A Statistical Bandwith Sharing Perspective on Buffer Sizing

Jordan Augé
(joint work with James Roberts in Orange Labs)

December 4th, 2008, LIP6, Paris

Buffer sizing issue

- The *Bandwidth Delay Product* dimensioning rule is not appropriate for high speed links
  - Buffer $\sim$ link rate x RTT
  - eq. 2.5Gb for 10Gb/s links (RTT=250ms)

Buffer sizing issue

- The *Bandwidth Delay Product* dimensioning rule is not appropriate for high speed links
  - Buffer $\sim$ link rate x RTT
  - eq. 2.5Gb for 10Gb/s links (RTT=250ms)
- Buffer size could be reduced if we assume a large number of flows, or paced TCP flows
  - Appenzeler et al. : Buffer $\sim 1/\sqrt{\text{number of flows}}$
  - Raina and Wishik : Buffer $\sim$ 20 packets

Buffer sizing issue

- The *Bandwidth Delay Product* dimensioning rule is not appropriate for high speed links
  - Buffer $\sim$ link rate x RTT
  - eq. 2.5Gb for 10Gb/s links (RTT=250ms)
- Buffer size could be reduced if we assume a large number of flows, or paced TCP flows
  - Appenzeler et al. : Buffer $\sim 1/\sqrt{\text{number of flows}}$
  - Raina and Wishik : Buffer $\sim$ 20 packets
- But these assumptions are unrealistic

Traffic characteristics at flow level

- Finite size flow arrivals according to a stochastic process
- The number of flows in progress is a random value

Traffic characteristics at flow level

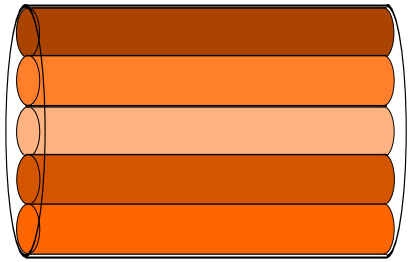
- Finite size flow arrivals according to a stochastic process
- The number of flows in progress is a random value
- The flow peak rate is an essential characteristic and determines a typical traffic mix
  - Most flows have a peak rate much less than the link rate
  - A small number of flows have a high peak rate and dynamically share link bandwidth

Link utilization regimes

a transparent regime
the sum of the peak rates of the flows is less than capacity with high probability

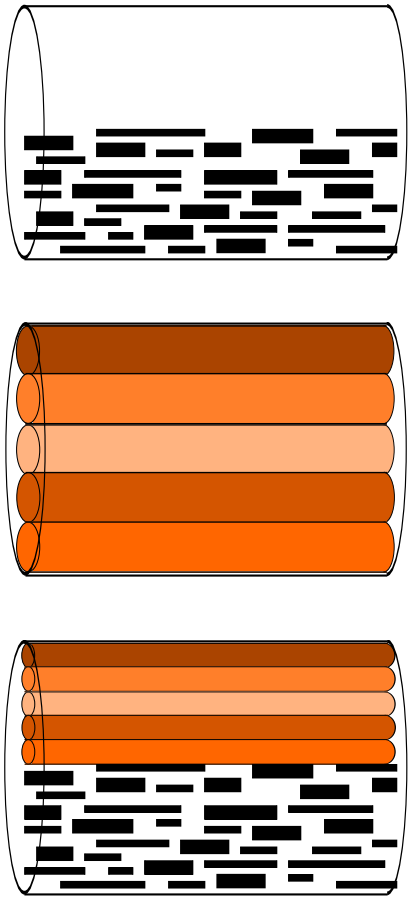
Link utilization regimes

a transparent regime
the sum of the peak rates of the flows is less than capacity with high probability



an elastic regime
all the competing flows share the link bandwidth

Link utilization regimes

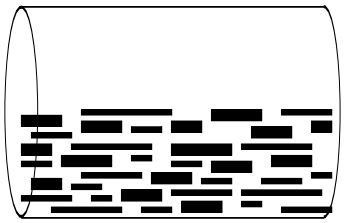


a transparent regime
the sum of the peak rates of the flows is less than capacity with high probability

an elastic regime
all the competing flows share the link bandwidth

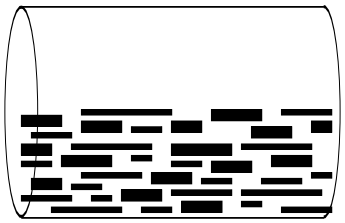
a intermediate elastic regime
the majority of flows are peak rate limited but share the bandwidth with flows using all the residual bandwidth

Buffer sizing in the transparent regime



- Packets arrivals are locally Poisson

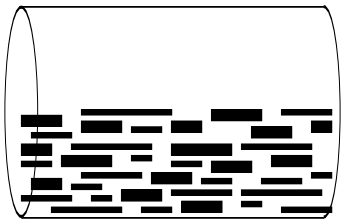
Buffer sizing in the transparent regime



- Packets arrivals are locally Poisson

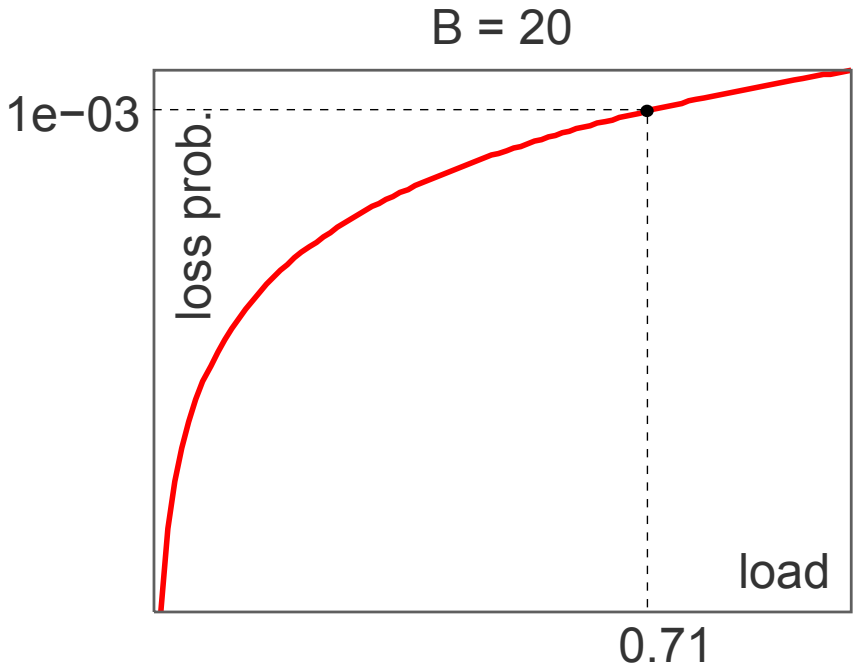
- M/G/1 or even M/M/1 approximation is a useful guideline
- Buffer size B, approximate packet loss probability is $\epsilon = \rho^B$

Buffer sizing in the transparent regime

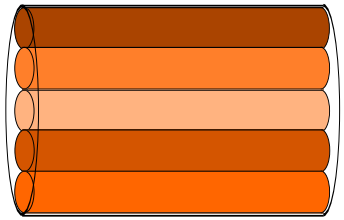


- Packets arrivals are locally Poisson

- M/G/1 or even M/M/1 approximation is a useful guideline
- Buffer size B, approximate packet loss probability is $\epsilon = \rho^B$

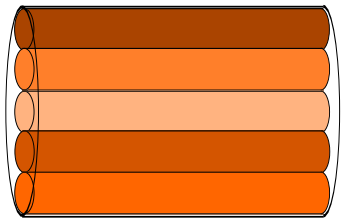


Buffer sizing in the elastic regime



- Processor sharing model illlustrates flow performance
  - Poisson flow arrivals
  - Perfect fair sharing

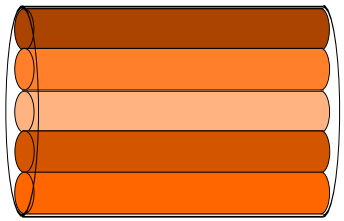
Buffer sizing in the elastic regime



- Processor sharing model illlustrates flow performance
  - Poisson flow arrivals
  - Perfect fair sharing

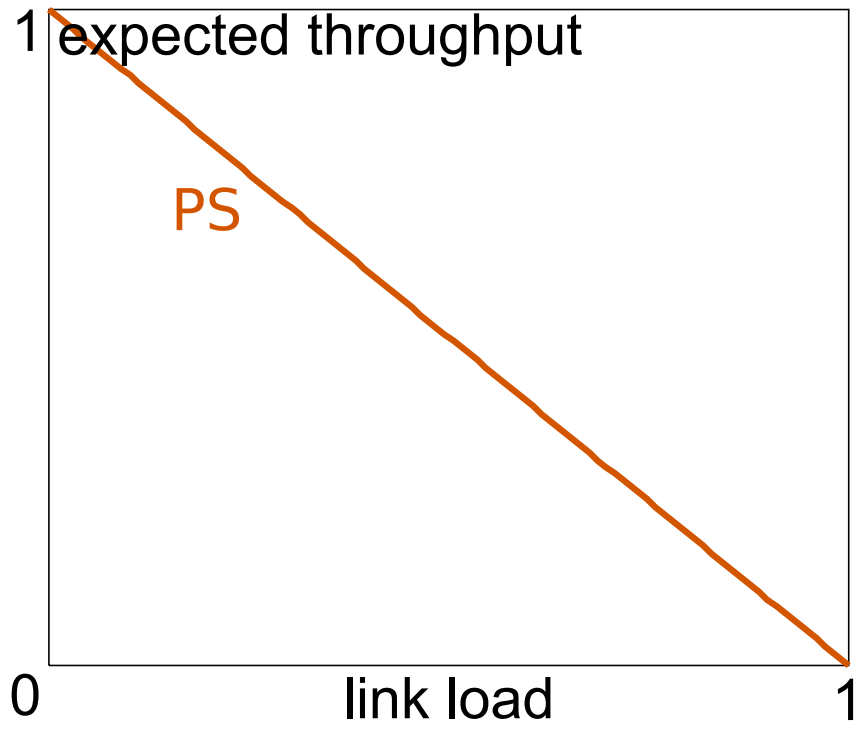
- The number of flows has a geometric distribution (M/M/1)
- Most of the time, only 1 or 2 flows : BDP necessary for standard TCP
- *normalized expected flow throughput* is a convenient performance indicator $= \frac{C}{E[\#\text{flows}]} = (1 - \rho)$

Buffer sizing in the elastic regime



- Processor sharing model illlustrates flow performance
  - Poisson flow arrivals
  - Perfect fair sharing

- The number of flows has a geometric distribution (M/M/1)
- Most of the time, only 1 or 2 flows : BDP necessary for standard TCP
- *normalized expected flow throughput* is a convenient performance indicator $= \frac{C}{E[\#\text{flows}]} = (1-\rho)$



Buffer sizing in the intermediate elastic regime

- Flows with small peak rates behave as a background traffic

Buffer sizing in the intermediate elastic regime

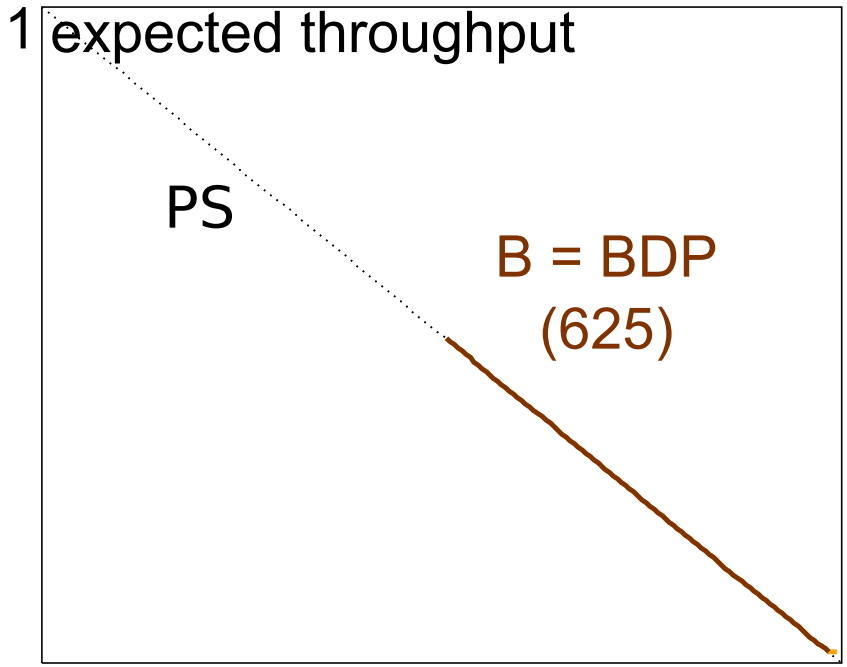
- Flows with small peak rates behave as a background traffic
- TCP does not fully utilizes available bandwidth

Buffer sizing in the intermediate elastic regime

- Flows with small peak rates behave as a background traffic
- TCP does not fully utilizes available bandwidth
- $\phi(N)$ equals utilization of residual bandwidth realized by N TCP

Buffer sizing in the intermediate elastic regime

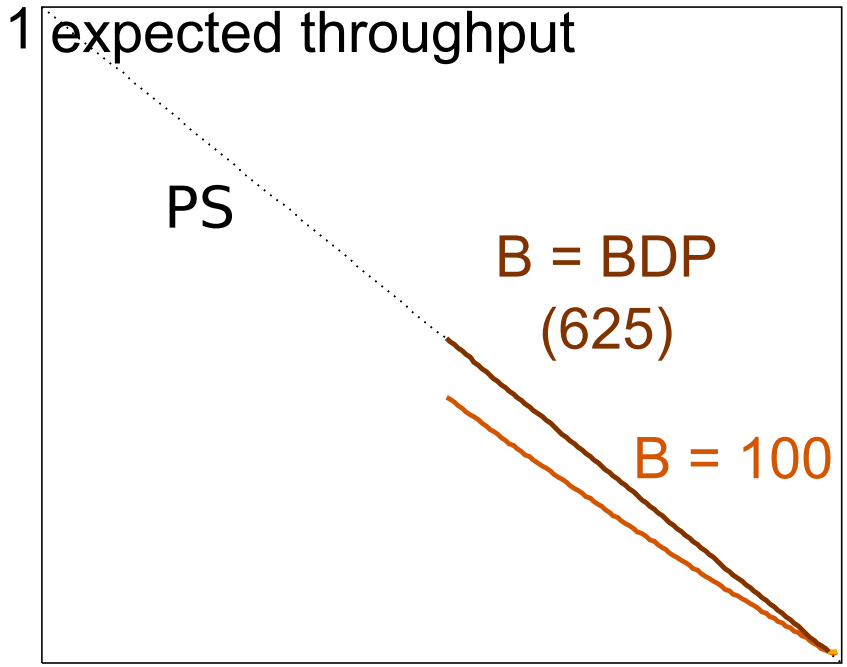
- Flows with small peak rates behave as a background traffic
- TCP does not fully utilizes available bandwidth
- $\phi(N)$ equals utilization of residual bandwidth realized by N TCP
- Evaluation of $\phi(N)$ values with simulations flows



- background load = 50%

Buffer sizing in the intermediate elastic regime

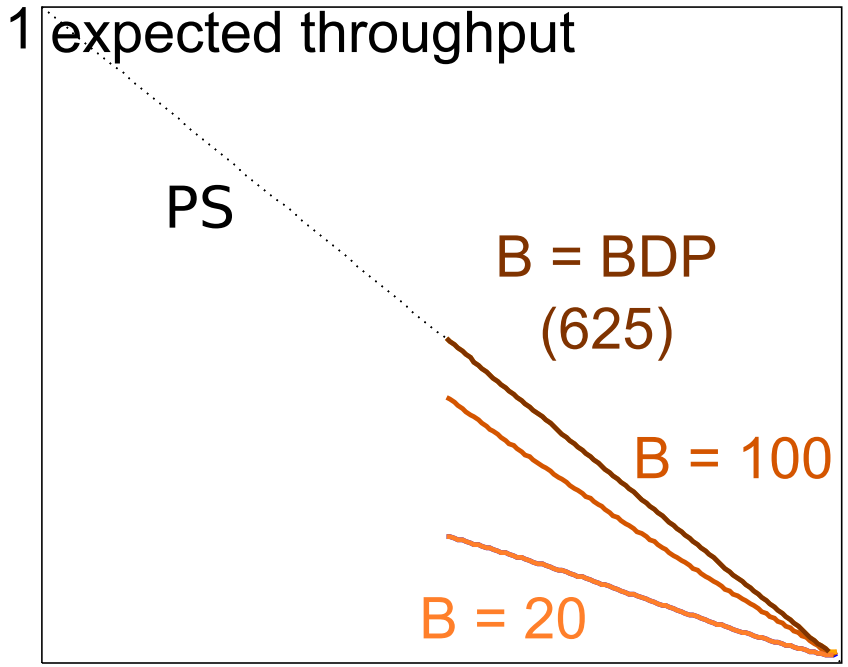
- Flows with small peak rates behave as a background traffic
- TCP does not fully utilizes available bandwidth
- $\phi(N)$ equals utilization of residual bandwidth realized by N TCP
- Evaluation of $\phi(N)$ values with simulations flows



- background load = 50%

Buffer sizing in the intermediate elastic regime

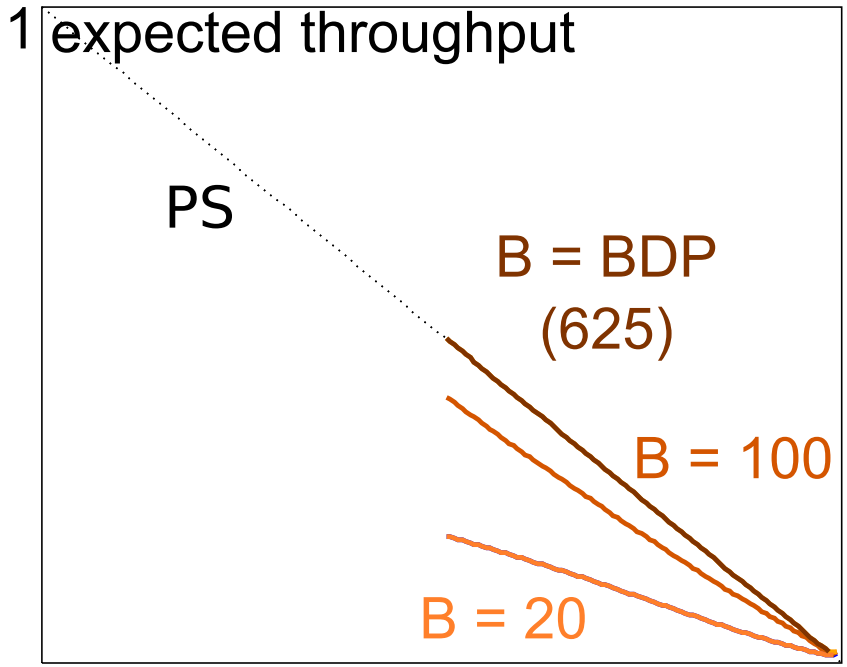
- Flows with small peak rates behave as a background traffic
- TCP does not fully utilizes available bandwidth
- $\phi(N)$ equals utilization of residual bandwidth realized by N TCP
- Evaluation of $\phi(N)$ values with simulations flows



- background load = 50%

Buffer sizing in the intermediate elastic regime

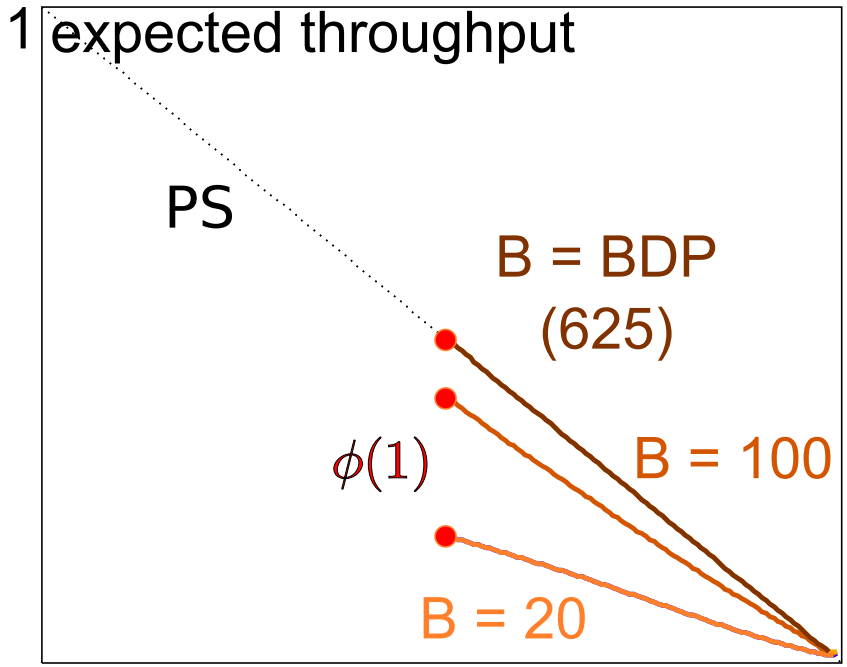
- Flows with small peak rates behave as a background traffic
- TCP does not fully utilizes available bandwidth
- $\phi(N)$ equals utilization of residual bandwidth realized by N TCP
- Evaluation of $\phi(N)$ values with simulations flows



- background load = 50%
- performance decreases with higher background load or higher link capacity

Buffer sizing in the intermediate elastic regime

- Flows with small peak rates behave as a background traffic
- TCP does not fully utilizes available bandwidth
- $\phi(N)$ equals utilization of residual bandwidth realized by N TCP
- Evaluation of $\phi(N)$ values with simulations flows

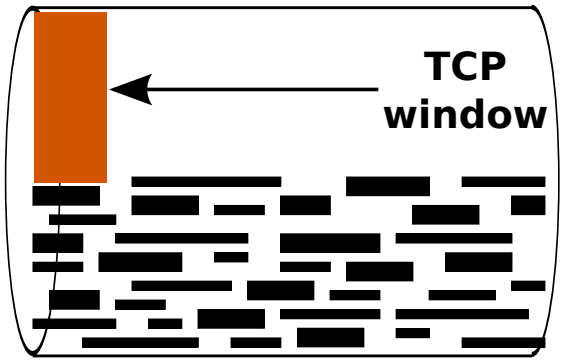


- background load $= 50\%$
- performance decreases with higher background load or higher link capacity
- linear decrease of throughput conditioned by $\phi(1)$

TCP performance of one foreground flow

- TCP packet emissions are controlled by the congestion window

TCP performance of one foreground flow

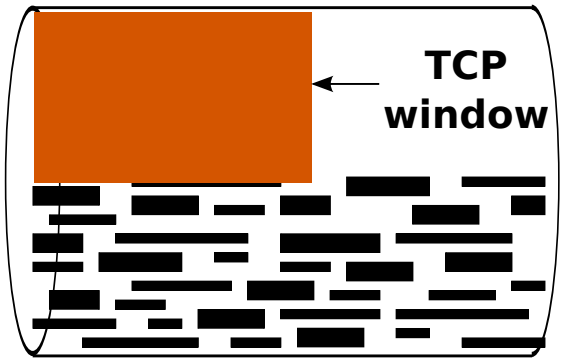


- TCP packet emissions are controlled by the congestion window

TCP performance of one foreground flow

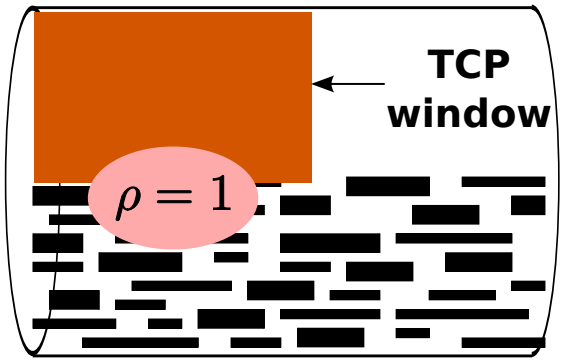
- TCP packet emissions are controlled by the congestion window

TCP performance of one foreground flow



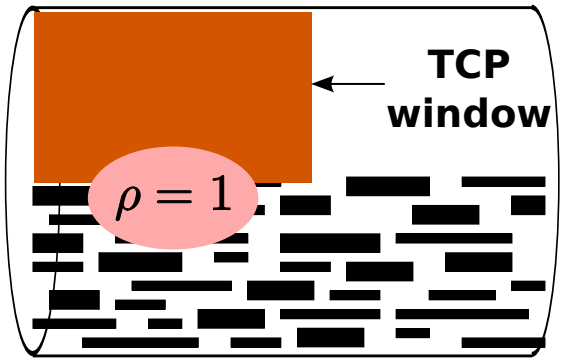
- TCP packet emissions are controlled by the congestion window

TCP performance of one foreground flow



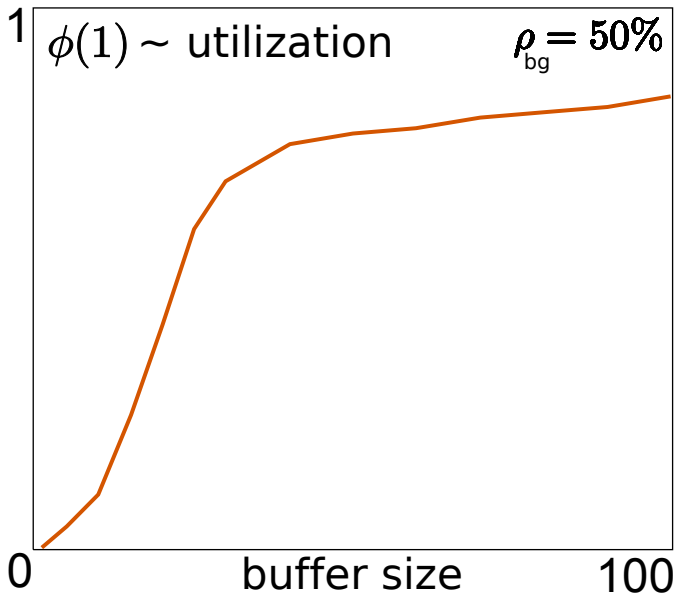
- TCP packet emissions are controlled by the congestion window
- Load is close to 1 during windows emissions
- Unstable behaviour of the queue which fills up, leading to a packet loss and halving of TCP window

TCP performance of one foreground flow



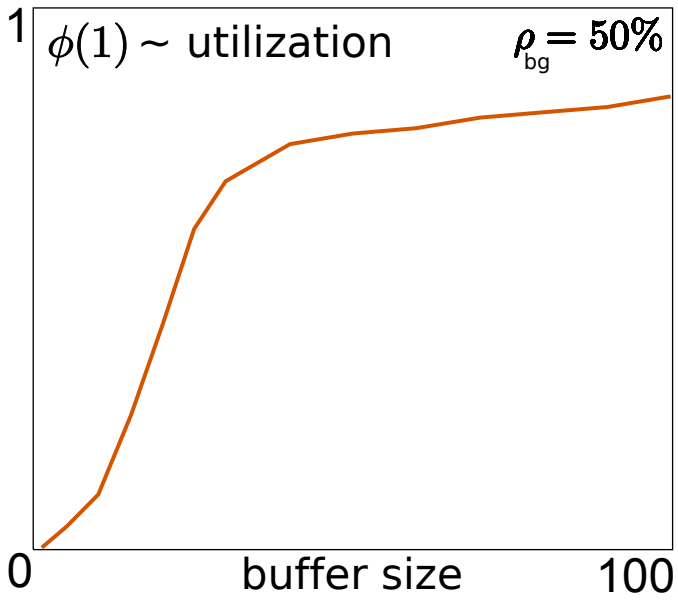
- TCP packet emissions are controlled by the congestion window
- Load is close to 1 during windows emissions
- Unstable behaviour of the queue which fills up, leading to a packet loss and halving of TCP window
- Large buffers are necessary to avoid the impact of the background traffic

Empirical buffer sizing (1/2)



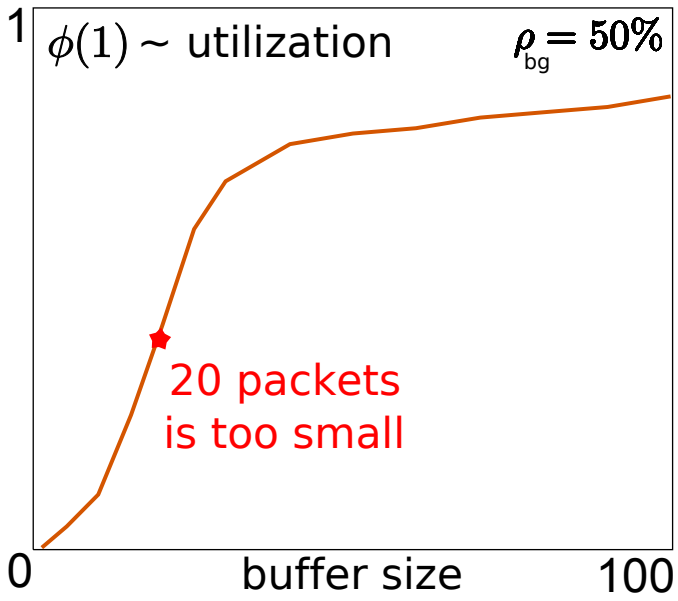
- fixed background load

Empirical buffer sizing (1/2)



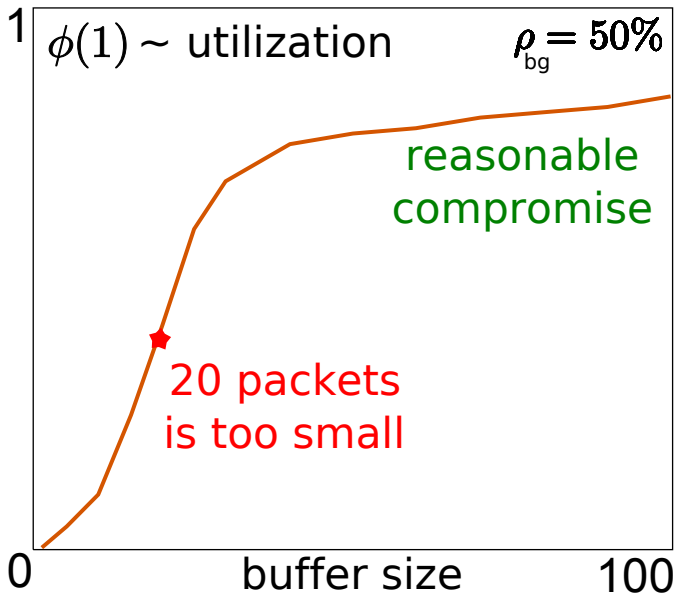
- fixed background load
- Two behaviors for $\phi(1)$ according to buffer size
  - for small buffer sizes, low values of $\phi(1)$
  - for higher buffer sizes, $\phi(1)$ tend to 100%, reached for the BDP

Empirical buffer sizing (1/2)



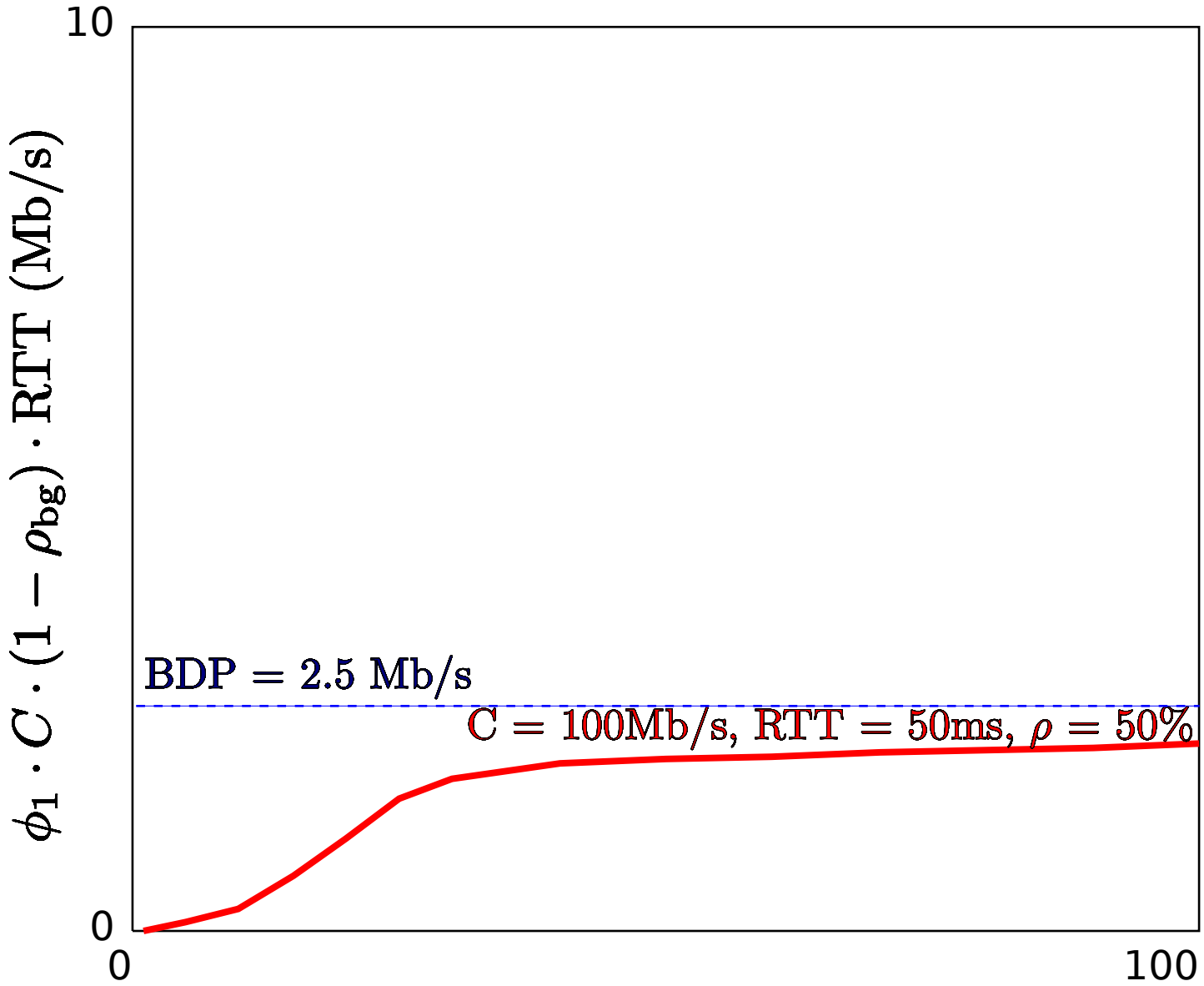
- fixed background load
- Two behaviors for $\phi(1)$ according to buffer size
  - for small buffer sizes, low values of $\phi(1)$
  - for higher buffer sizes, $\phi(1)$ tend to 100%, reached for the BDP

Empirical buffer sizing (1/2)

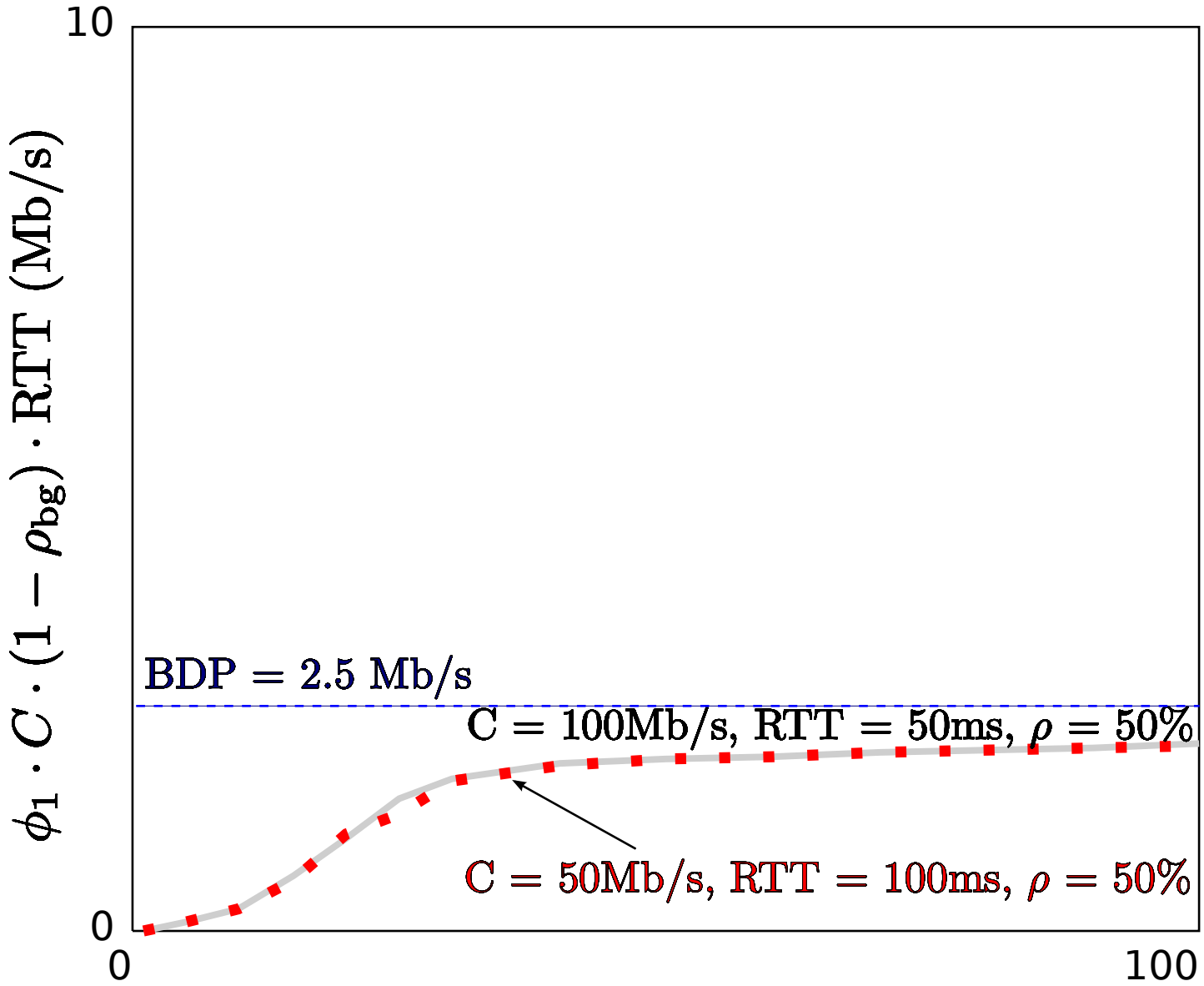


- fixed background load
- Two behaviors for $\phi(1)$ according to buffer size
  - for small buffer sizes, low values of $\phi(1)$
  - for higher buffer sizes, $\phi(1)$ tend to 100%, reached for the BDP
- buffers should be sized to avoid the low utilization zone

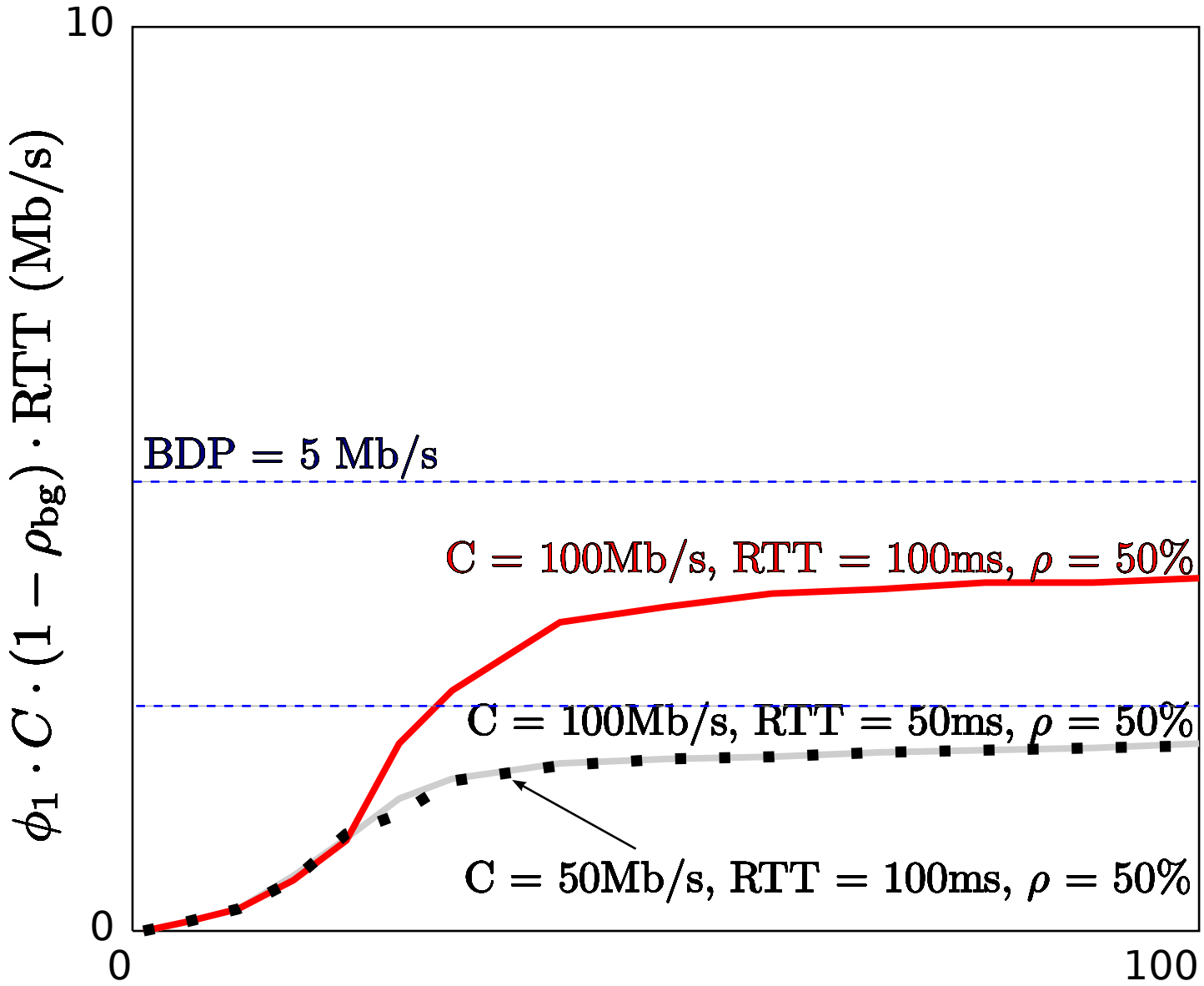
Empirical buffer sizing (2/2)



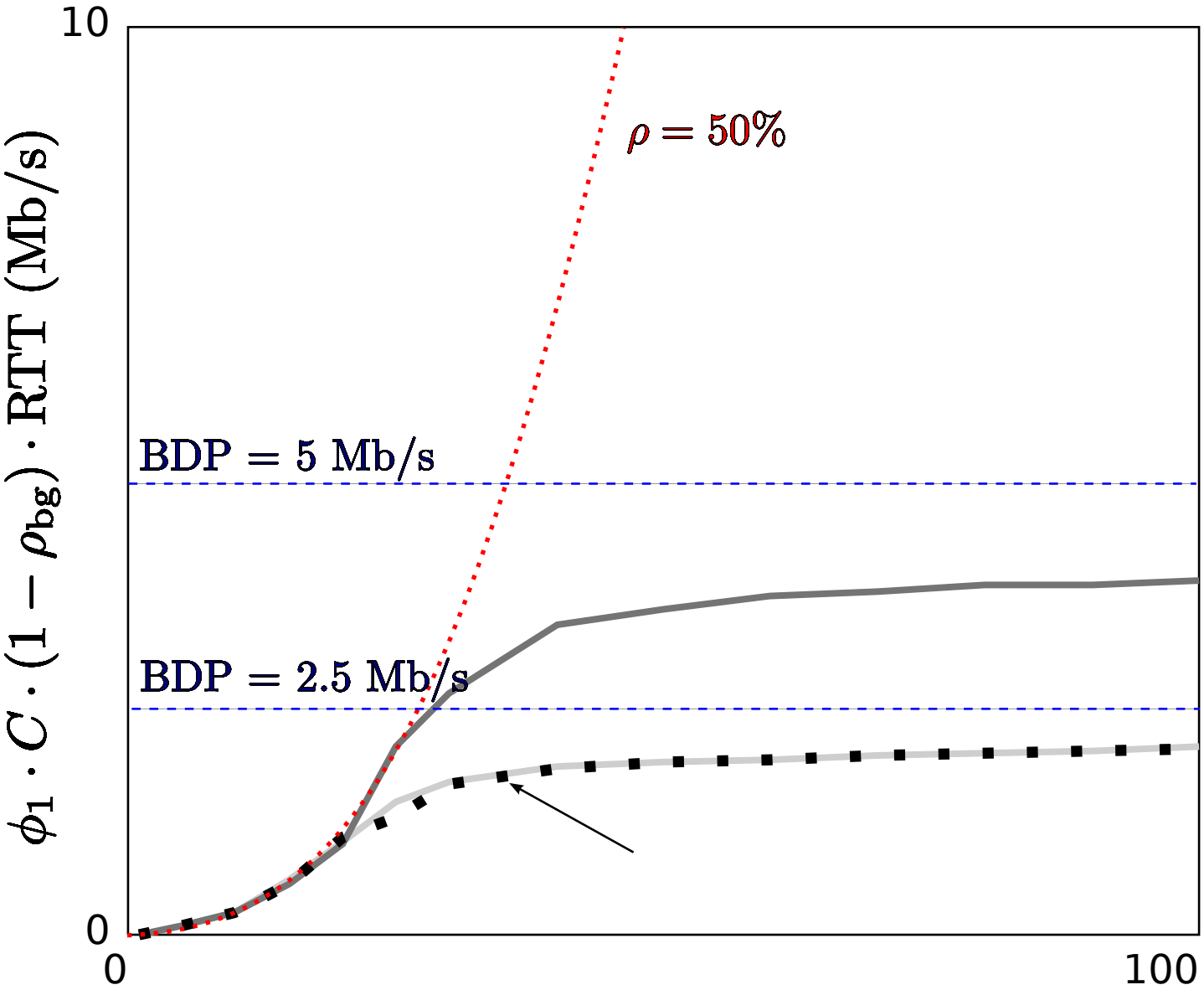
Empirical buffer sizing (2/2)



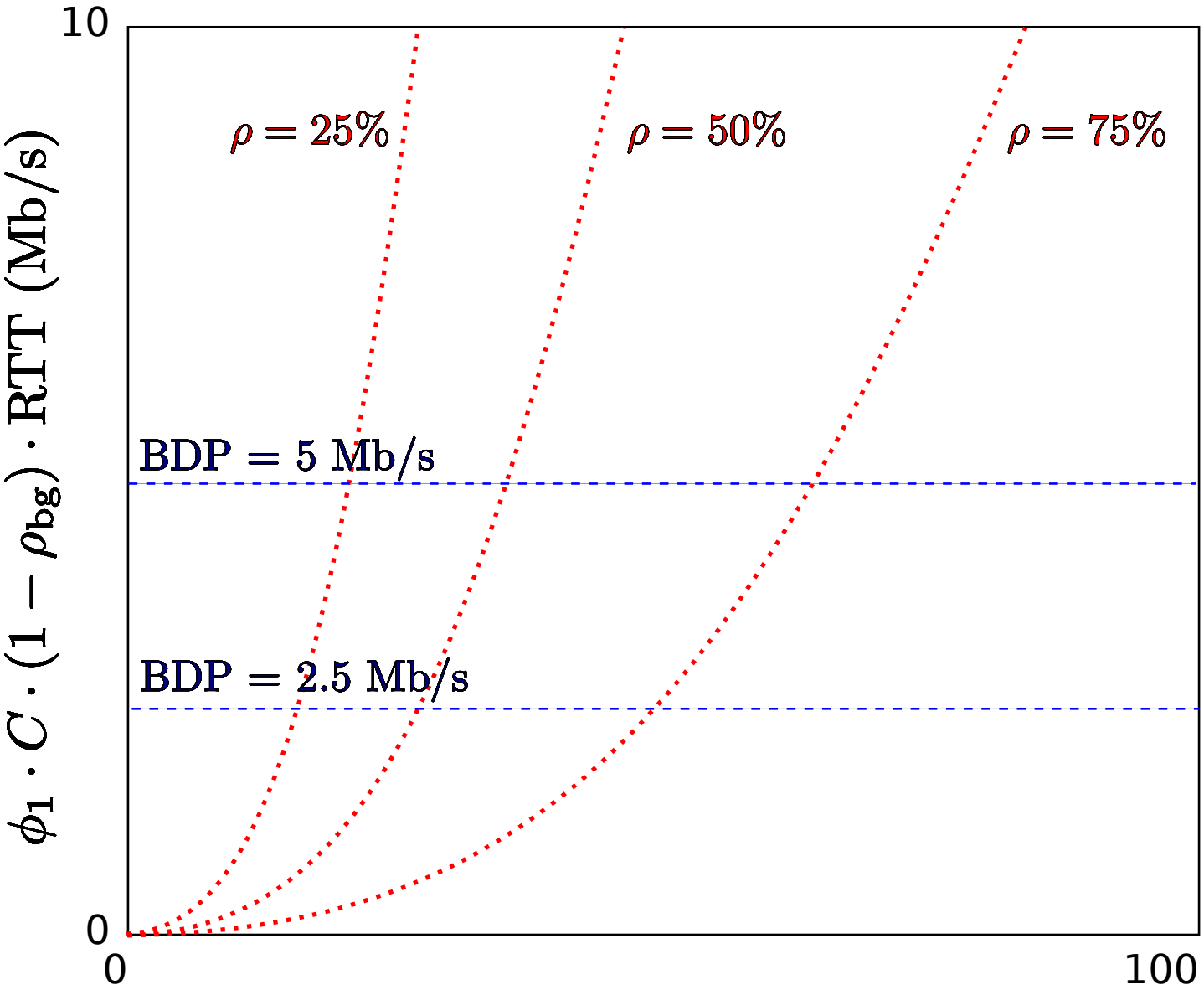
Empirical buffer sizing (2/2)



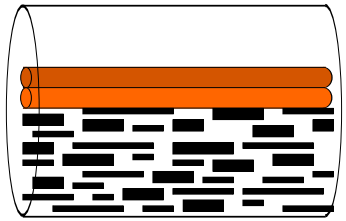
Empirical buffer sizing (2/2)

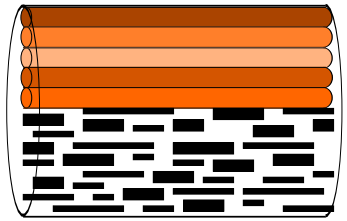


Empirical buffer sizing (2/2)



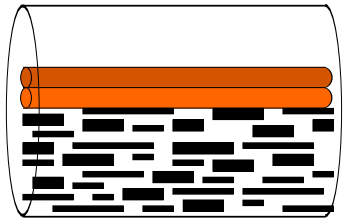
Flows with limited peak rates

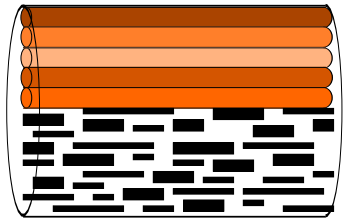




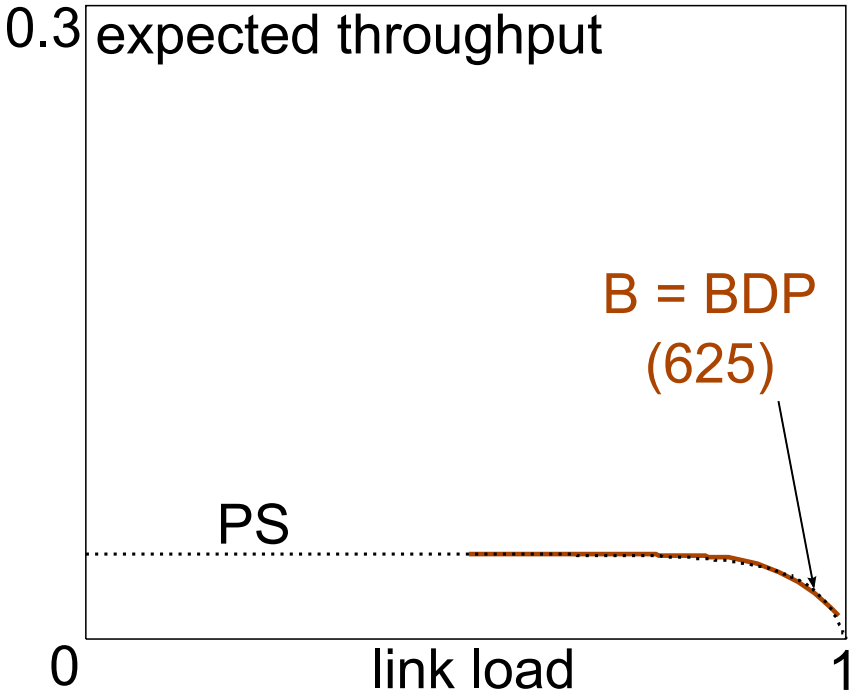
- Several flows are needed to saturate the link
- We are in a transparent regime up to high loads

Flows with limited peak rates

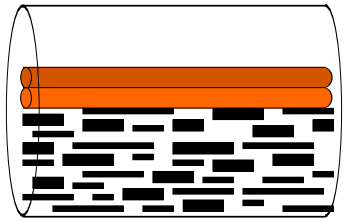


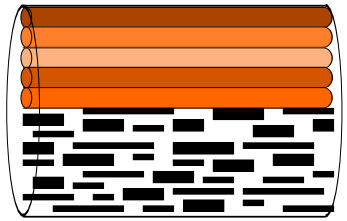


- Several flows are needed to saturate the link
- We are in a transparent regime up to high loads

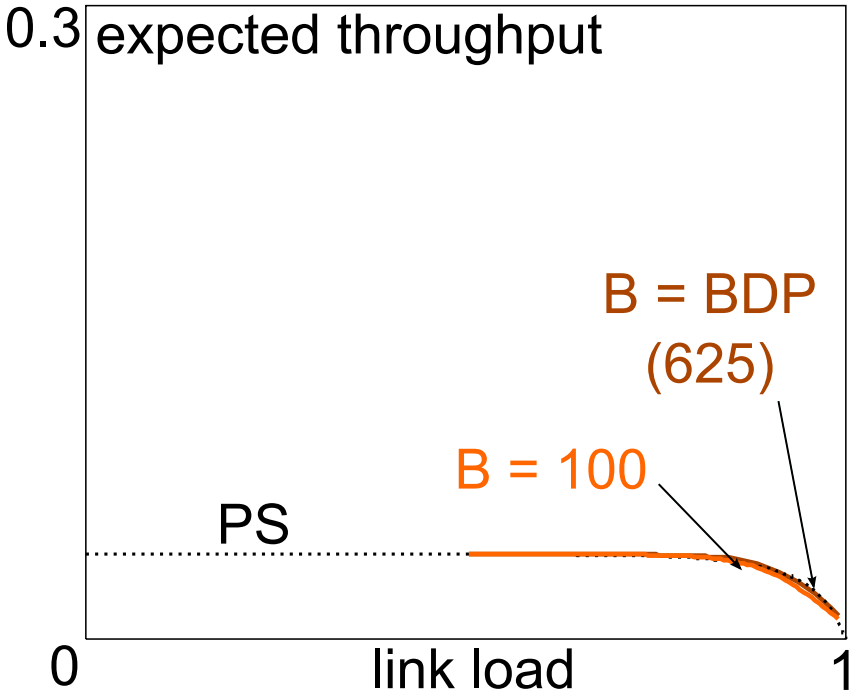


Flows with limited peak rates

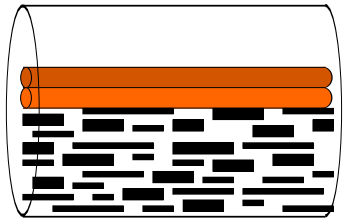


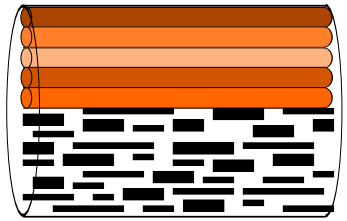


- Several flows are needed to saturate the link
- We are in a transparent regime up to high loads

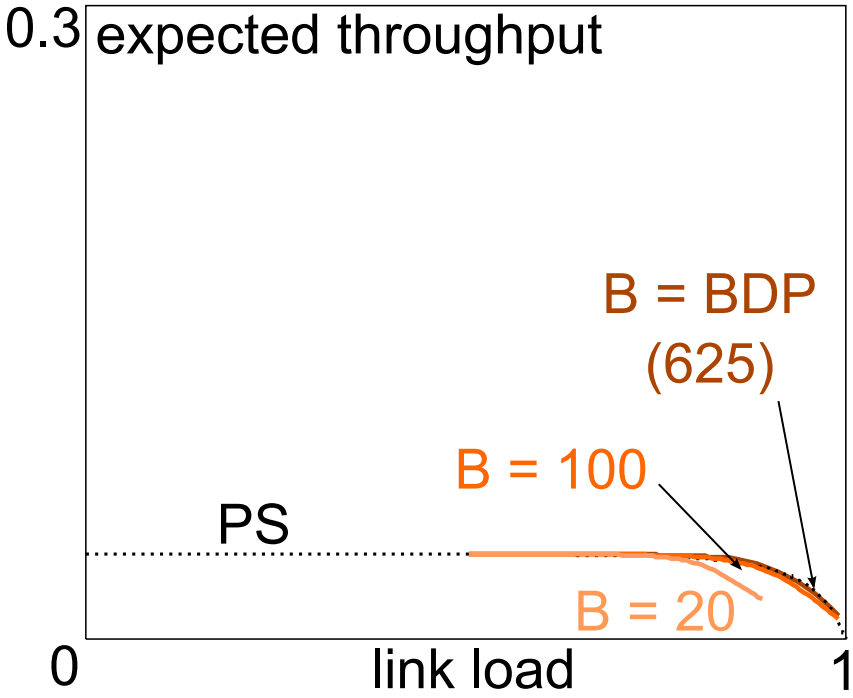


Flows with limited peak rates

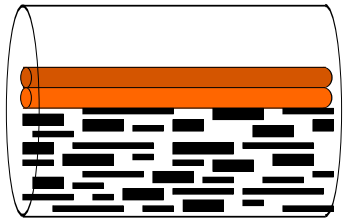


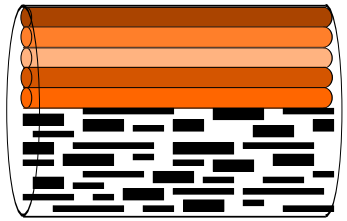


- Several flows are needed to saturate the link
- We are in a transparent regime up to high loads

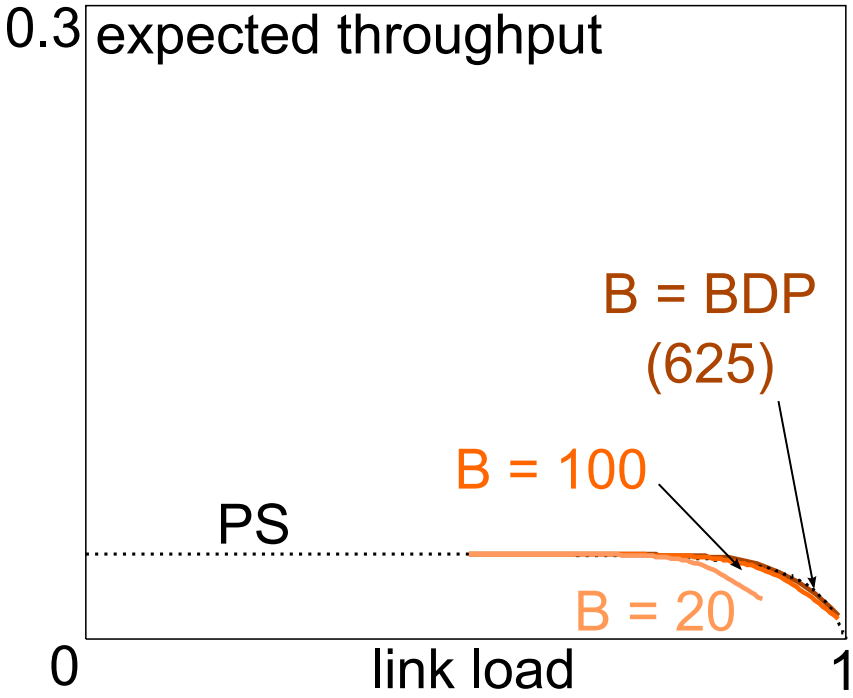


Flows with limited peak rates





- Several flows are needed to saturate the link
- We are in a transparent regime up to high loads



- Small buffers are sufficient up to high loads

Conlusion

- Relation between buffer size and realized performance clearly depends on assumed traffic characteristics
- The most significant is the mix of flow peak rates

Conlusion

- Relation between buffer size and realized performance clearly depends on assumed traffic characteristics
- The most significant is the mix of flow peak rates
- 20 packet suffice in transparent regime. . .

Conlusion

- Relation between buffer size and realized performance clearly depends on assumed traffic characteristics
- The most significant is the mix of flow peak rates
- 20 packet suffice in transparent regime...
- ...but are too few for high peak rate flows

Conlusion

- Relation between buffer size and realized performance clearly depends on assumed traffic characteristics
- The most significant is the mix of flow peak rates
- 20 packet suffice in transparent regime...
- ...but are too few for high peak rate flows
- Though the BDP is not necessary

Conlusion

- Relation between buffer size and realized performance clearly depends on assumed traffic characteristics
- The most significant is the mix of flow peak rates
- 20 packet suffice in transparent regime...
- ...but are too few for high peak rate flows
- Though the BDP is not necessary
- We proposed an empirical buffer sizing guideline depending on background load

Conlusion

- Relation between buffer size and realized performance clearly depends on assumed traffic characteristics
- The most significant is the mix of flow peak rates
- 20 packet suffice in transparent regime...
- ...but are too few for high peak rate flows
- Though the BDP is not necessary
- We proposed an empirical buffer sizing guideline depending on background load

---

Thanks ! Questions are welcome.

Contact : Jordan Augé <jordan.auge@free.fr>