Analysis of TCP with Several Bottleneck Nodes

Chadi Barakat and Eitan Altman
INRIA Sophia Antipolis - FRANCE

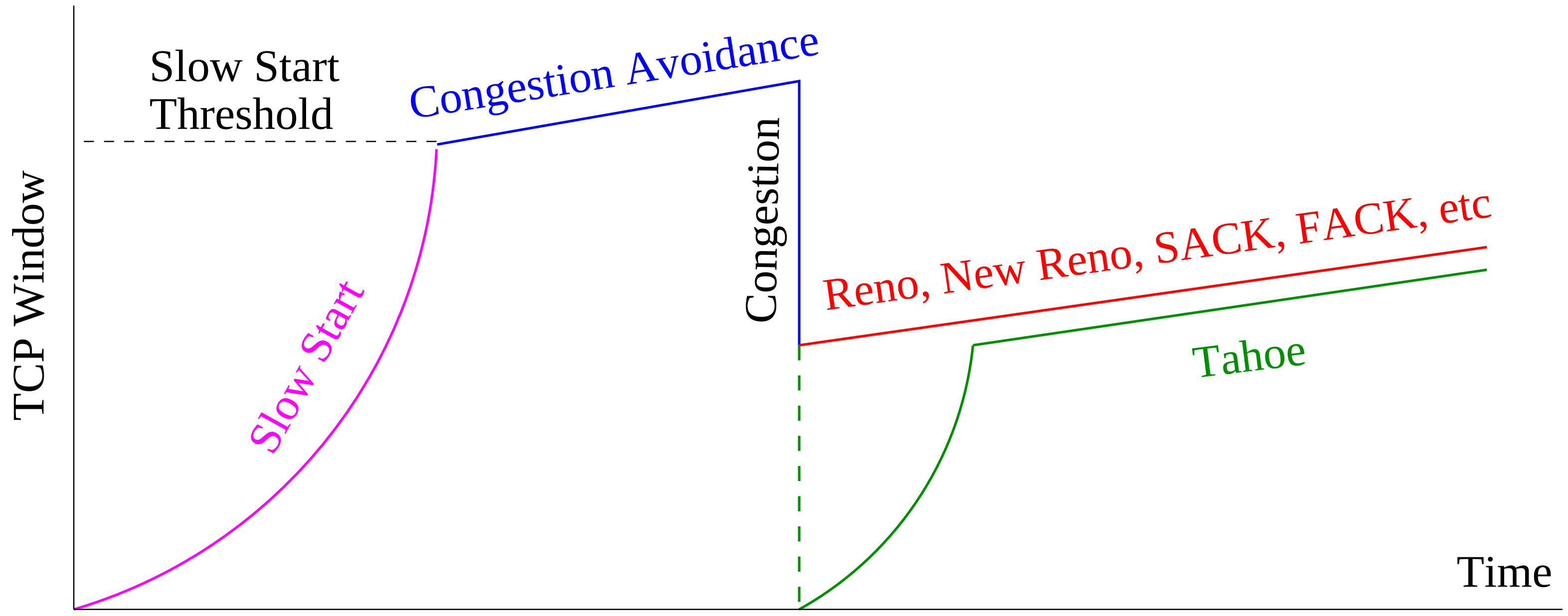
{Chadi.Barakat,Eitan.Altman}@sophia.inria.fr

Monday, December 6 1999

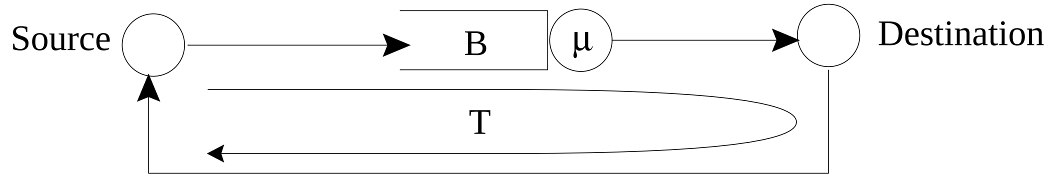
Organization of the talk

- Background on TCP congestion control algorithms.
- Insufficiency of the single node network model.
- Definition of a general network model for TCP analysis.
- Study of TCP algorithms with the general model.
- Simplification of the general model to a two-node model.
- Impact of network parameters on TCP performance.
- Conclusions and general guidelines.

Background on TCP



On TCP modeling: The single node network model



- The connection path is modeled only by the main bottleneck node.
- But, TCP is bursty (especially during Slow Start) and a queue can build up in some other nodes in addition to the main bottleneck.
- Different performance can be obtained if the buffers in the other nodes overflow before that of the main bottleneck. Also, the other nodes may absorb the ovrload of TCP bursts on the main bottlebneck.

A more general model is required ...

A general network model for TCP analysis

- Due to the reliance on the ACK clock, TCP transmits packets in long bursts at up to twice the bottleneck bandwidth (the twice appears during slow start, when the receiver acknowledges every packet and when ACKs are not lost).
- Let N_μ denote a router with outgoing rate μ.
- Nodes upstream N_μ and having a rate less than 2μ must be considered.

$$\mu < \mu_n < \mu_{n-1} < \ldots < \mu_1 < 2\mu$$

TCP behavior during Congestion Avoidance

– The single node model predicts a queue building in N_μ at a rate of one packet per RTT until the window reaches $W_{max} = B + \mu T$.

– But, an increase in the window by one packet every RTT results in a transmission rate $R_{CA} = \frac{W+1}{W}\mu > \mu$.

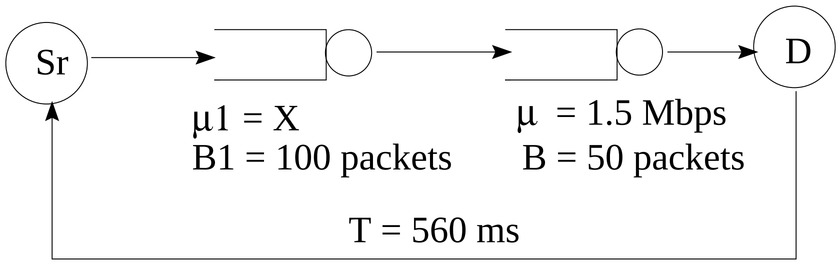
– Although this rate is very close to μ, the queue will not build up entirely in N_μ if one of the upstream nodes has an outgoing rate less than R_{CA}.

– Depending on the buffer size in this upstream node, the result can be:

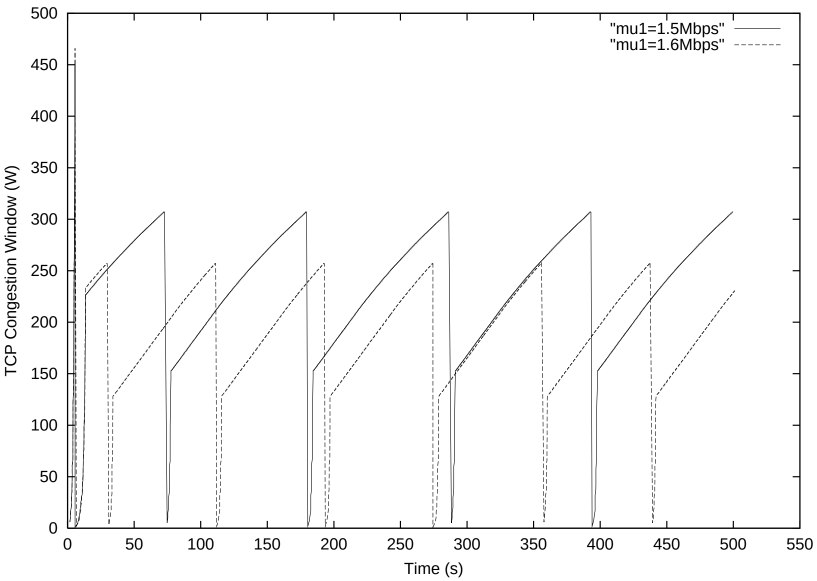
  * A smaller W_{max}, then a poorer performance.
  * Or a larger W_{max}, then a better performance.

A simulation example

Using the ns simulator developped at LBNL:



Two values for X: 1.5 Mbps and 1.6 Mbps.



TCP behavior during Slow Start

- The main objective of this phase is to increase quickly the congestion window until reaching the Slow Start threshold.
- This fast increase results in burstiness. It has been shown, in case of the single node model, that a buffer size larger than $\mu T/3$ is required in the main bottleneck to absorb this burstiness and to reach without losses a W_{th} equal to half the pipe size.
- But, this burstiness of Slow Start increases the impact of upstream nodes on the performance.
- Again, losses in the main bottleneck may be avoided.
  Or losses not predicted by the single node model may appear.

The overflow window

– Let W_B be the window at which losses occur during Slow Start. W_{th} must be set to less than W_B in order to avoid losses during Slow Start and to improve the performance. In contrast to the single node model, we calculate W_B as a function of all network parameters.

– Calculation of W_B:

  * Packets are transmitted in long bursts that double every RTT.
  * The burst size (in packets) required to overflow a buffer:

$$\begin{array}{ll} S_1 = 2\mu B_1/(2\mu - \mu_1) & \text{for } B_1, \\ S_i = \mu_{i-1} B_i/(\mu_{i-1} - \mu_i) & \text{for } B_i \ (i = 2 \ldots n), \\ S = \mu_n B/(\mu_n - \mu) & \text{for } B. \end{array}$$

The overflow window

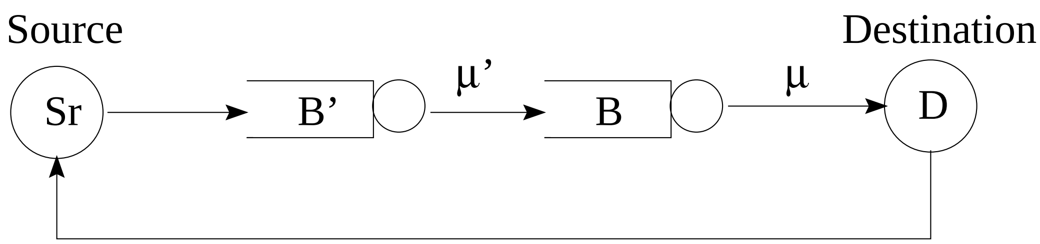
The first buffer that will overflow is the one with the smallest burst size. Thus,

$$W_B \simeq \min(S_1, S_2, \ldots, S_n, S)$$

- In contrast to the single node model, W_B is a function not only of B and μ but also of B_i and μ_i $(i = 1 \ldots n)$.
- Even if B is large compared to μT, W_B may be smaller than W_{th} due to a small buffer in an upstream node. Unpredicted losses will then appear.
- The single node model underestimates S (then W_B) since it supposes that the input rate at B is equal to 2μ while it can be constrained by an upstream node. The result is in an overestimation of the required buffer size.

Simplification of the general model

The main bottleneck (μ, B) is required for the modeling of Congestion Avoidance (W_{max}). And upstream nodes are required for the modeling of Slow Start. We keep the main bottleneck and we substitute the upstream nodes by the node between 1 and n having the smallest burst size.

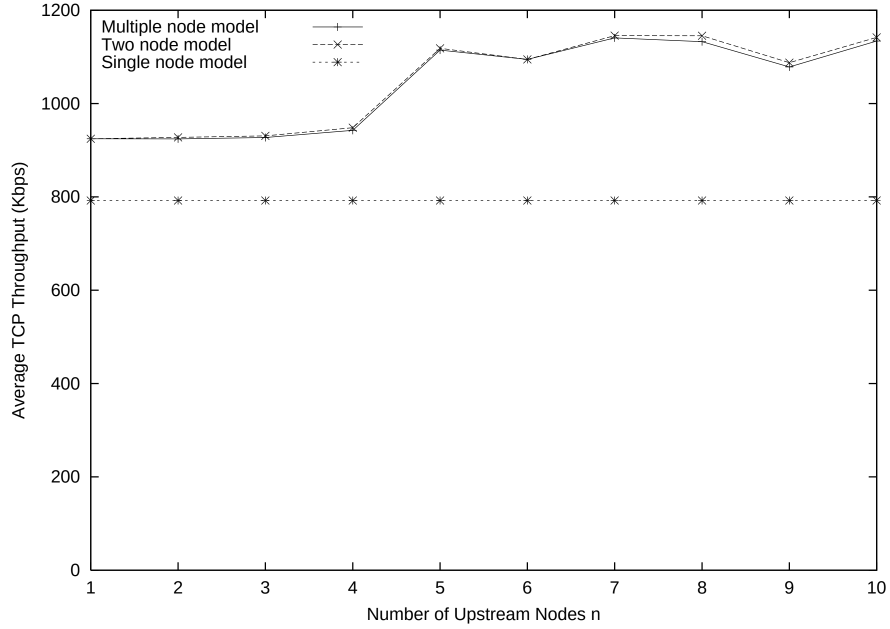


$$\mu' = \mu_n, \; B' \text{ verfies } S' = 2\mu B'/(2\mu - \mu') = \min_{i=1\ldots n}(S_i)$$

Validation of the simplified model

We change the number of upstream nodes from 1 to 10, and we distribute the inter-upstream node bandwidth equally on $[\mu, 2\mu]$.

For a long TCP Tahoe connection, the single node model predicts always an overflow in B during Slow Start at a small W_B whereas our models predict an overflow at a larger W_B until $n = 4$ and a disappearance of the overflow after that.



Impact of network parameters on TCP performance

- Due to the closeness of R_{CA} to μ, the behavior of TCP in the Congestion Avoidance mode is only a function of B and μ and it is identical to that predicted by the single node model.

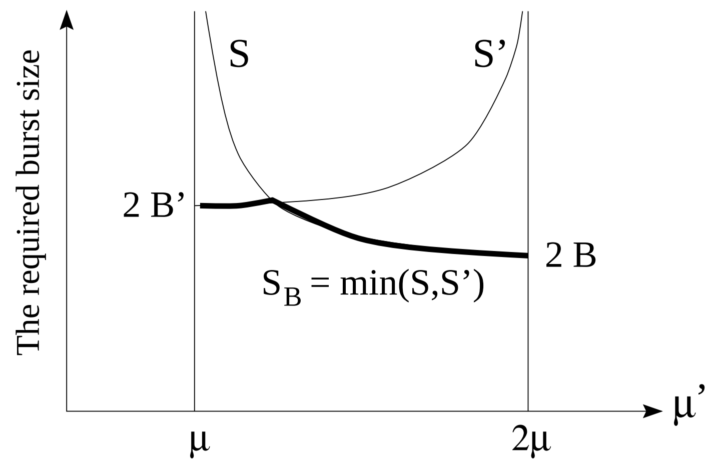
- The behavior of TCP during Slow Start depends on the position of W_B w.r.t. W_{th} where $W_B = \min(S, S')$ is a function of B, B', μ and the value of μ' between μ and 2μ.

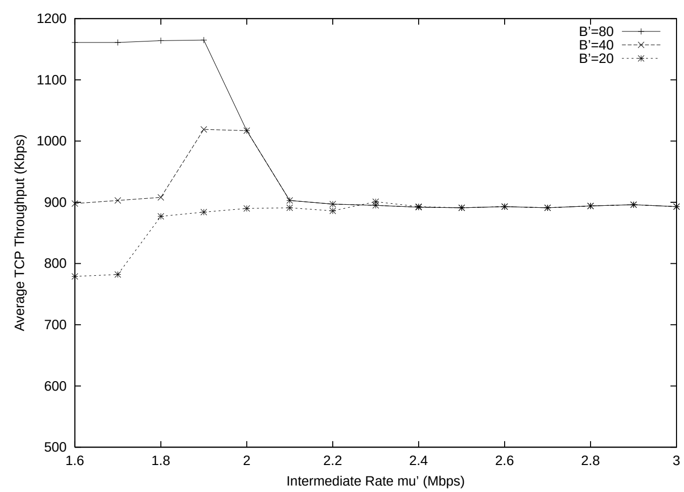
- Given that the minimum value of S and S' are $2B$ and $2B'$ respectively, the impact of μ' can be eliminated by taken:

$$2B > W_{th} \text{ and } 2B' > W_{th}$$

Impact of μ' on the performance

This impact exists when one of the two previous conditions is not satisfied. While μ' decreases, the situation improves and then worsen.





The case of a small B that it is not able to absorb Slow Start burstiness alone.

Conclusions

- The nodes upstream the main bottleneck must be considered.

- A minimum buffer size is required in an upstream node in case its outgoing rate fells below 2μ. To account for all the values of μ', B' as B, must be chosen larger than half the Slow Start threshold.

For a W_{th} function of B (e.g. half the pipe size), any increase in B must be accompanied by a less important increase in B'.