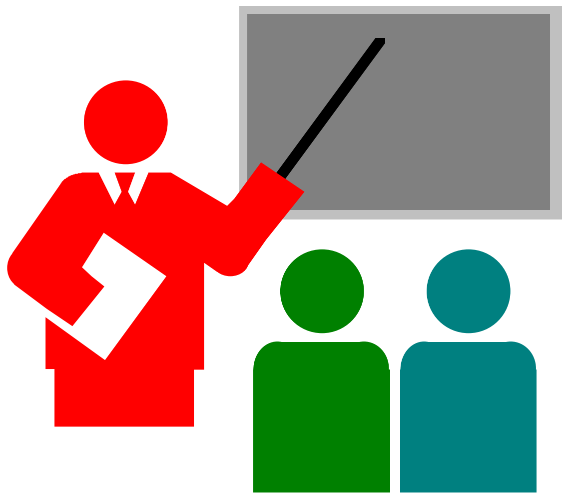


Flocking Effect

- All cells of a VC are often seen together.
- There is clustering of sources.
- Not all sources are seen all the time.



Summary



- Performance of ABR depends on RTT, the switch algorithm and its parameters
- For modified ERICA, 4*RTT buffers are sufficient
- For ERICA+, queue can be controlled to any desired level
- There is a efficiency, buffer size, and fairness tradeoff

REFERENCES

All our past ATM forum contributions, papers and presentations can be obtained on-line at http://www.cis.ohio-state.edu/~jain/

[1] R. Jain, S. Kalyanaraman, R. Viswanathan, R. Goyal,"A Sample Switch Algorithm," ATM Forum/95-0178R1, February 1995.

[2] R. Jain, S.Kalyanaraman, R. Goyal, S.Fahmy, F. Lu, S.M. Srinidhi, "TCP/IP over ABR(Was: TBE and TCP/IP Traffic)," ATM Forum/96-0177R1

[3] S. Kalyanaraman, R. Jain, S.Fahmy, R. Goyal, F.Lu, S.M. Srinidhi, "Performance of TCP/IP over ABR," submitted to Globecom'96.