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 cannot 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.

<table>
<thead>
<tr>
<th># of Sources</th>
<th>RTT</th>
<th>Feedback Delay</th>
<th>Maximum Queue</th>
<th>Total Throughput</th>
<th>Efficiency</th>
<th>Fairness</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>30</td>
<td>10</td>
<td>1010597=0.95*RTT</td>
<td>104.89</td>
<td>83.78</td>
<td>1.0000</td>
</tr>
<tr>
<td>10</td>
<td>30</td>
<td>10</td>
<td>1014460=1.31*RTT</td>
<td>105.84</td>
<td>84.54</td>
<td>1.0000</td>
</tr>
<tr>
<td>15</td>
<td>30</td>
<td>10</td>
<td>1015073=1.36*RTT</td>
<td>107.13</td>
<td>85.57</td>
<td>1.0000</td>
</tr>
</tbody>
</table>
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:
  \[ Q_{\text{max}} = a \times \text{RTT} + b \times \text{Averaging Interval} + c \times \text{Feedback delay} + d \times \text{VBR} \]
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

- Ensures *no-starvation* for all classes even under overload.
- Each class has an *allocation* = Guaranteed under overload
- Some classes need minimum delay ⇒ 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.
All links 155 Mbps. Lengths \( x = 1000, 500, 200, 50, 1 \) km

If VBR background, \( y \) ms on, \( y \) ms off, start at \( t = 2\)ms
\( y=100, 30, 10, 1 \) 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) = \[ \left\lfloor \frac{TBE}{Nrm} \right\rfloor \]
  - 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 = \( \frac{(\sum x_i)^2}{n \sum x_i^2} \)
  Where \( x_i \) = throughput of the \( i \)th TCP source
  *ABR loses another 6% for RM cells.
Effect of RTT

<table>
<thead>
<tr>
<th># of Sources</th>
<th>RTT</th>
<th>Feedback Delay</th>
<th>Maximum Queue</th>
<th>Total Throughput</th>
<th>Efficiency</th>
<th>Fairness</th>
</tr>
</thead>
<tbody>
<tr>
<td>15</td>
<td>15</td>
<td>5</td>
<td>12008=2.18*RTT</td>
<td>108.00</td>
<td>86.26</td>
<td>0.9995</td>
</tr>
<tr>
<td>15</td>
<td>6</td>
<td>2</td>
<td>6223=2.82*RTT</td>
<td>109.99</td>
<td>87.85</td>
<td>0.9999</td>
</tr>
<tr>
<td>15</td>
<td>1.5</td>
<td>0.5</td>
<td>1596=2.89*RTT</td>
<td>110.56</td>
<td>88.31</td>
<td>1.0000</td>
</tr>
</tbody>
</table>

- Maximum queue length approaches 3*RTT, particularly if RTT is medium
TCP/IP over ABR in LANs

- Given a switch algorithm (modified ERICA):
  \[ Q_{\text{max}} = a \times \text{RTT} + b \times \text{Averaging Interval} + c \times \text{Feedback delay} + d \times VBR \]
- In WANs: RTT is the dominant factor
- In LANs: RTT and Feedback delays are small, averaging interval dominates

<table>
<thead>
<tr>
<th>Averaging Interval</th>
<th>RTT</th>
<th>F/b Delay</th>
<th>Maximum Queue</th>
<th>Total Thruput</th>
<th>Efficiency</th>
<th>Fairness</th>
</tr>
</thead>
<tbody>
<tr>
<td>10ms,500cells</td>
<td>1.5</td>
<td>0.5</td>
<td>2511=3<em>RTT+1.71</em>AI</td>
<td>109.46</td>
<td>87.43</td>
<td>1.00</td>
</tr>
<tr>
<td>10ms,1000cells</td>
<td>1.5</td>
<td>0.5</td>
<td>2891=3<em>RTT+1.24</em>AI</td>
<td>109.23</td>
<td>87.24</td>
<td>1.00</td>
</tr>
<tr>
<td>10ms,500cells</td>
<td>0.03</td>
<td>0.01</td>
<td>2253=4.5*AI</td>
<td>109.34</td>
<td>87.33</td>
<td>1.00</td>
</tr>
<tr>
<td>10ms,1000cells</td>
<td>0.03</td>
<td>0.01</td>
<td>3597=3.6*AI</td>
<td>109.81</td>
<td>87.71</td>
<td>0.99</td>
</tr>
</tbody>
</table>
## Effect of Feedback Delay

- Smaller feedback delay $\Rightarrow$ Smaller queues

<table>
<thead>
<tr>
<th># of Sources</th>
<th>RTT</th>
<th>Feedback Delay</th>
<th>Maximum Queue</th>
<th>Total Throughput</th>
<th>Efficiency</th>
<th>Fairness</th>
</tr>
</thead>
<tbody>
<tr>
<td>15</td>
<td>30.02</td>
<td>0.01</td>
<td>719=0.07*RTT</td>
<td>105.94</td>
<td>84.62</td>
<td>1.0000</td>
</tr>
<tr>
<td>15</td>
<td>10.02</td>
<td>10</td>
<td>5259=1.43*RTT</td>
<td>109.31</td>
<td>87.31</td>
<td>0.9999</td>
</tr>
<tr>
<td>15</td>
<td>10.02</td>
<td>0.01</td>
<td>709=0.19*RTT</td>
<td>108.69</td>
<td>86.81</td>
<td>1.0000</td>
</tr>
<tr>
<td>15</td>
<td>30.02</td>
<td>30</td>
<td>8701=0.79*RTT</td>
<td>106.78</td>
<td>85.29</td>
<td>1.0000</td>
</tr>
</tbody>
</table>

- Smaller feedback delay $\Rightarrow$ Smaller queues
High Frequency VBR: Problem

- Limit of $1 \times RTT$ 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 ⇒ Can lead to instability at high utilization levels.

<table>
<thead>
<tr>
<th>VBR On/Off</th>
<th>RTT</th>
<th>F/b Delay</th>
<th>Maximum Queue</th>
<th>Total Throughput</th>
<th>Efficiency</th>
<th>Fair-ness</th>
</tr>
</thead>
<tbody>
<tr>
<td>30 ms</td>
<td>30</td>
<td>10</td>
<td>$12359=1.12\times RTT$</td>
<td>69.60</td>
<td>92.65</td>
<td>0.9967</td>
</tr>
<tr>
<td>100 ms</td>
<td>30</td>
<td>10</td>
<td>$13073=1.18\times RTT$</td>
<td>63.85</td>
<td>85.00</td>
<td>0.9987</td>
</tr>
<tr>
<td>10 ms</td>
<td>30</td>
<td>10</td>
<td>diverges</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 ms</td>
<td>30</td>
<td>10</td>
<td>diverges</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
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

<table>
<thead>
<tr>
<th>Scheme</th>
<th>RTT</th>
<th>F/b Delay</th>
<th>Maximum Queue</th>
<th>Total Throughput</th>
<th>Efficiency</th>
<th>Fairness</th>
</tr>
</thead>
<tbody>
<tr>
<td>ERICA+</td>
<td>30</td>
<td>10</td>
<td>5435=0.49*RTT</td>
<td>69.22</td>
<td></td>
<td>0.9827</td>
</tr>
<tr>
<td>Target=70%</td>
<td>30</td>
<td>10</td>
<td>12359=1.12*RTT</td>
<td>50.52</td>
<td></td>
<td>0.9958</td>
</tr>
</tbody>
</table>
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/

