STQ Size
and
Priority Inversion

Jon Schuringa

May, 2003

Contents

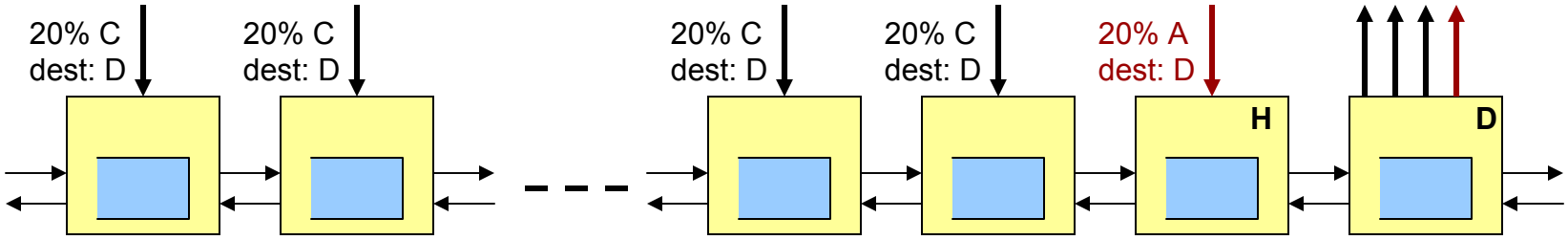
- The problem
- STQ size formula
- Implications
- A solution
- Conclusion

The Problem

- Problem:
  - Major priority inversion with long duration
  - Class A0, A1 and B are affected
  - In conservative and aggressive mode
- Reason:
  - A large, continues stream of class C traffic may build up on the transit path, and there is no mechanism to stop this traffic. The STQ reaches the full threshold at stations sending A or B traffic.

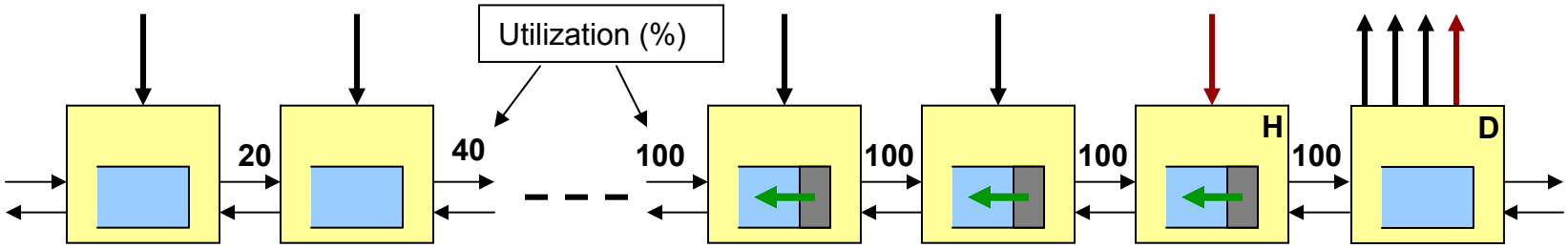
Example Scenario

1



Initially, all STQ‘s are empty

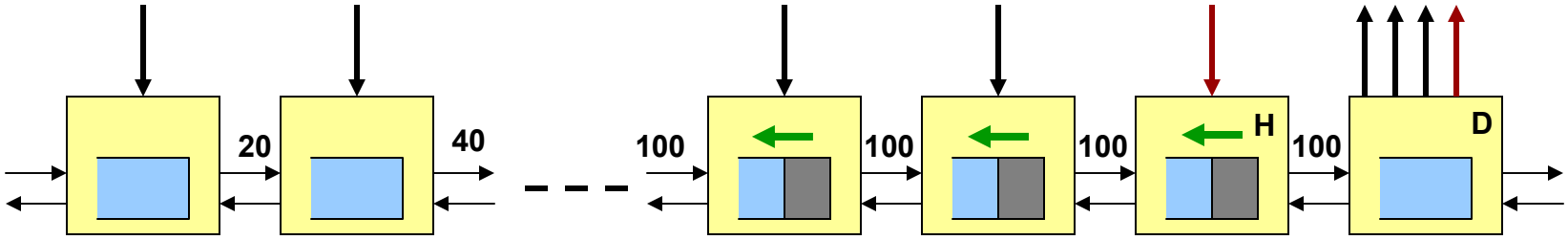
2



All stations send at 20% line rate, STQ‘s start to fill at stations receiving a lot of traffic

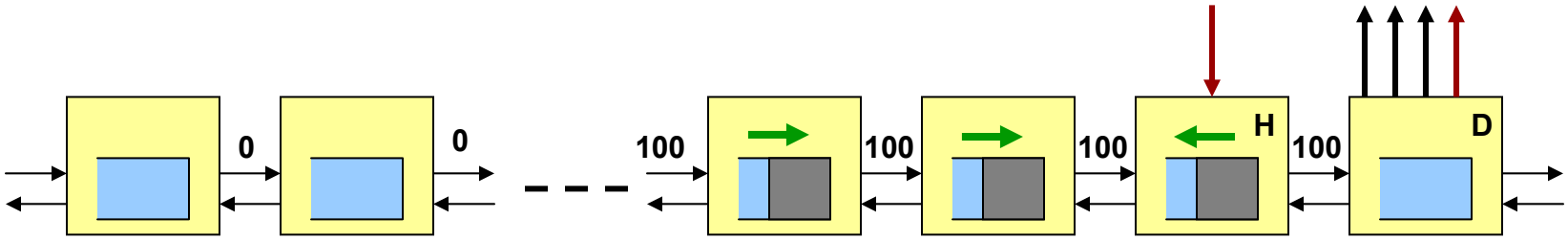
Example Scenario (cont.)

3



STQLowThreshold(s) reached, congestion advertised through FCM‘s, STQ‘s still growing

4



FCM‘s received, no class C traffic added anymore (simplifying assumption). All STQ‘s are now flushed, except at the head station H: the STQ at H grows as long as it receives 100% from its upstream neighbors

STQ size

- How much traffic can be expected at the head station (aggressive):

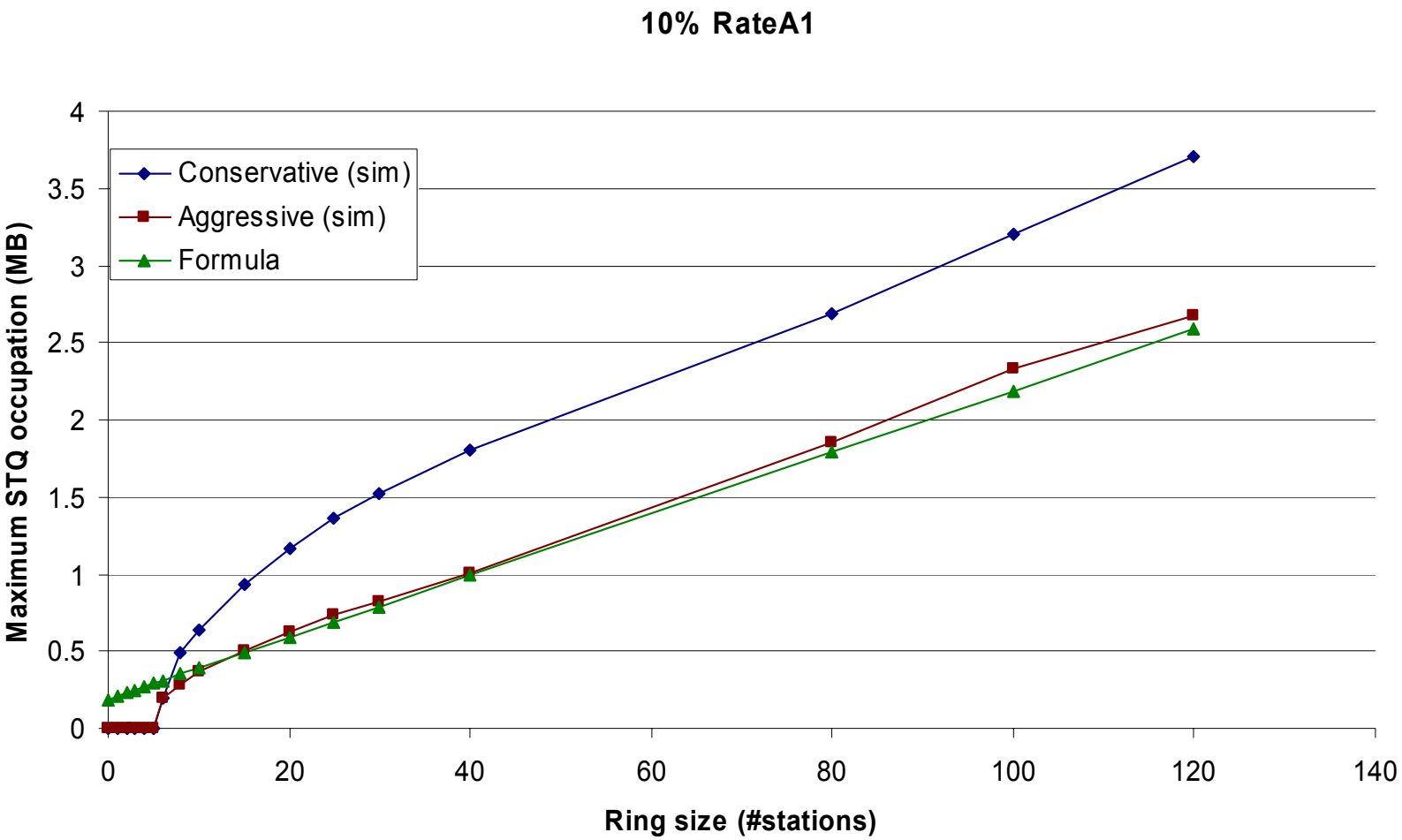
$$sizeSTQ \approx Min(STQLowThreshold + X, STQHighThreshold) * N * \frac{rateA}{LR} + C$$

$$C = STQLowThreshold + (STQHighThreshold - STQLowThreshold) * \frac{LR - rateA}{LR}$$

- N is number of stations, LR is line rate
- X depends on N, ring size, aggressive or conservative mode, aging parameters, advertisement interval, etc.
- Worst case:

$$sizeSTQ \approx STQHighThreshold * N * \frac{rateA}{LR} + C$$

STQ size (cont.)



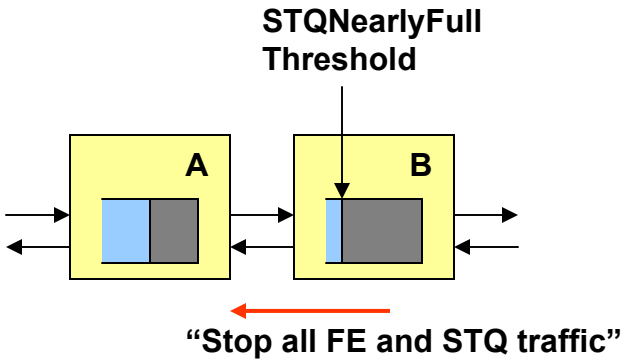
- STQLowThreshold 100KB, STQHighThreshold 200KB

Implications

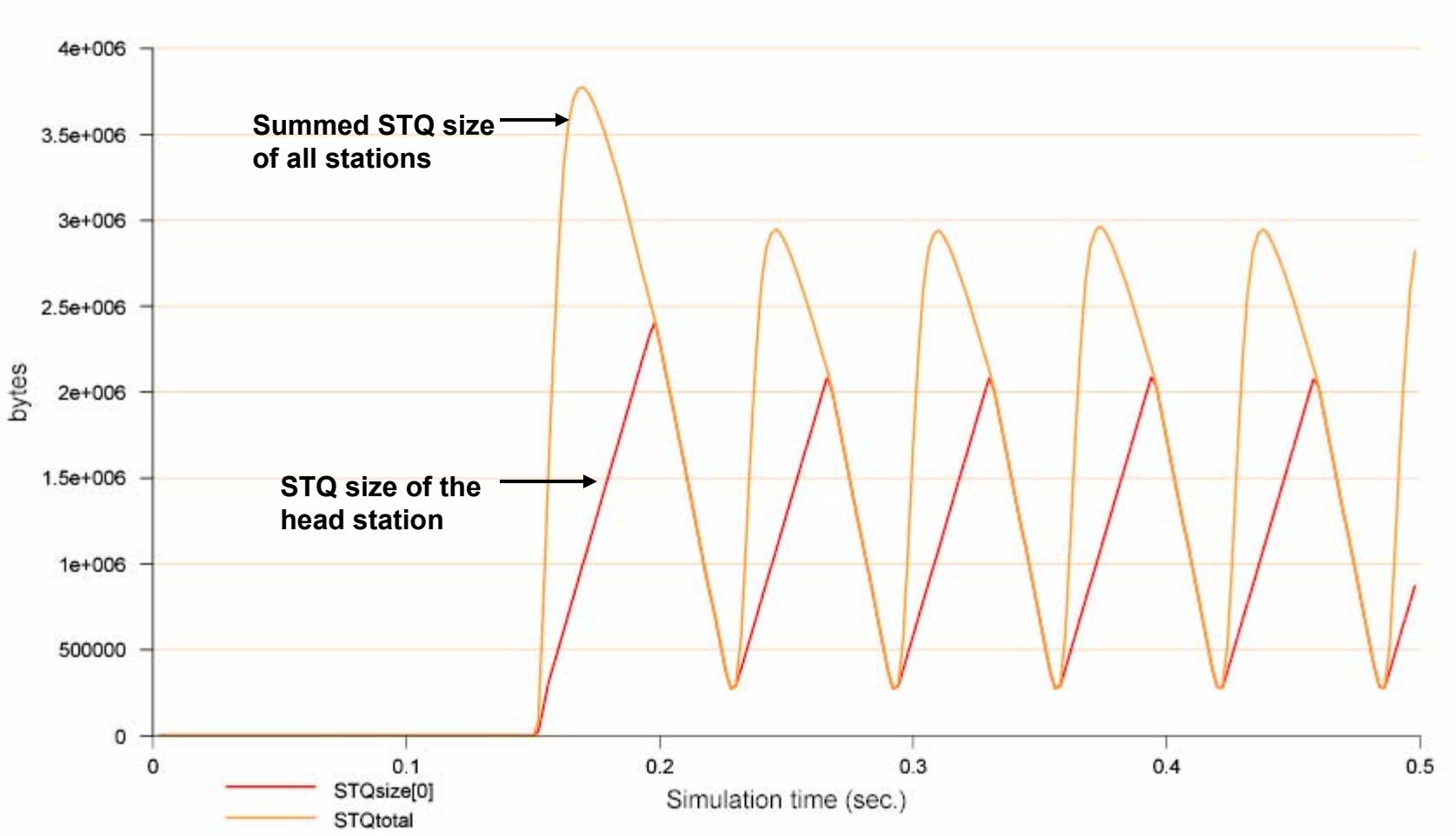
- STQ must be very large in large rings
- With the current draft and its default values, priority inversion may already occur in small rings (<16 stations in aggressive mode).
- Results in high class B and C delays
- This cannot be fixed by adjusting the rateA0 and rateA1 levels, since both are affected

A solution

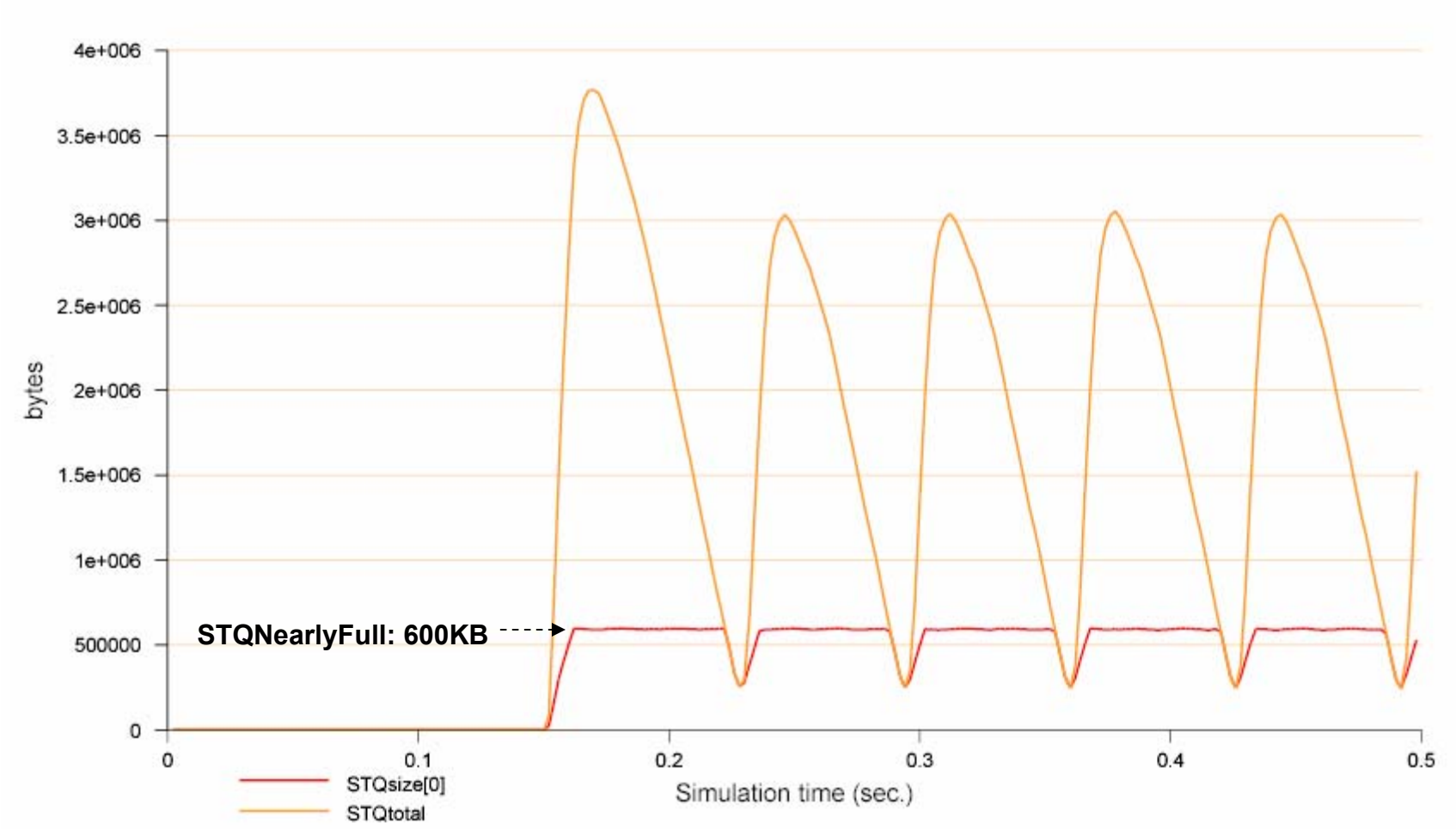
- Backpressure signal:
  - Tell upstream neighbor to stop sending from its STQ and all local sourced FE-traffic if the “STQNearlyFullThreshold“ is reached.
  - Advertised in FCM (1 bit needed)
  - Trivial and effective
  - STQ‘s can be much smaller
  - But:
    - Potential HOL blocking (traffic in A, destination B, is unnecessarily blocked)
    - Class B,C delay and jitter affected?



Without backpressure



With backpressure



Solution (cont.)

- Interestingly, the external observable RPR behavior remains unchanged with the backpressure signal (in the example scenario):
  - End to end delay and throughput unchanged
  - The transit traffic is now spread over multiple STQ‘s. The total amount of queued traffic remains the same.

Conclusions

- With backpressure signal:
  - No major priority inversion
  - STQ‘s can be smaller
  - Different sized STQ‘s on the same ringlet is no problem anymore
  - Easy to give an upper bound for the STQ size occupation (equal for aggressive and conservative mode)
- To do:
  - prove correctness and study possible impact on throughput, delay and jitter
  - Use a shaper instead of on/off backpressure