Stochastic Network Calculus

From Theory to Applications

Jörg Liebeherr

University of Toronto

(Tutorial ITC)

Roadmap

- Goal: Attempt a coherent self-contained story anchored in a few important questions:
  - Analysis of “real” traffic
  - Analysis of buffered link
  - Scaling of performance metrics (in size of network)
  - Quality of upper and lower bounds
  - Statistical Multiplexing
- Goal: No prerequisites (on the subject)
- Goal: Discuss interesting applications and case studies
- Non-goal: Comprehensive literature survey

Outline

1. Network Traffic Models
2. A Network Calculus Primer
   - Buffered links (+ min-plus algebra)
   - Envelopes and service curves
   - Performance bounds
   - Min-plus linear system theory
   - Case study: Bandwidth estimation
3. Stochastic Network Calculus
   - Statistical multiplexing gain
   - Statistical envelopes and sample path envelopes
   - Statistical end-to-end analysis
   - Assessment of the state-of-the-art (beyond hype and myths)
4. Application areas
   - Wireless networks
   - Data centers
   - Smart grids

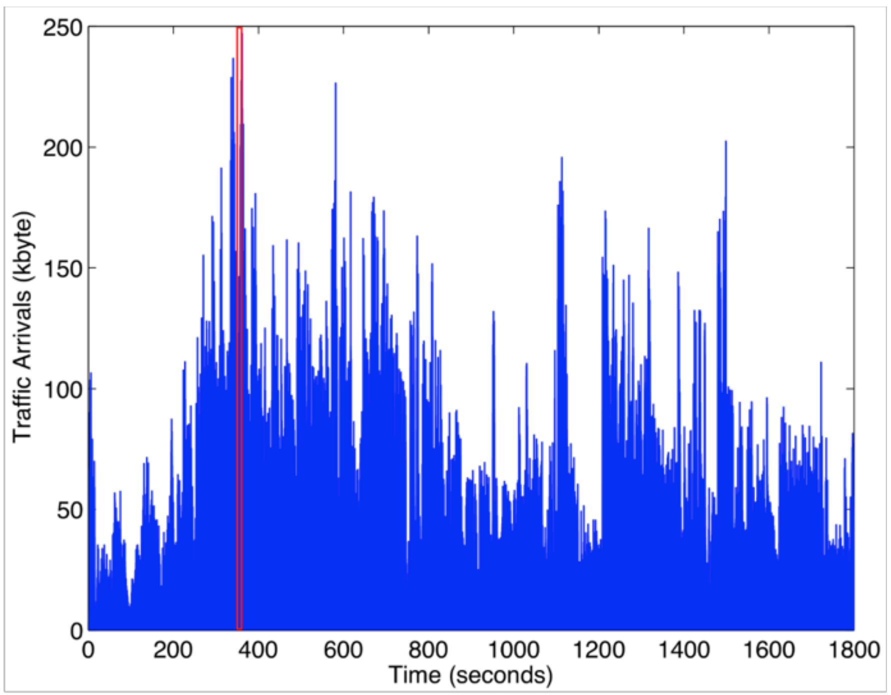
MODELLING NETWORK TRAFFIC

Network Traffic

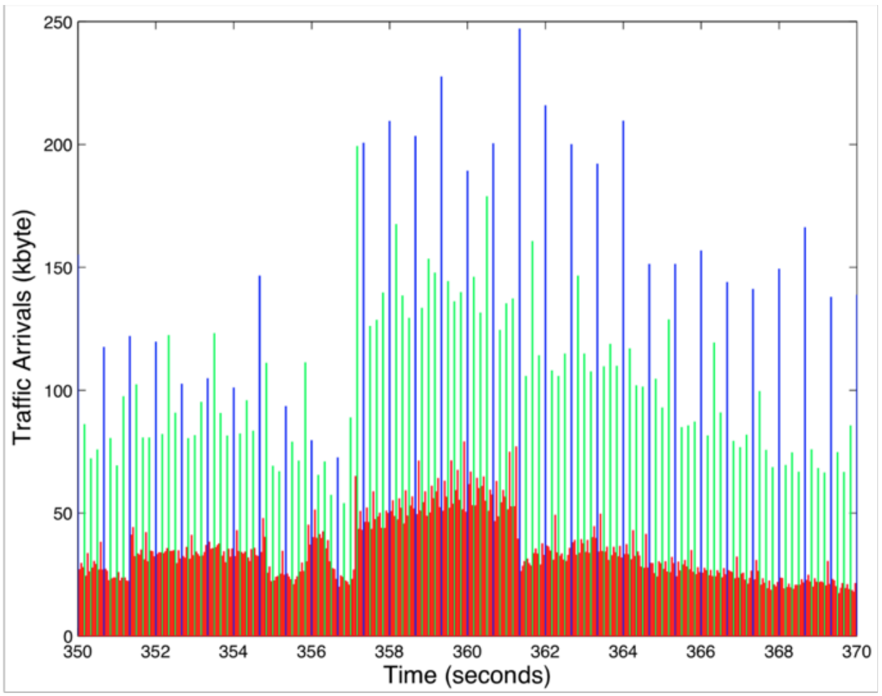
HD Movie

Harry Potter movie with HD encoding

- Codec: H.264 SVC
- Frames per second: 24 fps @ 1920x1088



30 minutes

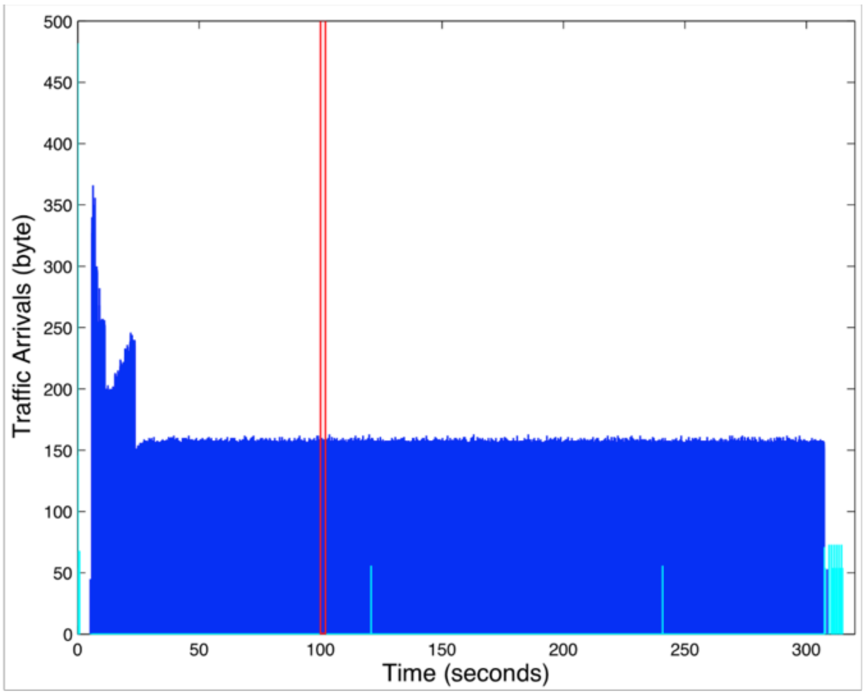


20 seconds

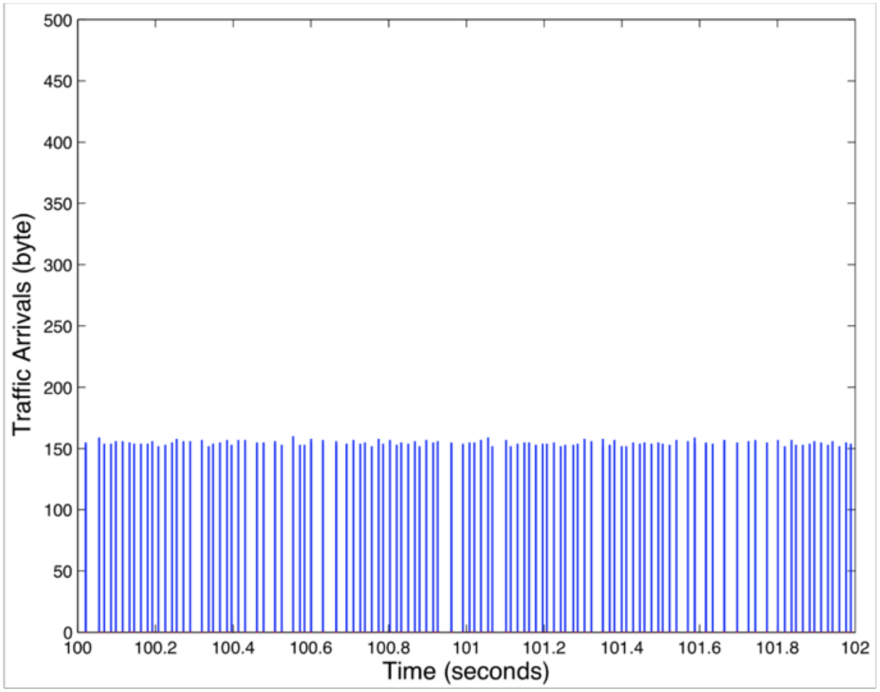
Skype Voice Call: 6 minutes

- SVOPC encoding, one direction of 2-way call

Dark blue: UDP traffic
Light blue: TCP traffic



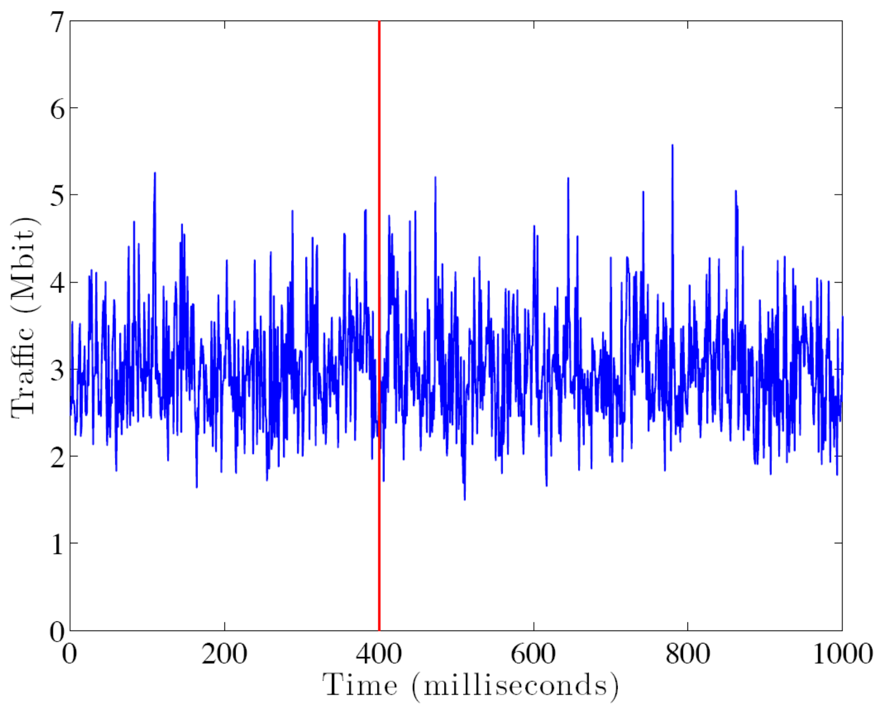
6 minutes



2 seconds

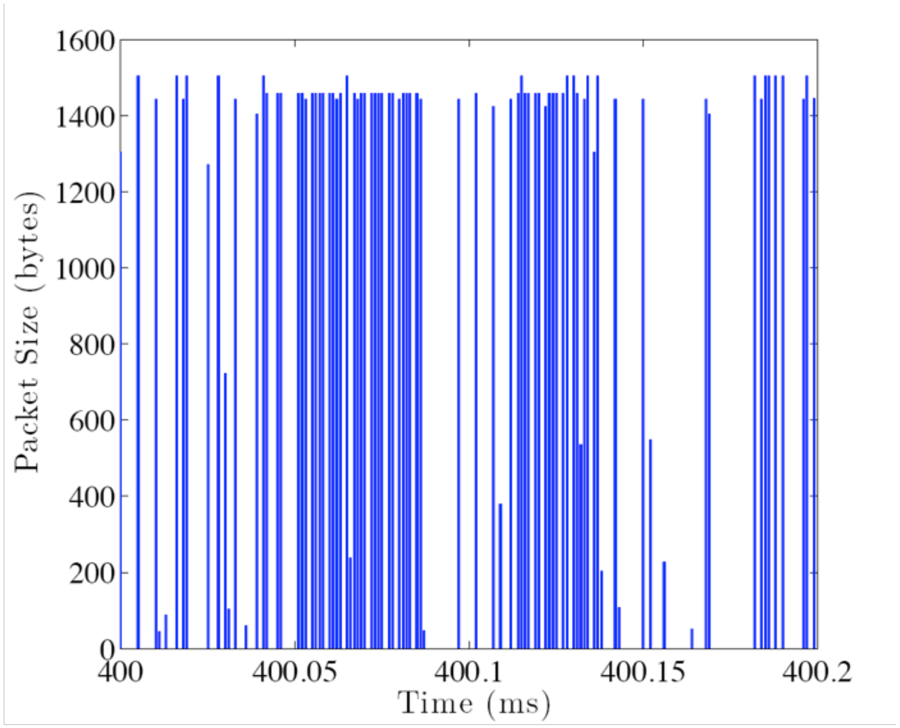
Internet Traffic: 10 Gbps link

- Backbone link of a Tier-1 Internet Service provider
- ~430,000 packets packets in 1 second



1 second

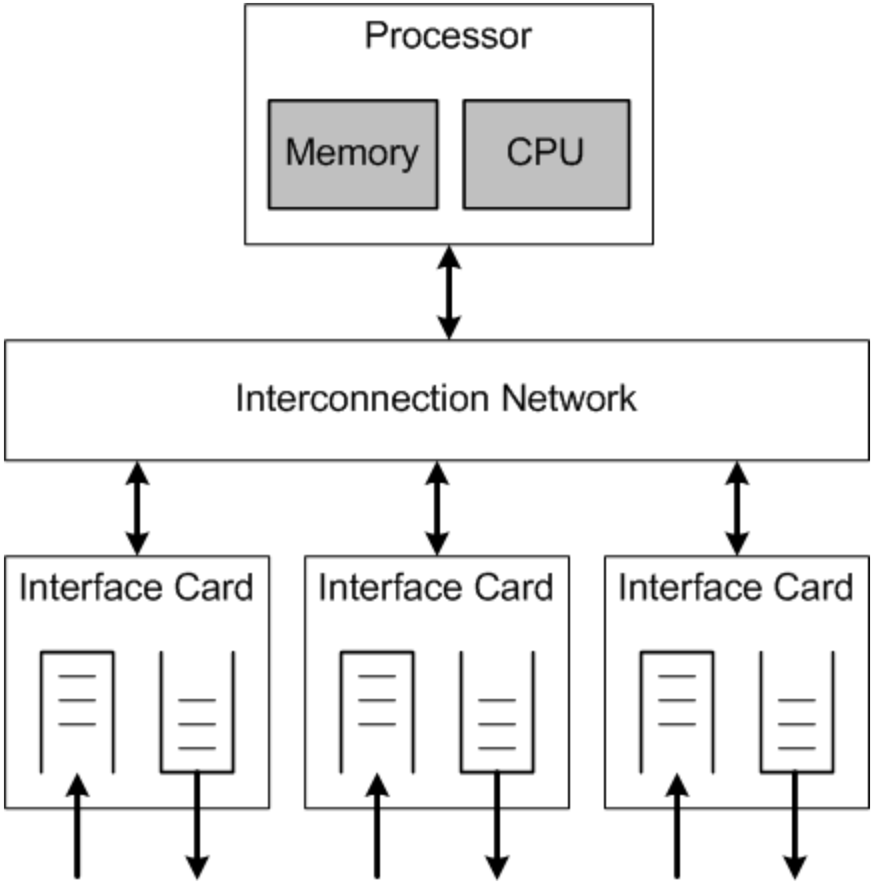
(One data point is the traffic in one millisecond)



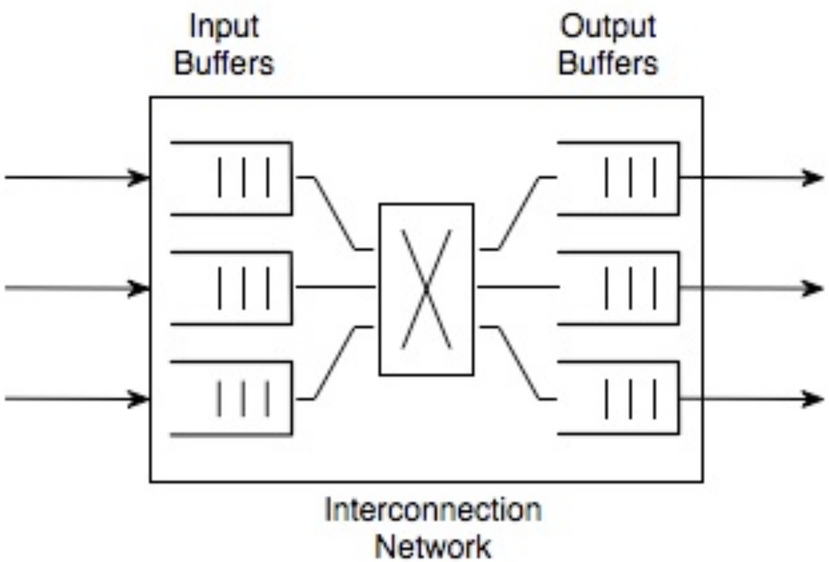
2 μs

Switch Model

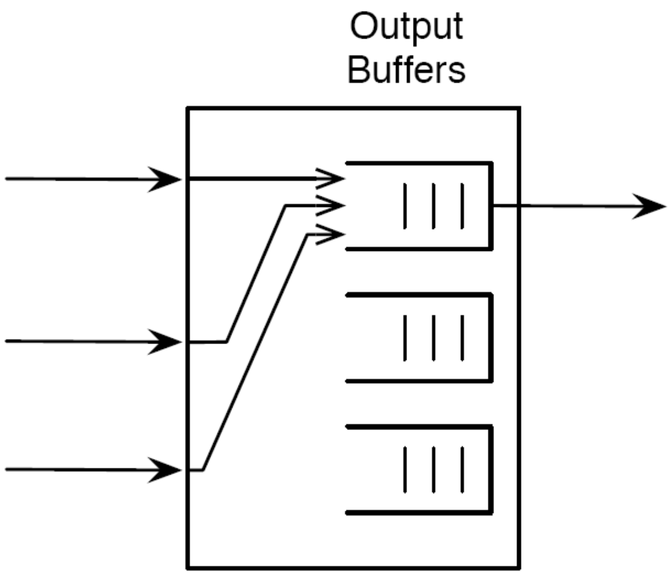
Components of a Packet Switch



Modeling a Packet Switch

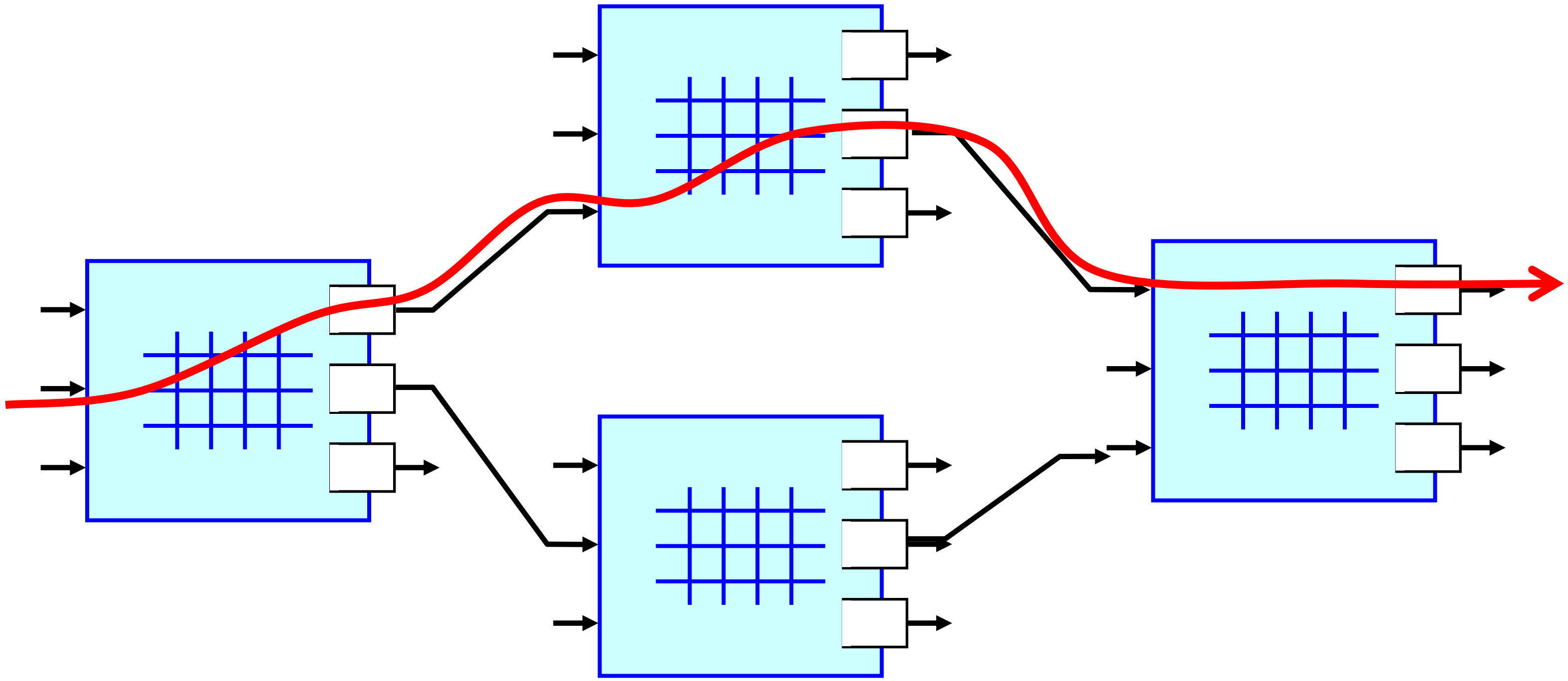


Model with input and
output buffers



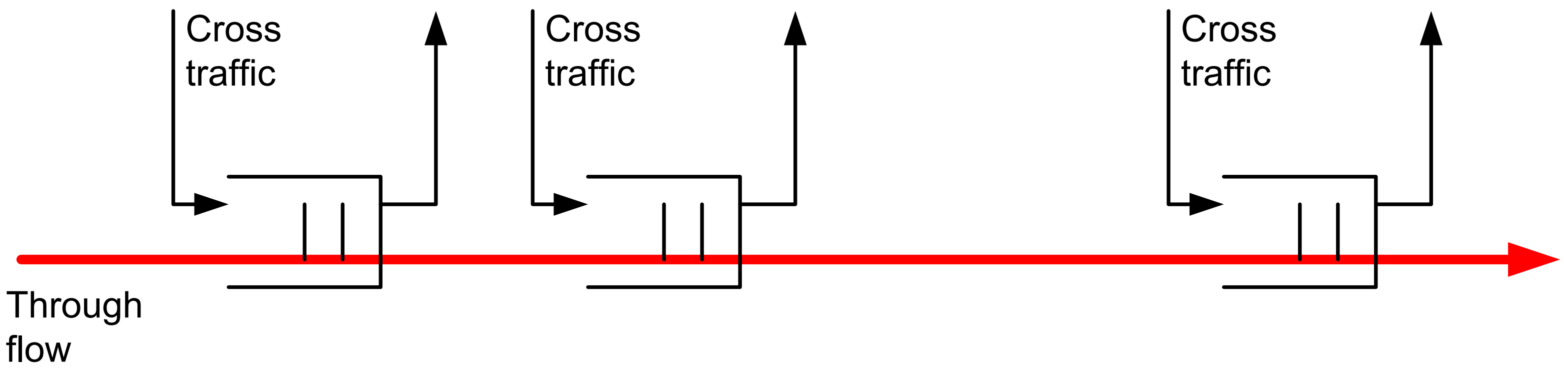
Simplified model
(only output buffers)

A Path of Network Switches



- Each switch is modeled by the traversed output buffer
- Output buffer is shared with other traffic ("cross traffic")

A Path of Network Switches

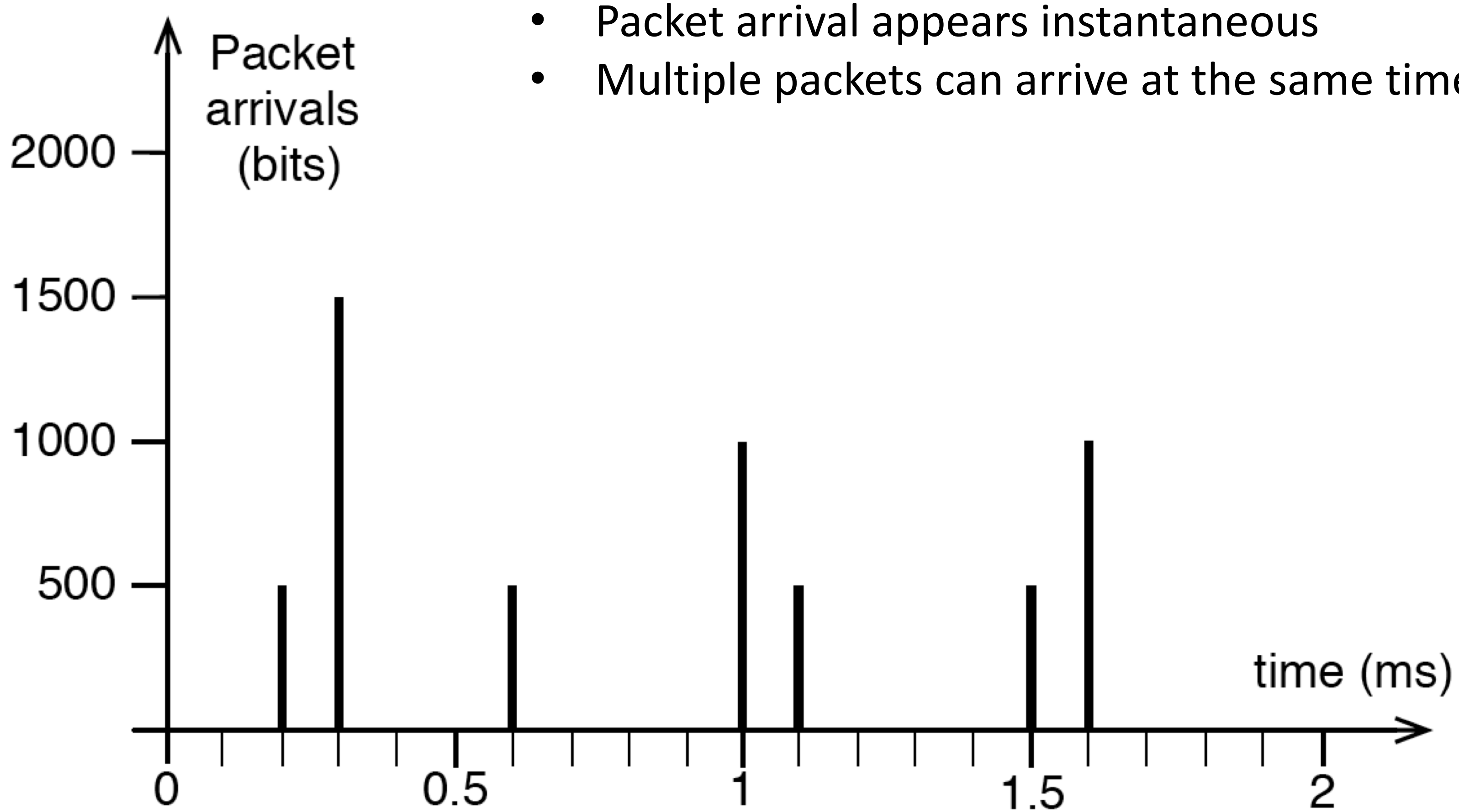


- Each switch is modeled by the traversed output buffer
- Output buffer is shared with other traffic ("cross traffic")

Arrival Model

Arrival Scenarios

- Packet arrives to buffer only after all bits of the packet have been received
- Packet arrival appears instantaneous
- Multiple packets can arrive at the same time



Modeling Traffic Arrivals

- We write arrivals as functions of time:

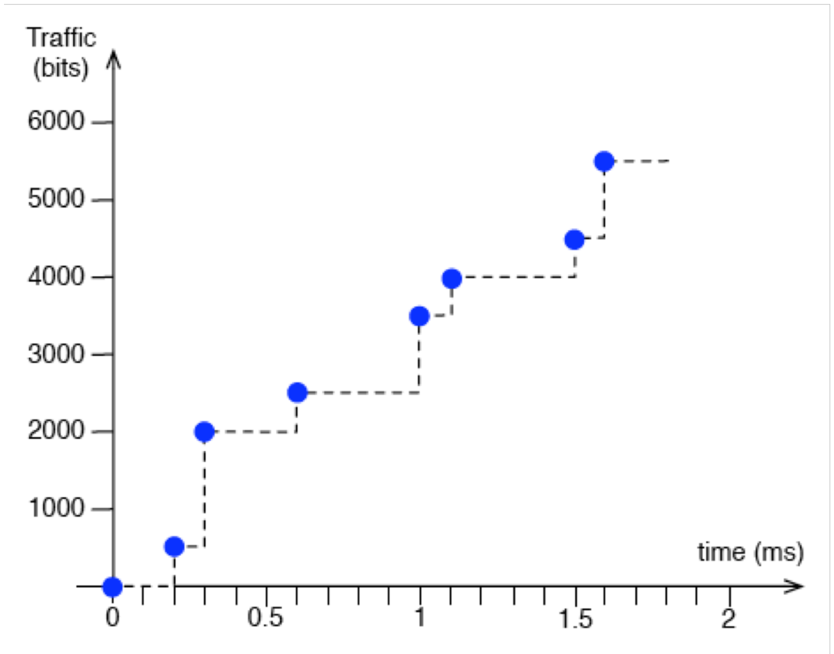
  A(t) : Arrivals until time t, measured in bits

  We assume that A(0)=0 for $t \leq 0$.

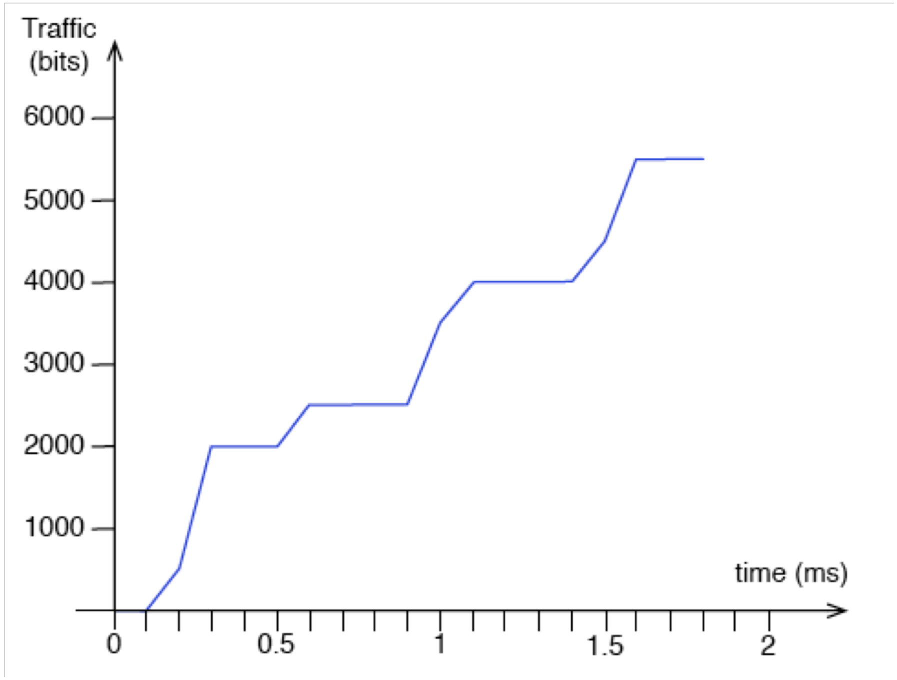
- To plot the arrival function, there are a number of choices:
  1. **Continuous time or Discrete time domain**
  2. **Discrete sized or fluid flow traffic**
  3. **Left-continuous or right-continuous**

Note: The presentation of stochastic network calculus uses discrete-time for conciseness , and otherwise a (left-) continuous time domain

Discrete-sized vs. Fluid Flow



Traffic arrives
in multiples of bits
(discrete time, discrete-sized)

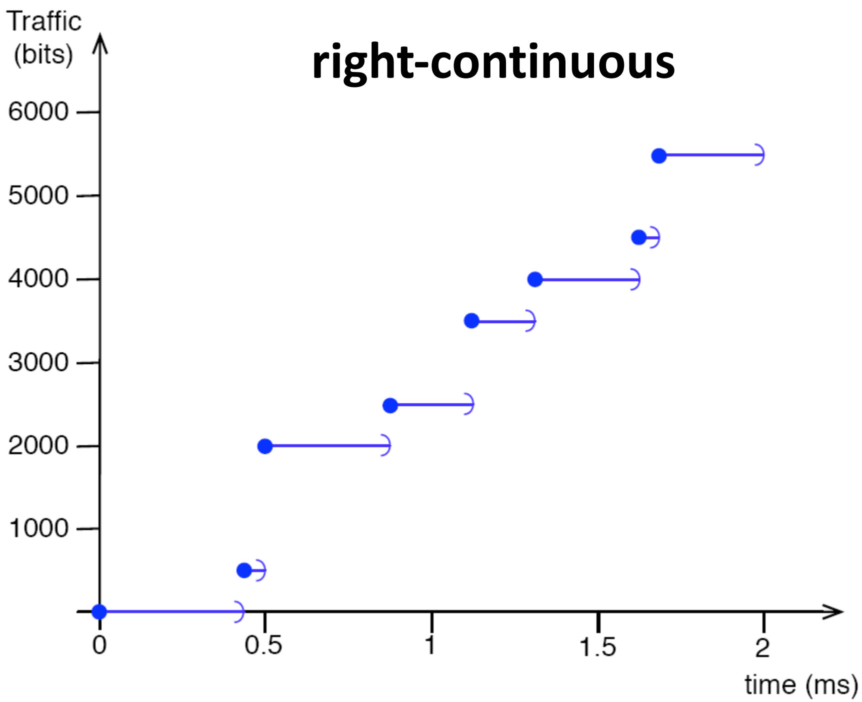


Traffic arrives like a fluid
(continuous time, fluid flow)

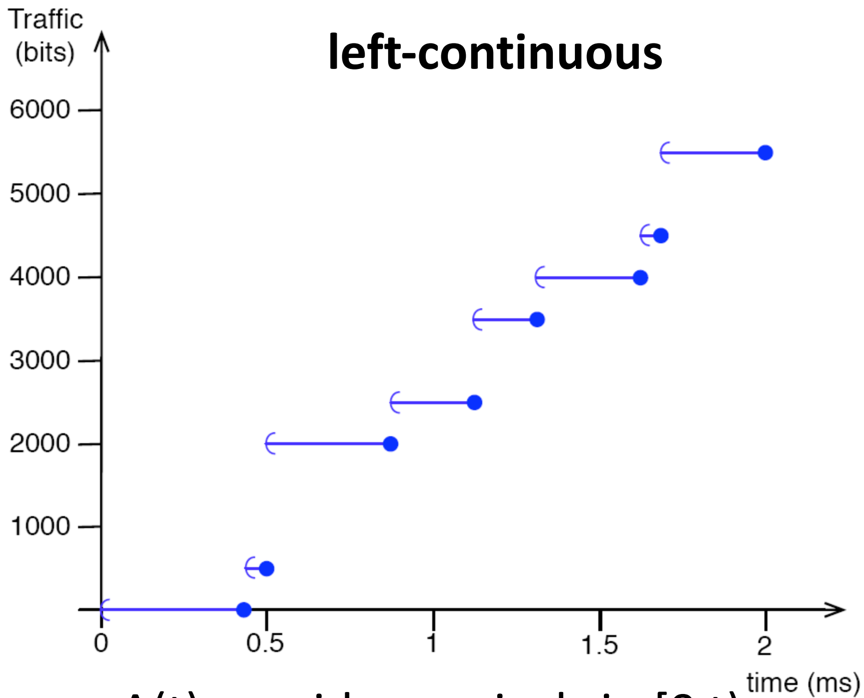
- It is often most convenient to view traffic as a fluid flow, that allows discrete sized bursts

Fluid Flow Traffic with Bursts

- Does A(t) include the arrival at time t, or not ?



A(t) considers arrivals in (0,t]



A(t) considers arrivals in [0,t)

(Note: A(0) = 0 !)

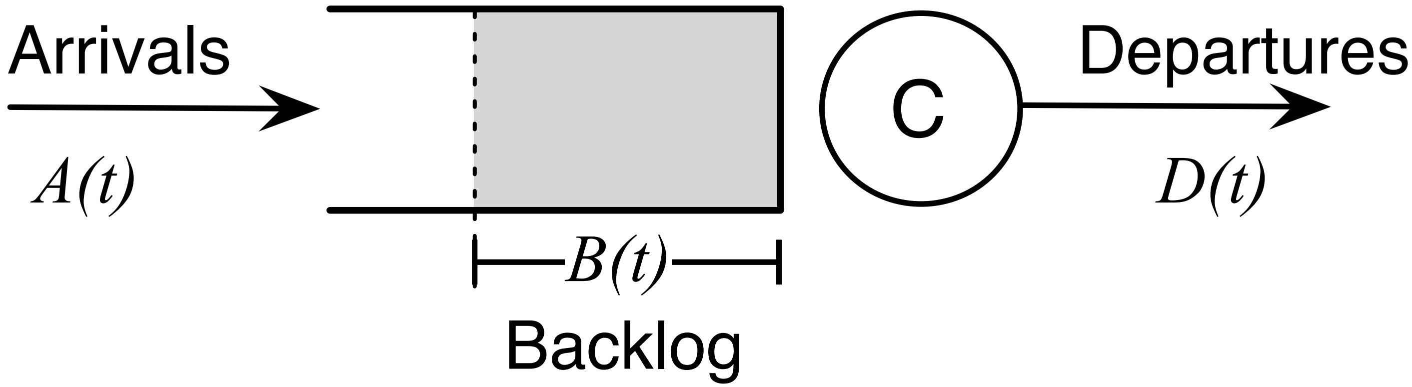
- Most people take a left-continuous interpretation

NETWORK CALCULUS PRIMER

Buffered Link

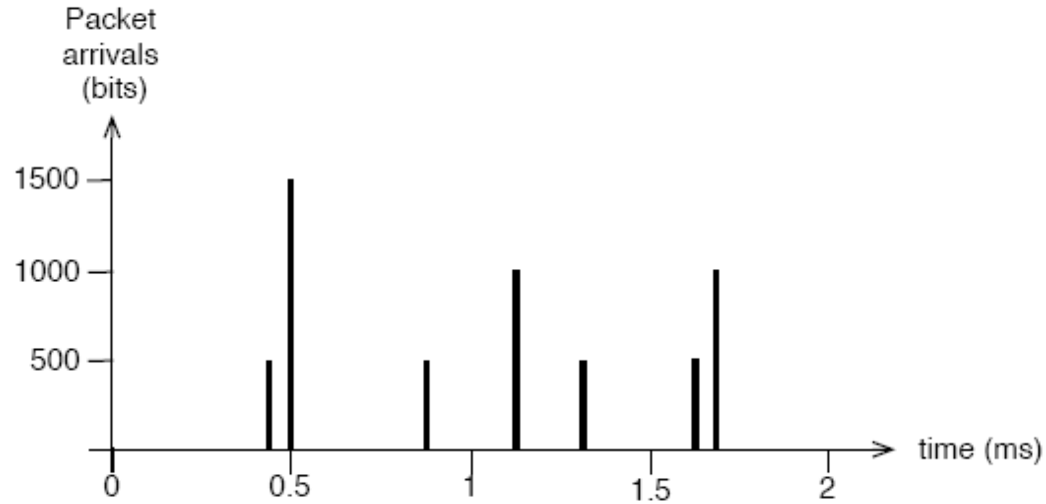
Buffered Link

- Standard model of an output buffer at a switch

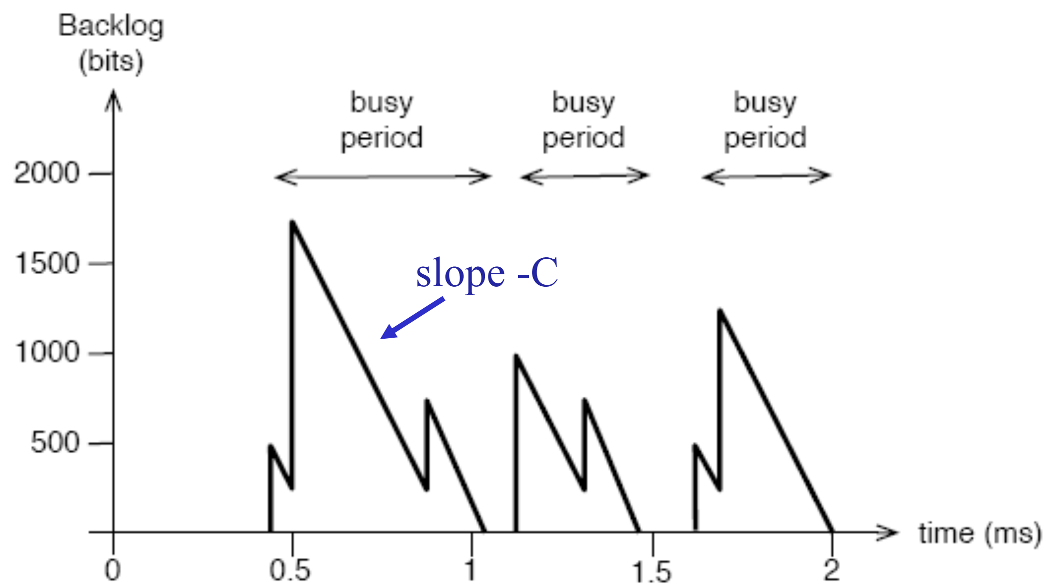


- Link rate of output is C
- Scheduler is **work-conserving** (always transmitting when there is a backlog)
- Transmission order is **FIFO**
- **Infinite Buffers** (no buffer overflows)

- Arrivals of packets to a buffered link



- Backlog at the buffered link



Definitions

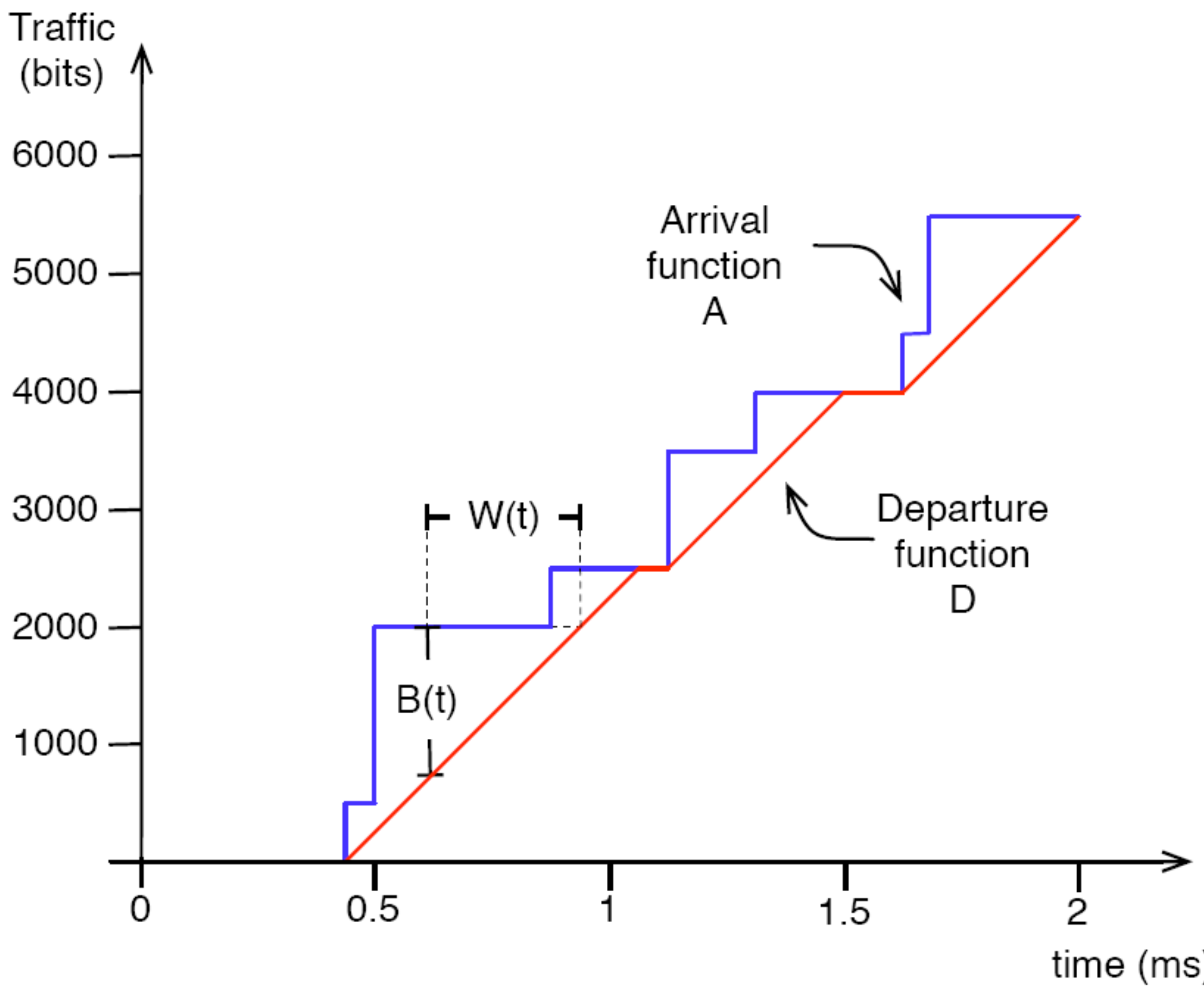
$A(t)$ Arrivals in $[0,t)$, with $A(t)=0$ if $t\le 0$

$D(t)$ Departures in $[0,t)$, with $D(t)\le A(t)$

$B(t)$ Backlog at t.
$B(t) = A(t) - D(t)$

$W(t)$ (Virtual) delay at t:
$W(t) = \inf\{y > 0 \mid D(t+y) \ge A(t)\}$

We write: $A(s,t) = A(t) - A(s)$



Traffic (bits)
6000
5000
4000
3000
2000
1000
0
0.5
1
1.5
2
time (ms)
Arrival function A
Departure function D
W(t)
B(t)

Reich's Backlog Equation (1958)

Reich's backlog equation: Given a left-continuous arrival function A and a buffered link with capacity C. Then for all $t \geq 0$ it holds that

$$B(t) = \sup_{0 \leq s \leq t} \{A(t) - A(s) - C(t-s)\} \ .$$

Meet the min-plus convolution

Use Reich's backlog equation to describe departures:

$$\begin{aligned} D(t) &= A(t) - B(t) \\ &= A(t) - \sup_{0 \le s \le t} \{A(t) - A(s) - C(t-s)\} \\ &= \inf_{0 \le s \le t} \{A(s) + C(t-s)\} \ . \end{aligned}$$

We call this operation (min-plus) convolution ($\otimes$)

Define: $S(t) = \begin{cases} Ct\,, & t > 0 \\ 0\,, & t \le 0 \end{cases}$

Then: $\boxed{D(t) = A \otimes S(t)}$

Introduction to Network Calculus

(min, +) Algebra

Processes

A Process is a function of time, $F : \mathbb{R} \to \mathbb{R} \cup \{\infty\}$

- non-negative : $F(t) \geq 0$
- non-decreasing : $F(t+s) \geq F(t)$ for $s > 0$
- one-sided (causal) : $F(t) = 0$ for $t \leq 0$

$\mathcal{F}$ = all non-negative, non-decreasing processes
$\mathcal{F}_o$ = all non-negative, non-decreasing, **and** causal processes

Meet the $(\min, +)$ algebra

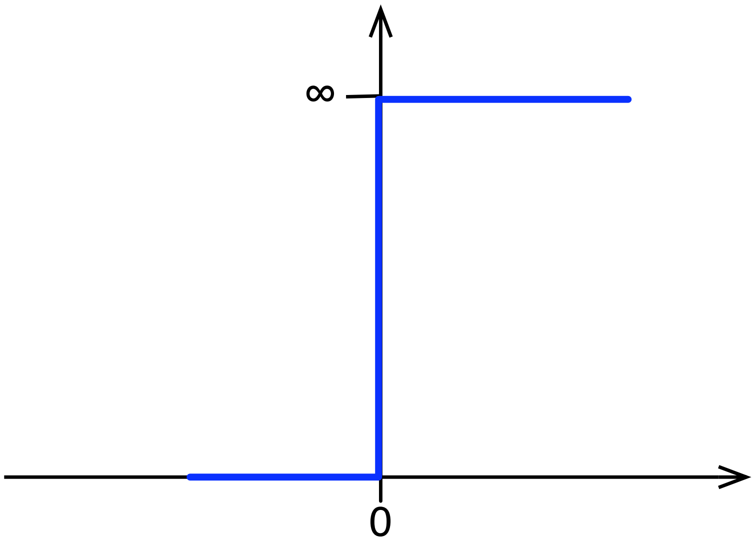
Conventional Algebra		Min-Plus Algebra
Addition $(+)$	$\longrightarrow$	Minimum $(\min, \wedge)$
Multiplication $(\cdot)$	$\longrightarrow$	Addition $(+)$

$\int_1^\infty F(t)\, dt$	$\longrightarrow$	$\inf_{t \geq 1} \{F(t)\}$
$\int_{\mathbb{R}} F(s) G(t-s)\, ds$	$\rightarrow$	$\inf_{s \in \mathbb{R}} \{F(s) + G(t-s)\}$
$(= F * G(t))$	$\rightarrow$	$(= F \otimes G(t))$

Burst and Delay Functions

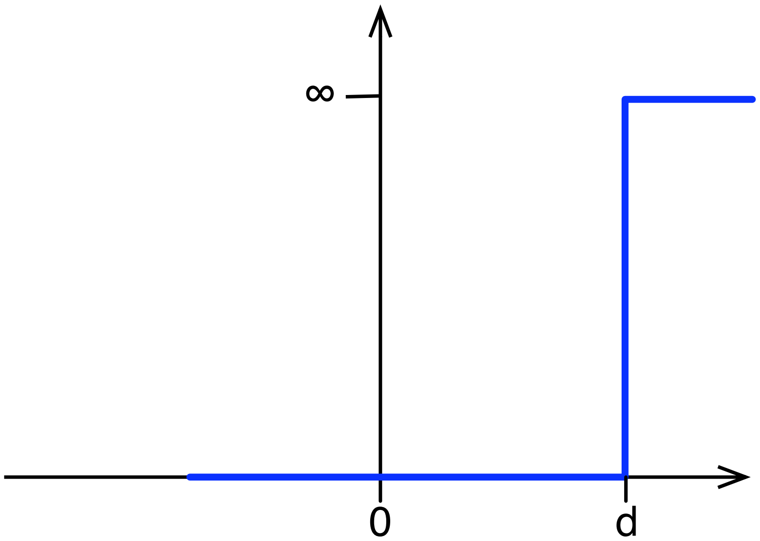
Burst function

$$\delta(t) = \begin{cases} \infty\,, & t > 0 \\ 0\,, & t \leq 0 \end{cases}$$



Delay function

$$\delta_d(t) = \delta(t - d)$$



Properties of min-plus convolution

$F, G, H \in \mathcal{F}_o$:

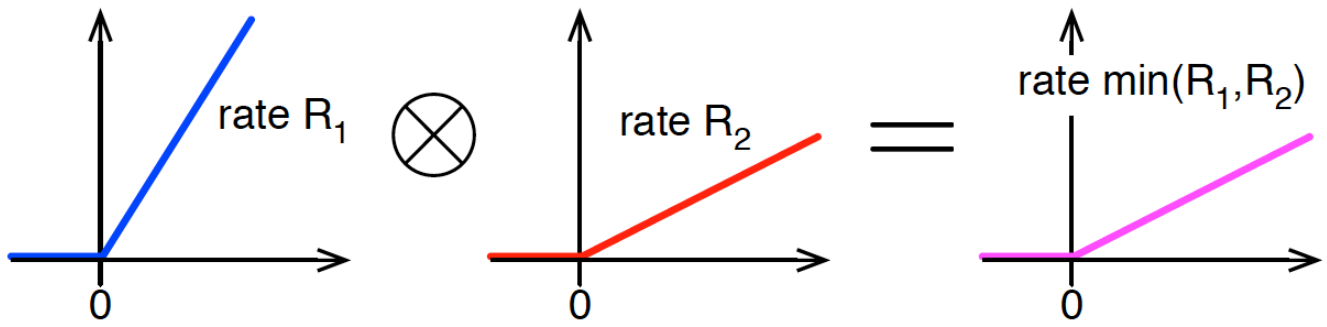
- Closure. $F \otimes G \in \mathcal{F}_o$.
- Associativity. $(F \otimes G) \otimes H = F \otimes (G \otimes H)$.
- Commutativity. $F \otimes G = G \otimes F$.
- Distributivity. $(F \wedge G) \otimes H = (F \otimes H) \wedge (G \otimes H)$.
- Neutral element. $F \otimes \delta = F$.
- Time shift. $F \otimes \delta_d(t) = F(t-d)$.
- Order preserving. If $F \leq G$ then $F \otimes H \leq G \otimes H$.
- Boundedness. $F \otimes G \leq F$, in particular, $F \otimes F \leq F$.

Except boundedness, properties hold for $F, G, H \in \mathcal{F}$

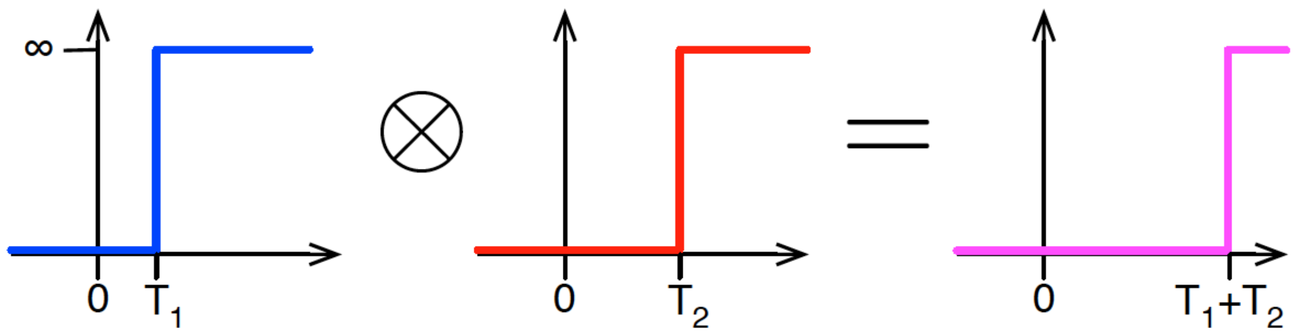
If $F, G \in \mathcal{F}_o$: $\boxed{\inf_{s \in \mathbb{R}} \{F(s) + G(t-s)\} = \inf_{0 \leq s \leq t} \{F(s) + G(t-s)\},}$

Useful min-plus convolutions

- Convolving rates yields the minimum rate:

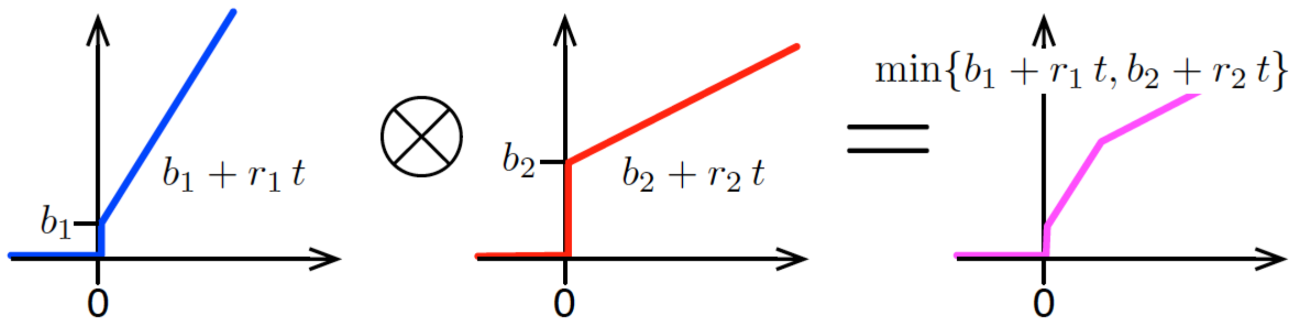


- Convolving delays yields the sum of delays:

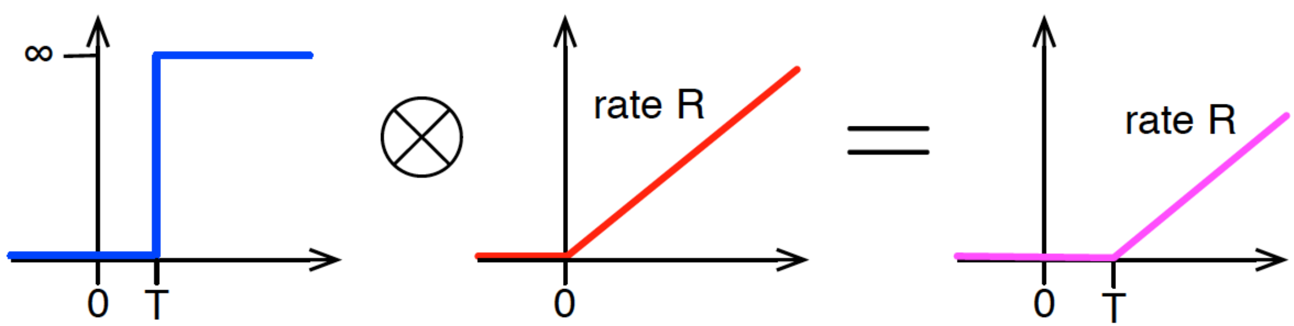


Useful min-plus convolutions

- Convolving affine functions is the minimum:



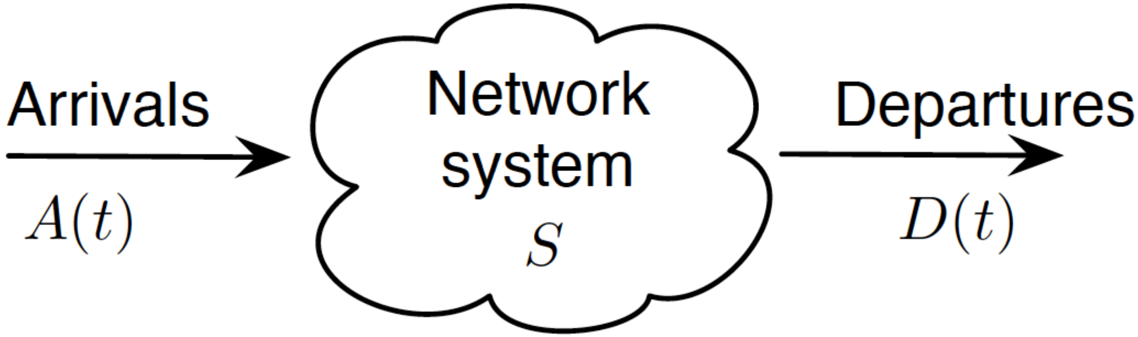
- Convolving delay and rate yields latency-rate:



Service Curves

Service Curves

- At buffered link: $S(t) = \begin{cases} Ct\,, & t > 0 \\ 0\,, & t \leq 0 \end{cases} \Rightarrow D(t) = A \otimes S(t)$
- Service curve is a generalization:



For arbitrary arrivals A and resulting departures D at a network system, a process $S \in \mathcal{F}_o$ is an *exact service curve*, if for all t,

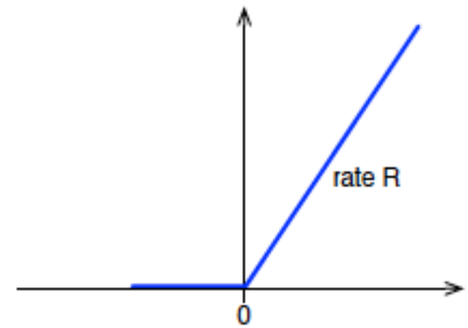
$$D(t) = A \otimes S(t)\,.$$

Lower service curve $\Rightarrow D(t) \leq A \otimes S(t)$

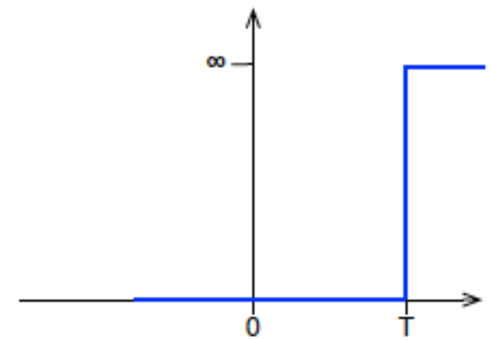
Upper service curve $\Rightarrow D(t) \geq A \otimes S(t)$

About Service Curves

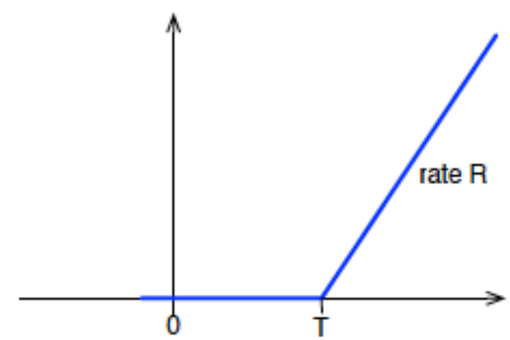
- Service curves express a service guarantee for arrivals.



- Network system does not need to be work-conserving

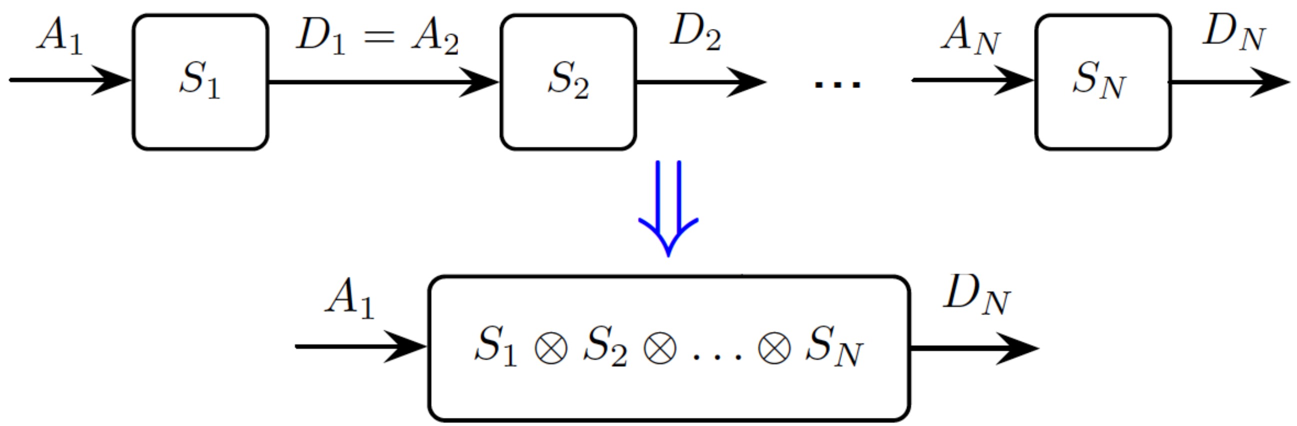


- Constant-rate server: $S(t) = [Rt]^+$
- Delay server: $S(t) = \delta_T(t)$
  (since $A \otimes \delta_T(t) = A(t - T)$!)
- Latency-rate server: $S(t) = R[t - T]^+$



- Many scheduling algorithms (e.g., Priority, Deadline-based) can be precisely described by service curves

Concatenation of Service Curves

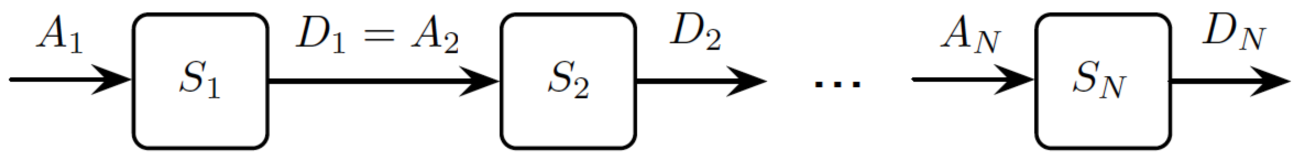


For a sequence of N service elements where the n-th element offers a lower/exact/upper service curve S_n $(n = 1, \ldots, N)$, the sequence as a whole offers an lower/exact/upper service curve

$$S_1 \otimes S_2 \otimes \ldots \otimes S_N$$

The service curve of the sequence is called *network service curve*.

Examples: Network Service Curves



Sequence of ...	$S_n(t)$	$S_1 \otimes S_2 \otimes \ldots \otimes S_N(t)$
constant rate servers	$C_n[t]^+$	$\min_n (C_n) [t]^+$
delay servers	$\delta_{d_n}(t)$	$\delta_{\sum_{n=1}^N d_n}(t)$
latency-rate servers	$C_n[t - d_n]^+$	$\min_n (C_n) [t - \sum_{n=1}^N d_n]^+$

Traffic Envelopes
and
Traffic Regulation

Meet the $(\min, +)$ deconvolution

For two processes $F, G \in \mathcal{F}$, the min-plus *deconvolution* $F \oslash G$ is

$$F \oslash G(t) = \sup_{s \geq 0} \{F(t+s) - G(s)\}$$

- If $F, G \in \mathcal{F}_o$, then $\sup_{s\geq 0} \{F(t+s) - G(s)\} = \sup_{s \in \mathbb{R}} \{F(t+s) - G(s)\}$.
- Weak properties: not closed , not associative, not commutative.

$F, G, H \in \mathcal{F}_o$

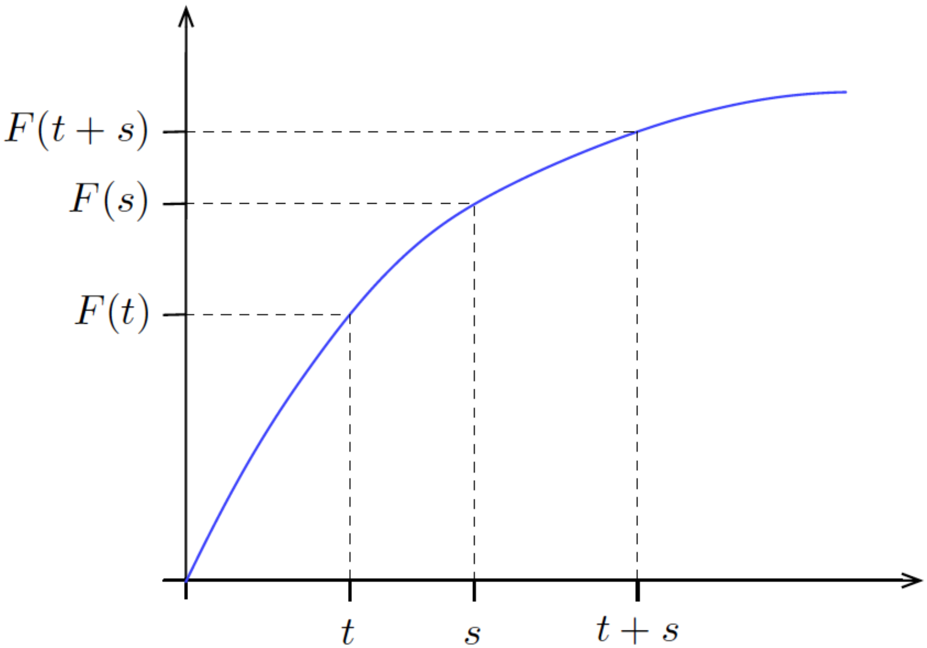
- **Composition of $\otimes$ and $\oslash$.** $(F \oslash G) \oslash H = F \oslash (G \otimes H)$.
- **Duality.** $F \leq G \otimes H$ if and only if $F \oslash H \leq G$.

Subadditivity

A function F is *subadditive*, if for all $s, t \in \mathbb{R}$

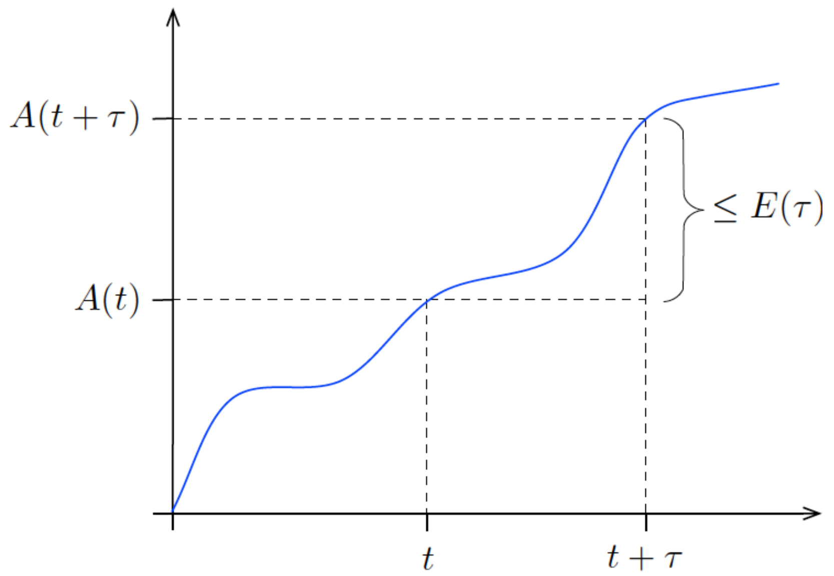
$$F(t+s) \leq F(t) + F(s) \ .$$

- F concave $\Rightarrow$ F subadditive
- F subadditive:
  - $F = F \otimes F$
  - $F = F \oslash F$
- $F \in \mathcal{F}_o$:
  - $F \oslash F \in \mathcal{F}_o$
  - $F \oslash F$ subadditive



Traffic Envelopes

Traffic envelopes bound traffic over intervals of time



A function E is a traffic envelope for A ($A \sim E$) if

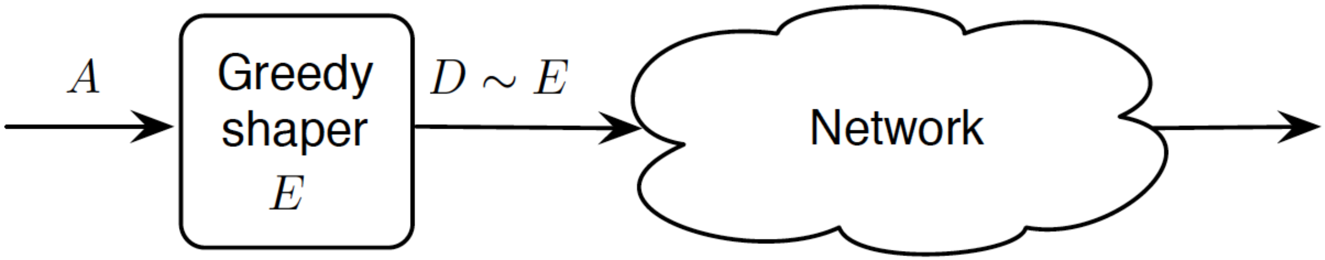
$$E(\tau) \geq A(t+\tau) - A(t)\,, \quad \forall \tau \geq 0\,, \forall t \in \mathbb{R}.$$

- $A \sim E \quad \Rightarrow \quad A = A \otimes E$
- Good envelopes are subbadditive

Traffic shaping (Greedy shaper)

Greedy shaper : A network element that

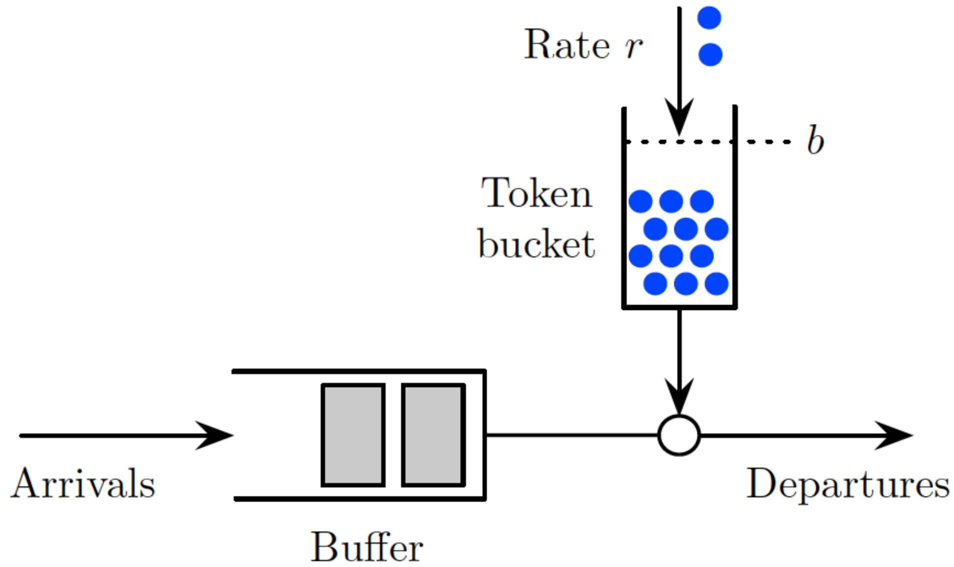
- limits arrivals to a network to a given specification (traffic envelope),
- buffers non-compliant traffic, and
- releases buffered traffic when it becomes compliant.



If E is subadditive, greedy shaper offers an exact service curve, i.e.,

$$D = A \otimes E$$

Example of greedy shaper: Token Bucket



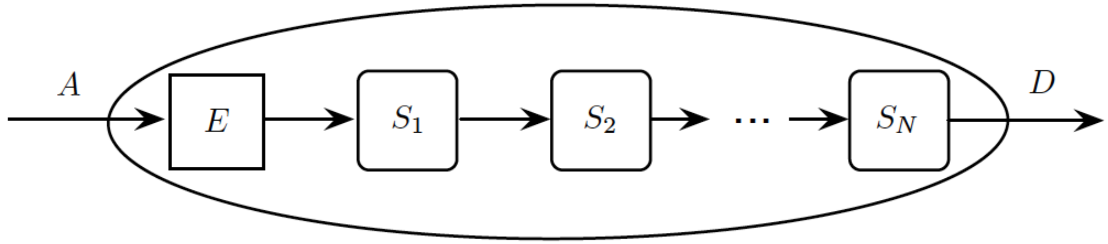
$$E(\tau) = \begin{cases} b + r\,\tau \ , & \tau > 0 \\ 0 \ , & \tau \le 0 \end{cases}$$

$\Longrightarrow$ $E(\tau)$ is an exact service curve for the token bucket

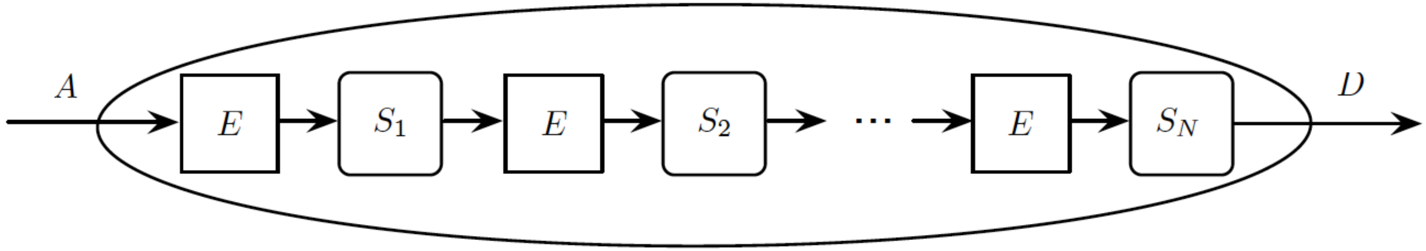
Concatenation of traffic shapers

The following networks offer the same service (as long as E is subadditive). Why?

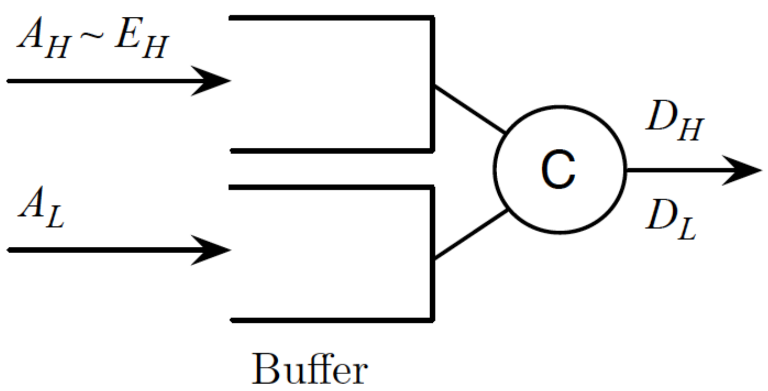
One traffic shaper at network ingress:



Traffic shaping at each node:



Leftover Service Curve



Given a buffered link with rate C, with two priorities. Low-priority traffic has arrival function A_L, and high-priority traffic has arrival function A_H with $A_H \sim E_H$. Then

$$S_L(t) = [Ct - E_H(t)]^+$$

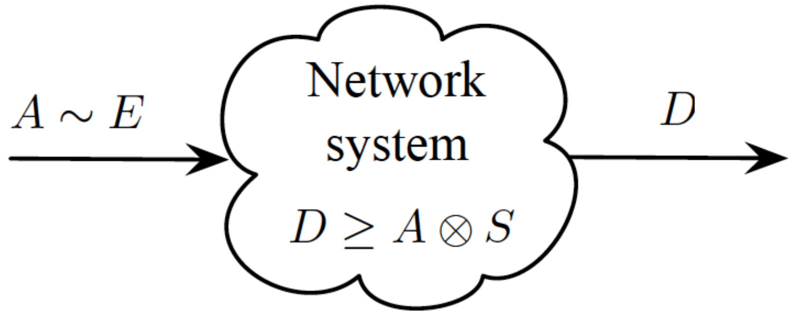
is a lower service curve for the low-priority traffic.

This leftover service curve is a (pessimistic) benchmark for the service experienced at a link with multiplexing.

Performance Bounds

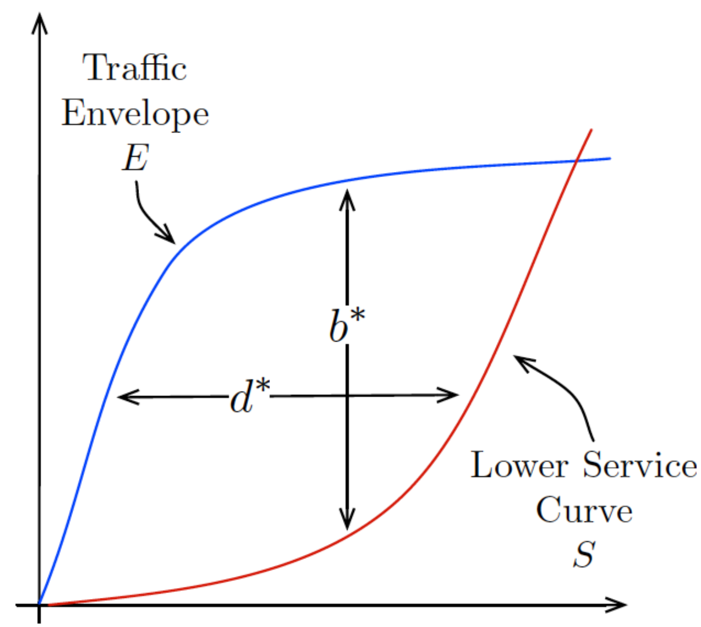
Three Performance Bounds

Given arrival function A with traffic envelope E, and a network system with lower service curve S:



1. Envelope for departures:
   $D \sim E \oslash S$ (for $t > 0$)
2. Backlog bound b^*:
   $B(t) \leq b^* = E \oslash S(0)$
3. Delay bound d^*: $W(t) \leq d^*$ with
   $d^* = \inf \{ d \geq 0 \,|\, E \oslash S(-d) \leq 0 \}$

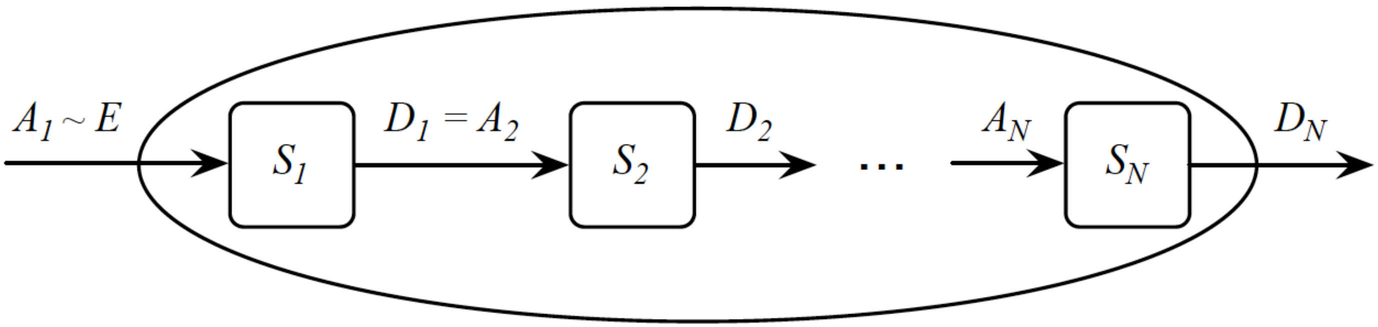
Bounds are tight for some $A \sim E$.



Proof of Backlog Bound

$$
\begin{aligned}
B(t) &= A(t) - D(t) \\
&= A(t) - \inf_{0\le s\le t}\{A(t-s) + S(s)\} \\
&= \sup_{0\le s\le t}\{A(t-s,t) - \mathcal{S}(s)\} \\
&\le \sup_{s\ge 0}\{E(s) - \mathcal{S}(s)\} \\
&= E \oslash \mathcal{S}(0)
\end{aligned}
$$

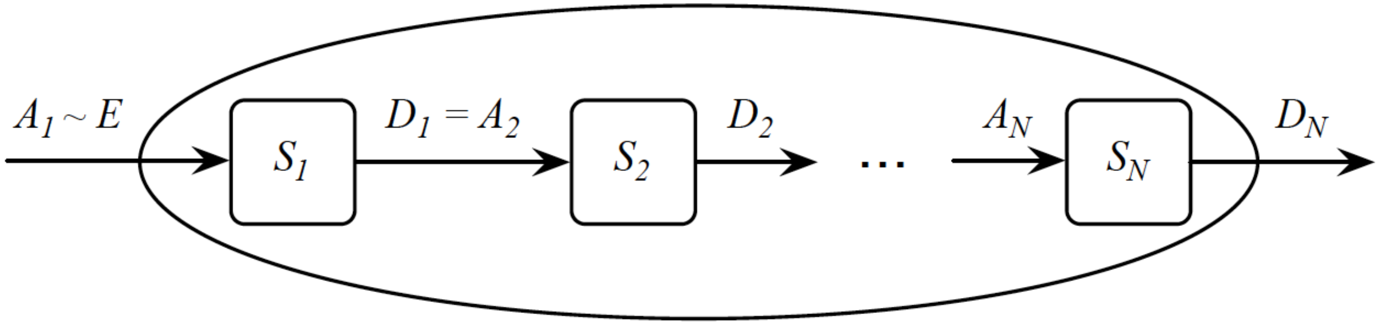
End-to-End bounds



- $A_1 \sim E$
- $D_n \geq A_n \otimes S_n$

$$E(t) = \begin{cases} 0\ , & t \leq 0 \\ b + rt\ , & t > 0 \end{cases}\ , \qquad S_n(t) = \begin{cases} 0\ , & t \leq T \\ R(t-T)\ , & t > T \end{cases}$$

Objective: Compute end-to-end backlog and delay bounds



Two Analysis Approaches:

1. Network Service Curve: Obtain $S^{net} = S_1 \otimes S_2 \otimes \ldots \otimes S_N$, and compute delay bound using E and S^{net}.

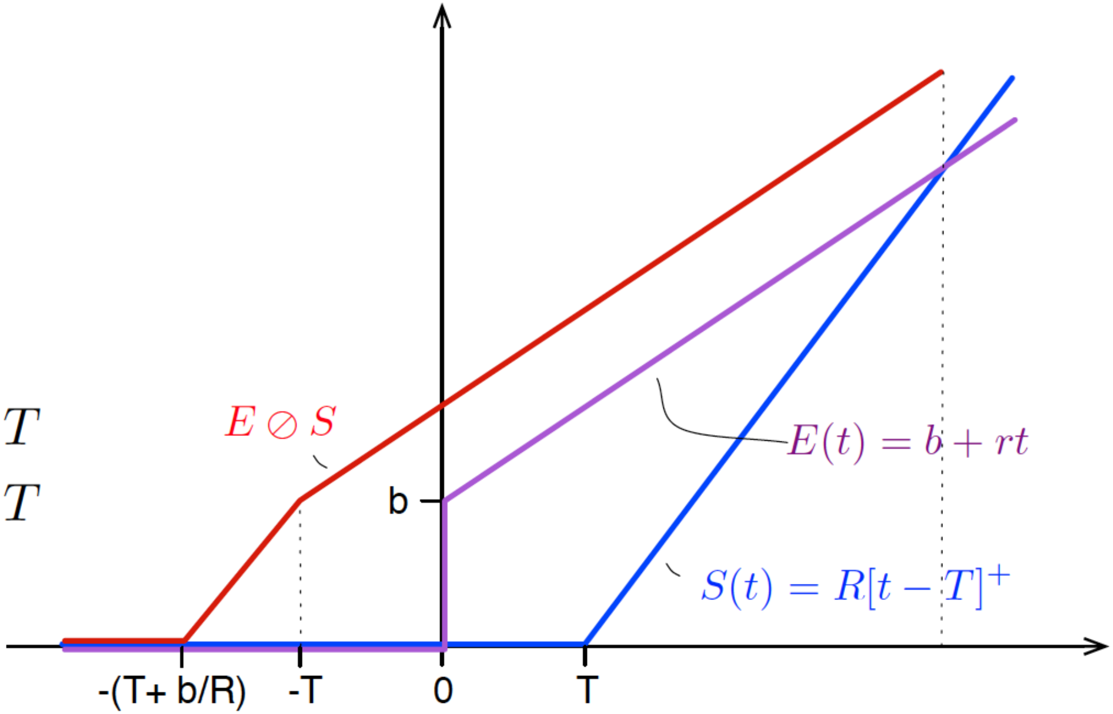
2. Add per-node bounds: Compute arrival envelopes iteratively at each node.
   - 1^{st} Node: Use envelope E and service curve S_1
   - 2^{nd} Node: Use $A_2 \sim E \oslash S_1$ and service curve S_2
   - n^{th} Node: Use $A_n \sim (\ldots(E \oslash S_1) \oslash S_2) \oslash \ldots) \oslash S_{n-1}$ and S_n

All we need is a deconvolution

Deconvolution of token bucket envelope with latency-rate service curve:

$$E(t) = \begin{cases} 0 \ , & t \le 0 \\ b + rt \ , & t > 0 \end{cases}$$

$$S(t) = \begin{cases} 0 \ , & t \le T \\ R(t-T) \ , & t > T \end{cases}$$



$$E \oslash S(t) = \begin{cases} 0 \ , & t \le -(T + \frac{b}{R}) \\ b + R(t+T) \ , & -(T+\frac{b}{R}) < t \le -T \\ b + r(t+T) \ , & t > -T \end{cases}$$

End-to-End bounds

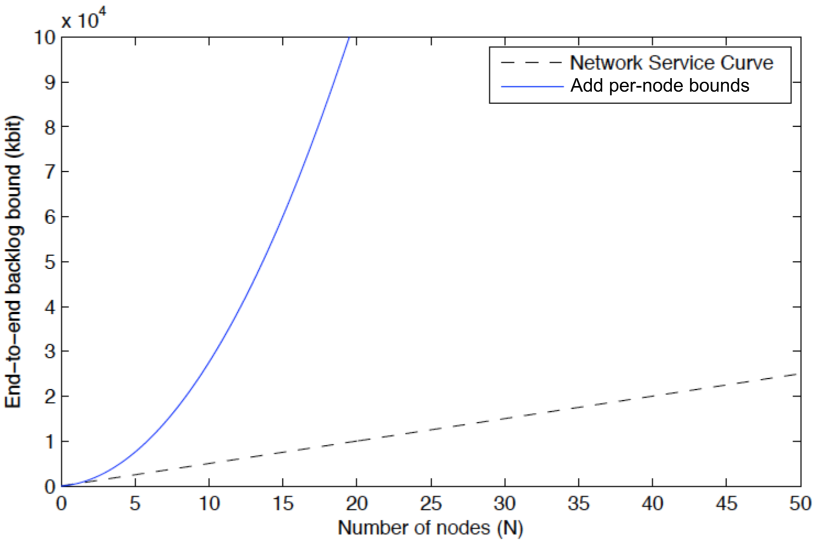
Results:

	Network Service Curve	Add Per-node Bounds
b^*_{net}	$b + NrT$	$Nb + (N^2 + N)\frac{rT}{2}$
d^*_{net}	$\frac{b}{R} + NT$	$\frac{Nb}{R} + NT + (N^2 - N)\frac{rT}{2R}$

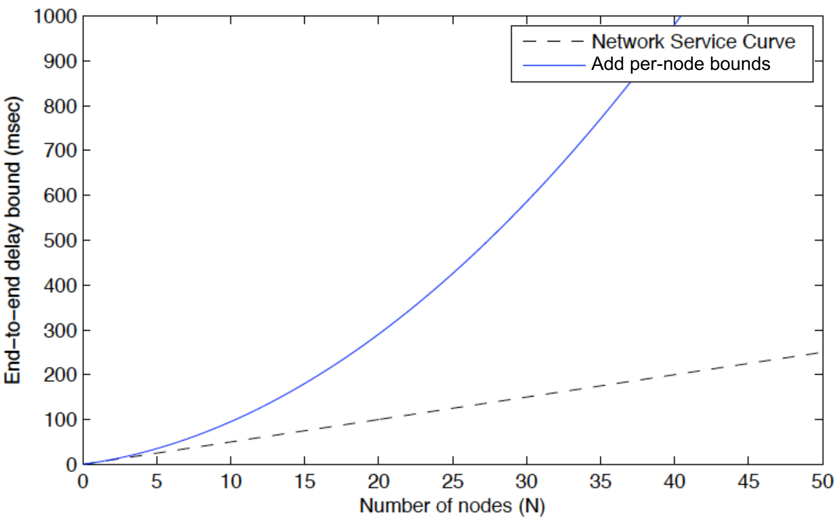
End-to-End Bounds

- $b = 10\ kb,$
- $r = 100\ kbps,$
- $T = 5\ ms,$
- $R = 500\ kbps$

End-to-End Backlog Bound



End-to-End Delay Bound



(min, +) Linear Systems

(Classical) System Theory

Linear Time Invariant (LTI) Systems

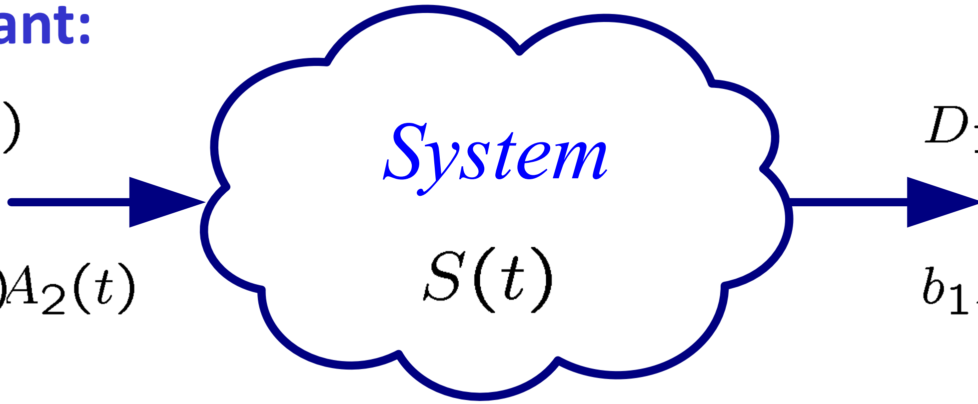
Linear:

Time invariant:

$A_1(t)$ $A(t)$ $A_2(t)$

Input

$b_1 A_1$ $A(t-\tau)$ $A_2(t)$



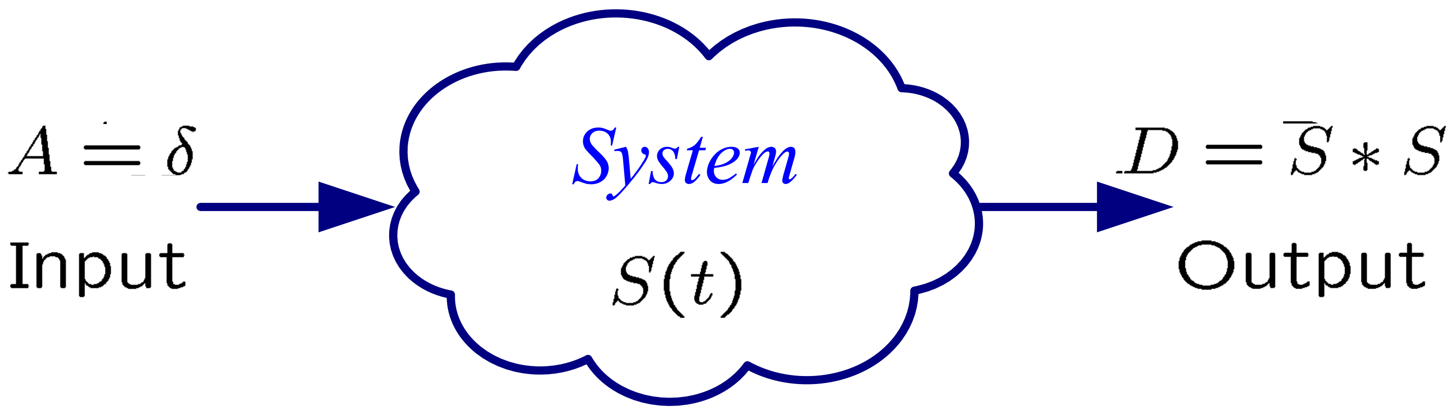
$D_1(t)$ $D(t)$ $D_2(t)$

Output

$b_1 D$ $D(t-\tau)$ $D_2(t)$

(Classical) System Theory

Linear Time Invariant (LTI) Systems



- If input is Dirac impulse, output is the impulse response S
- Output can be calculated from input and system response:

$$D(t) = \int_{-\infty}^{\infty} A(\tau) \cdot S(t-\tau)d\tau =: A * S(t)$$

(classical) convolution”

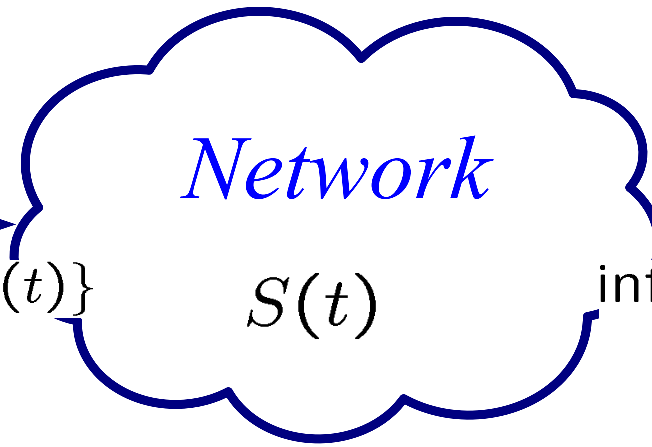
Min-Plus Linear System

min-plus linear:

time-invariant:

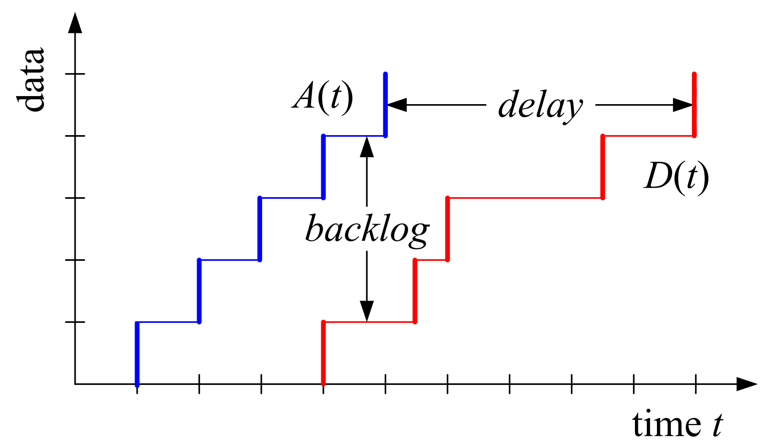
$A_1(t)$ $A_2(t)$ $A(t)$

$\inf\{b_1 + A_1(t), b_2 + A_2(t)\}$

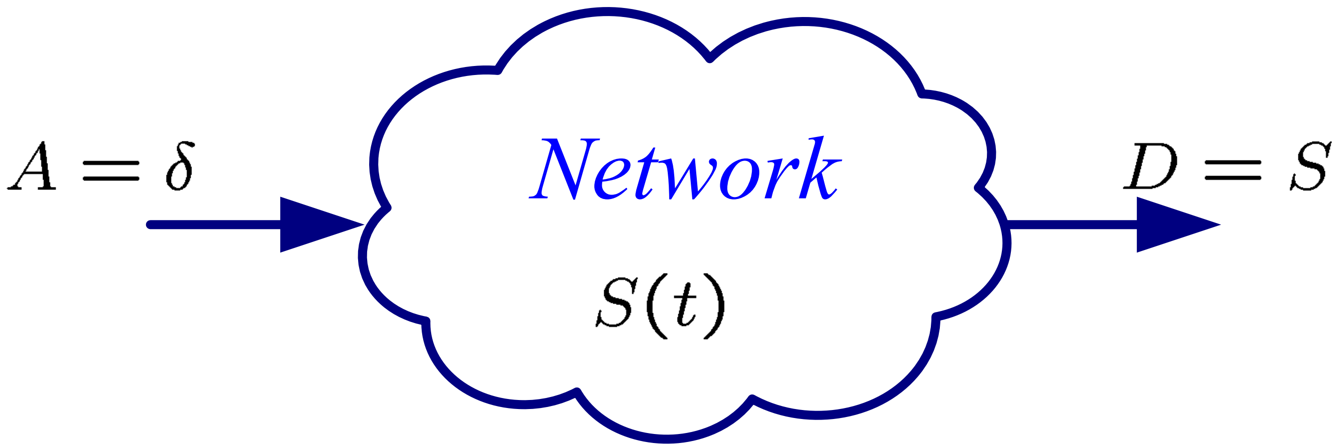


$D_1(t)$ $D_2(t)$ $D(t)$

$\inf\{b_1 + D_1(t), b_2 + D_2(t)\}$



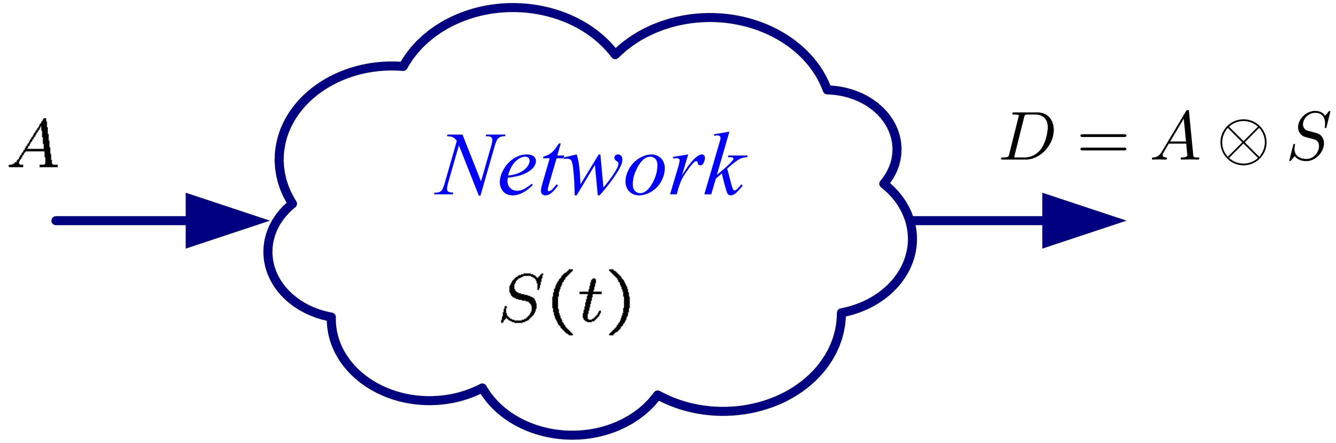
Min-Plus Linear Systems



- If input is burst function δ, output is the service curve S

$$\delta(t) = \begin{cases} \infty \,, & \text{if } t > 0 \,, \\ 0 \,, & \text{otherwise} \end{cases}$$

Min-Plus Linear Systems



- Departures can be calculated from arrivals and service curve:

$$D(t) = \inf_{\tau}\{A(\tau + S(t-\tau)\} = A \otimes S(t)$$

(min, +) convolution

Transforms

- Classical LTI systems

Time domain		Frequency domain
$f(t)$ $f \otimes g(t)$	Fourier transform $\longrightarrow$	$\mathcal{F}_f(\omega) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(t) e^{-i\omega t} dt$ $\mathcal{F}_{f*g}(\omega) = \mathcal{F}_f(\omega) \cdot \mathcal{F}_g(\omega)$

- Min-plus linear systems

Time domain		Rate domain
$f(t)$ $f \otimes g(t)$	Legendre transform $\longrightarrow$	$\mathcal{L}_f(r) = \sup_\tau \{r\tau - f(\tau)\}$ $\mathcal{L}_{f \otimes g} = \mathcal{L}_f + \mathcal{L}_g$

Properties:

(1) $\mathcal{L}(\mathcal{L}(f)) \leq f$. If f is convex: $\mathcal{L}(\mathcal{L}(f)) = f$

(2) If g convex, then $f \geq g \Leftrightarrow \mathcal{L}_f \leq \mathcal{L}_g$

(3) Legendre transforms are always convex

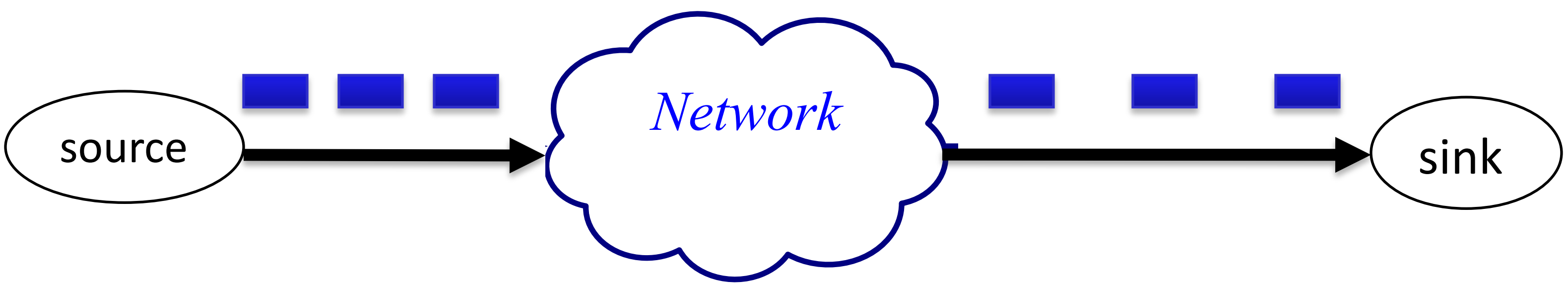
Characterizing non-linear systems

- Many networks are not min-plus linear
  - i.e., for some t: $D \neq A \otimes S$
- … but can be described by a **lower service curve**
  - such that for all t: $D(t) \geq A \otimes S(t)$
- Having a lower service curve is often enough, since it provides a lower bound on the service !!

Bandwidth Estimation (Case Study)

Probing a network with packet trains

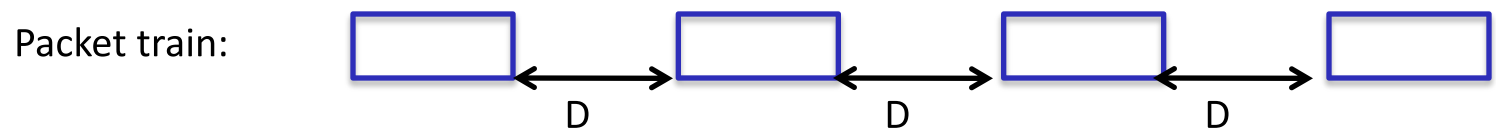
- A network probe consists of a sequence of packets (**packet train**)
- The packet train is from a source to a sink
- For each packet, a measurement is taken when the packet is sent by the source (arrival time), and when the packet arrives at the sink (departure time)



Edited slide from: V. Ribeiro, Rice. U, 2003

Rate Scanning Method

- Each packet trains is sent at a fixed rate r (in bits per second):
  - All packets in the train have the same size
  - Packets of packet train are sent with same distance



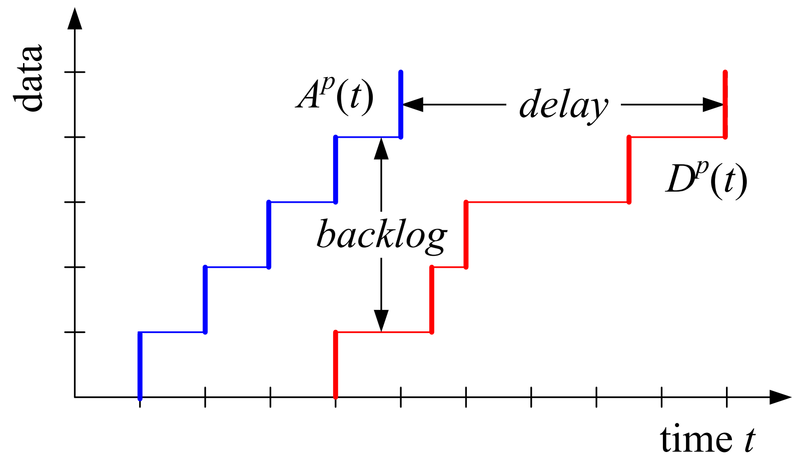
- Rate Scanning: Source sends multiple packet trains, each with a different rate r

Bandwidth Estimation in the Network Calculus

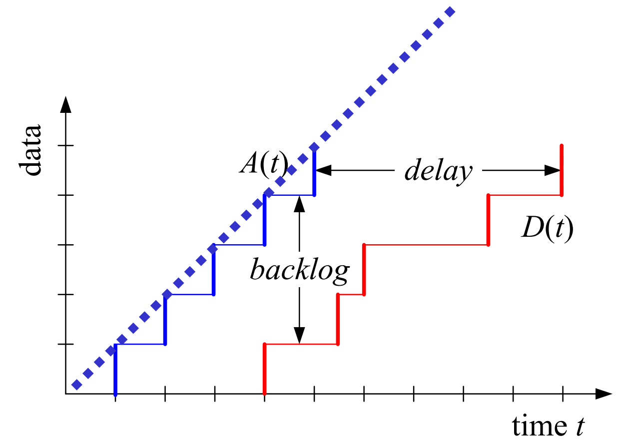
- View the network as a min-plus system that is either linear or nonlinear

Bandwidth estimation scheme:

1. Timestamp packets of packet train:
   $A^p(t)$- Send probes
   $D^p(t)$- Receive probes
2. Use probes to find an S that satisfies
   $$D(t) \geq A \otimes S(t)$$
   for all (A,D).
3. S is the estimate of the available bandwidth. The goal is to select as large as possible.



Rate Scanning (1): Theory



- Backlog: $B(t) = A(t) - D(t)$
- Max. backlog:

$$B_{max} = \sup_t \{A(t) - D(t)\}.$$

- If $A(t) = rt$ and $D(t) = A \otimes S(t)$, we can write this as:

$$\begin{aligned} B_{max}(r) &= \sup_t \{rt - \inf_\tau \{r\tau + S(t-\tau)\}\} \\ &= \sup_t \{\sup_\tau \{r(t-\tau) - S(t-\tau)\}\} \\ &= \sup_t \{rt - S(t)\} \\ &= \mathcal{L}_S(r) \end{aligned}$$

- Inverse transform: If S is convex we have

$$S(t) = \mathcal{L}(\mathcal{L}_S)(t) = \mathcal{L}_{B_{max}}(t) = \sup_r \{rt - B_{max}(r)\}$$

Rate Scanning (2): Algorithm

Init: $S(t) = 0$

Step 1: Transmit a packet train at rate r,

compute $B_{max}(r)$

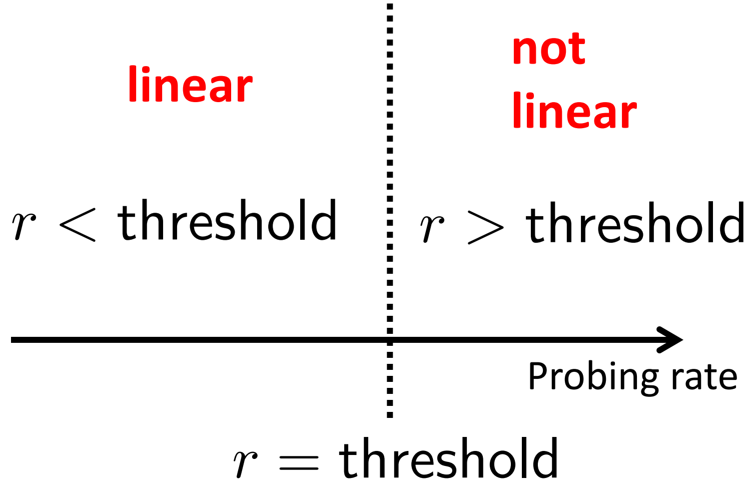
compute $S(t) = \mathcal{L}_{B_{max}}(t)$

Step 2: If estimate of S has improved, increase r and go to Step 1.

Non-Linear Systems

- When we exploit $D(t) = A \otimes S(t)$, we assume a min-plus linear system
- In non-linear networks, we can only find a lower service curve that satisfies $D(t) \geq A \otimes S(t)$

- We view networks as system that are linear at low load, and that become non-linear when network load exceeds a threshold.



- When increasing the probing rate, we eventually exceed the threshold at which the network becomes non-linear

Detecting Non-linearity

How to determine when the probed network becomes non-linear?

Backlog convexity criterion

- Suppose that we probe at constant rates $A(t) = rt$
- Legendre transform is always convex
- In a linear system, the max. backlog is the Legendre transform of the service curve:

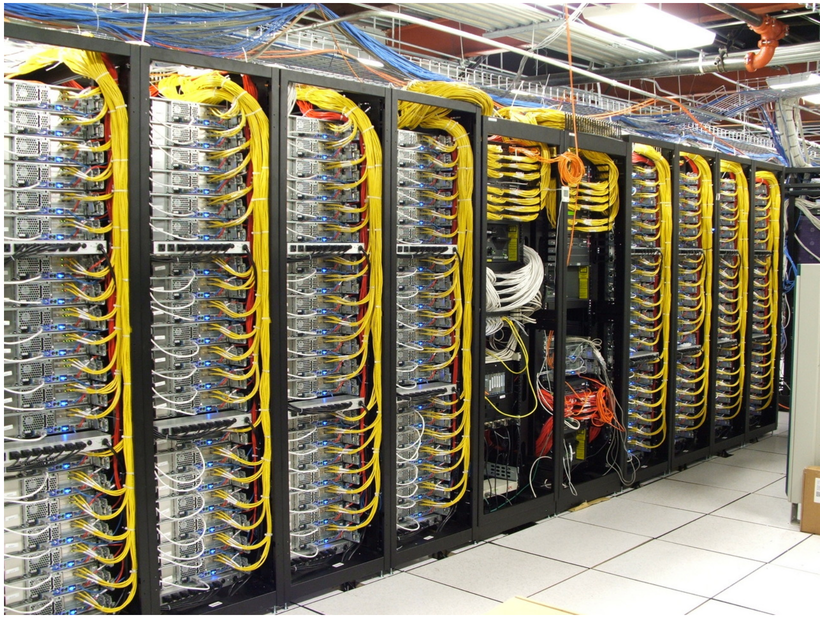
$$B_{max}(r) = \mathcal{L}_S(r)$$

→ When $B_{max}(r)$ is no longer convex the system is no longer linear

→ Stop (increasing the rate) when $B_{max}(r)$ has become non-convex

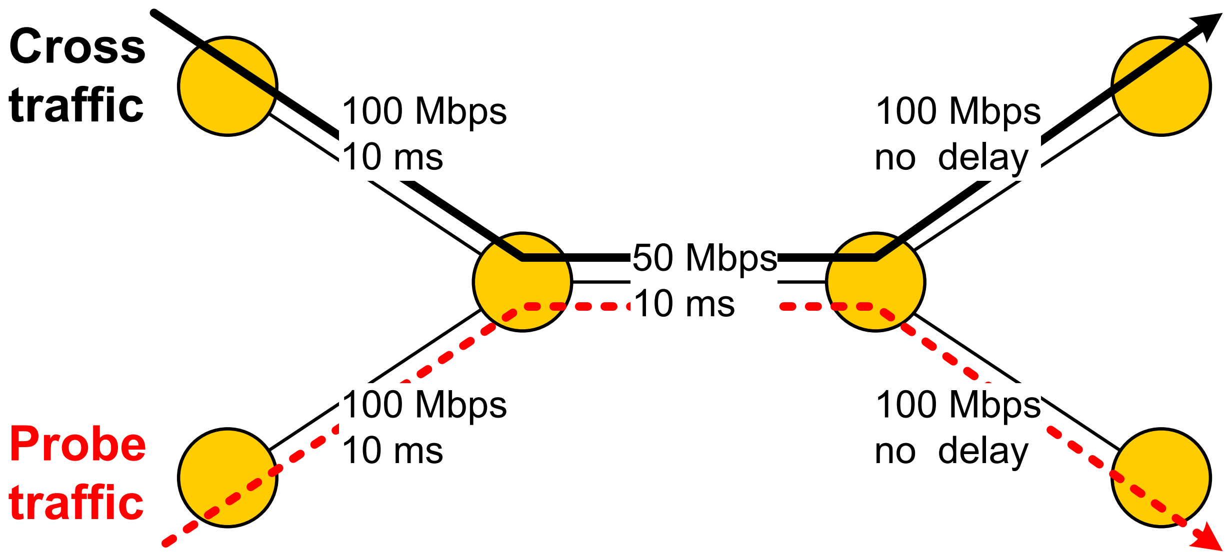
EmuLab Measurements

- Emulab is a network testbed at U. Utah
  - can allocate PCs and build a network
  - controlled rates and latencies



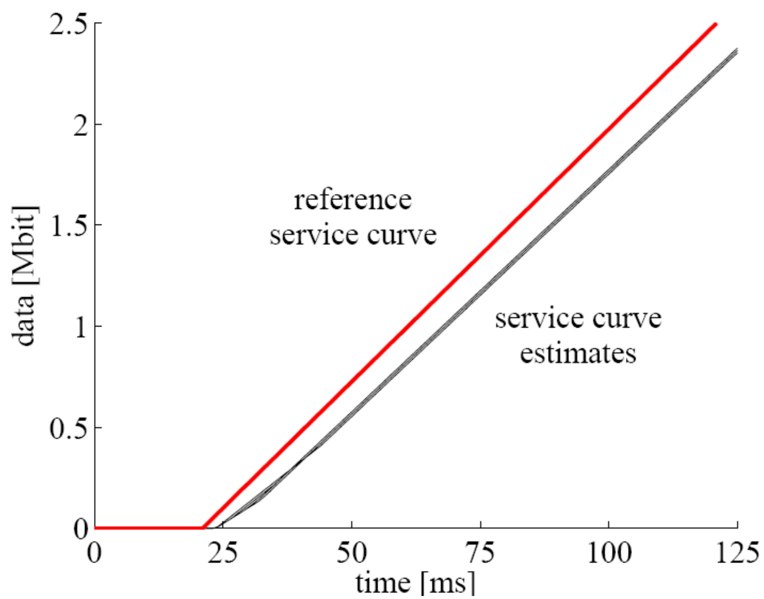
Dumbbell Network

- UDP packets with 1480 bytes (probes) and 800 bytes (cross)
- Cross traffic: 25 Mbps



Constant Bit Rate (CBR) Cross Traffic *(LFV10)*

- Cross traffic is sent at a constant rate (=CBR)

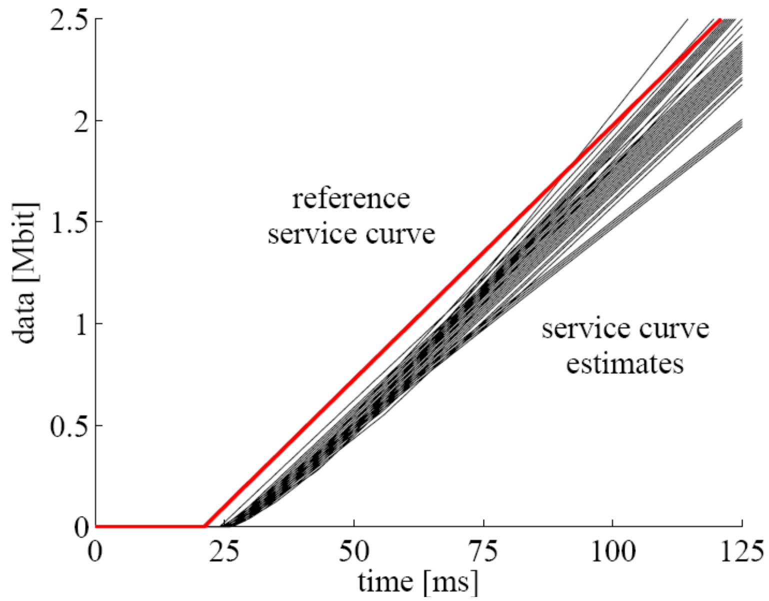


CBR cross traffic

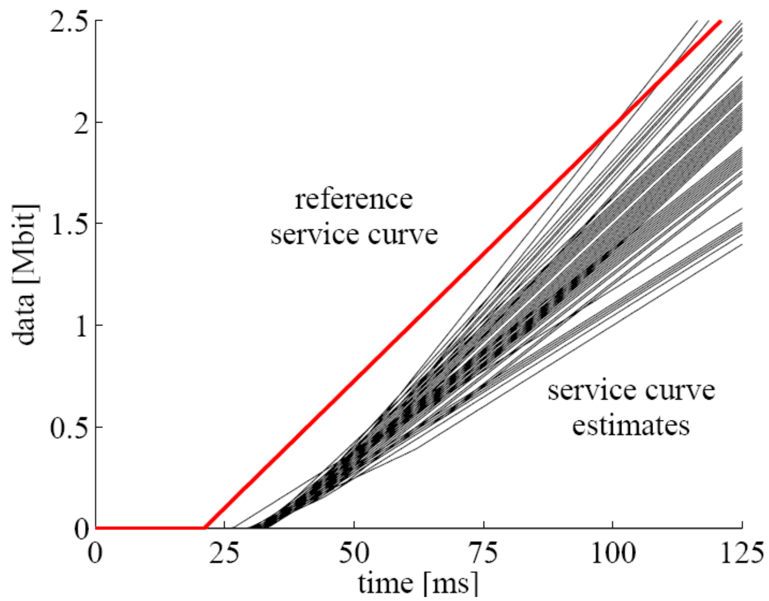
- The "reference service curve" (red) shows the ideal results. The "service curve estimates" shows the results of the rate scanning method
- Figure shows 100 repeated estimates of the service curve

Rate Scanning: Different Cross Traffic *(LFV10)*

- Exponential: random interarrivals, low variance
- Pareto: random interarrivals, very high variance



Exponential



Pareto

A Few Remarks on
Deterministic Network Calculus

Remarks on deterministic network calculus

- Comprehensive analysis framework for communication networks:
  - Applicable for an analysis that includes "real traffic"
  - Precise/exact description of traffic control algorithms: traffic shaping, scheduling

- Foundations in
  - Dioid algebra
  - Linear systems

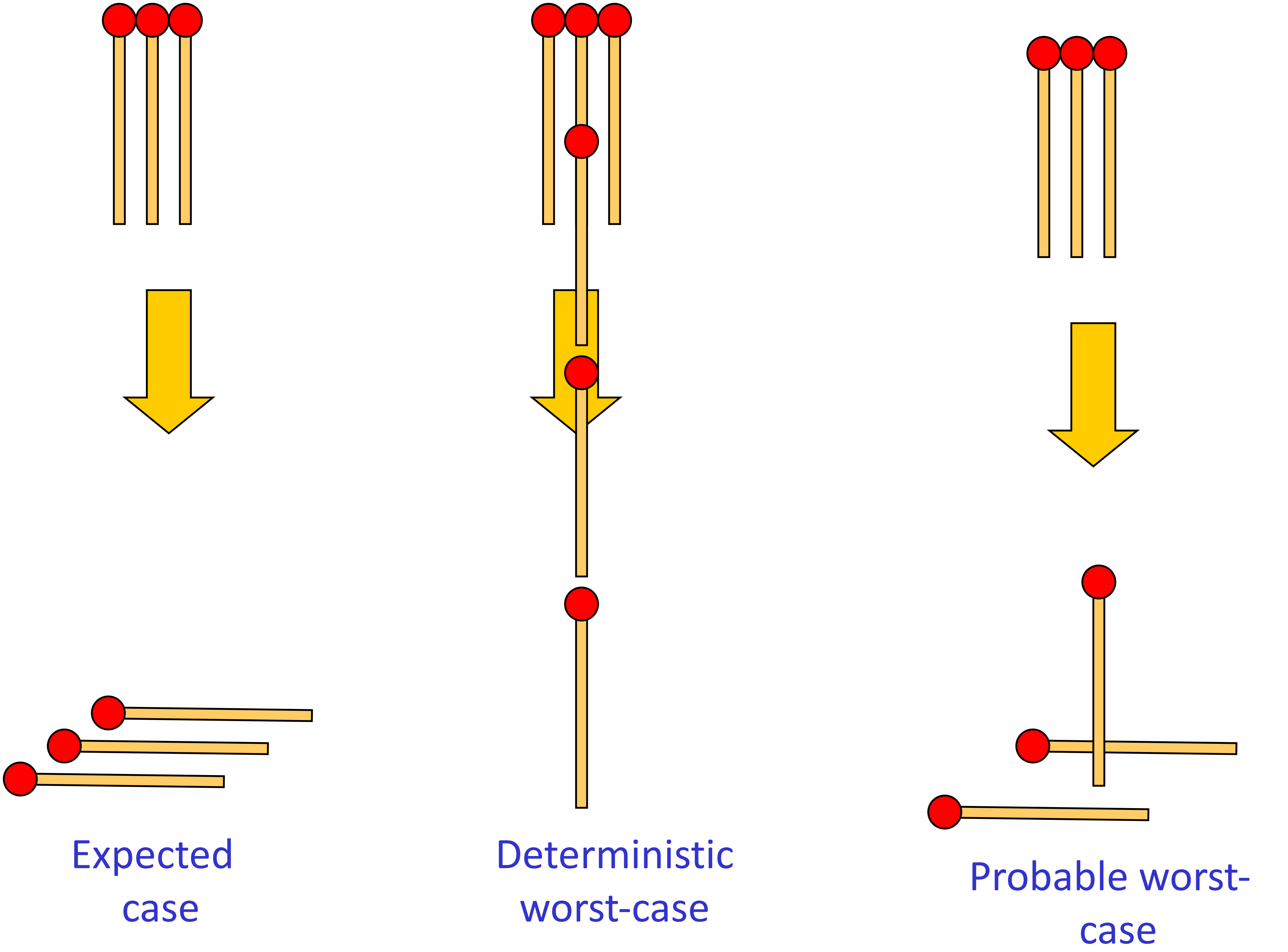
- Great for teaching due to the closeness of theory and practice (e.g., traffic models, traffic shaping, scheduling, etc.)

Mythbusters

- **Myth:** "Deterministic Network Calculus is always pessimistic"
  - Deterministic network calculus analysis of min-plus linear systems can be exact
  - Lower service curves of non-trivial scheduling algorithms can yield a tight (necessary and sufficient) analysis
  - Worst-case analysis is pessimistic (for average case). Therefore, if network calculus is used for worst-case analysis, results are accordingly

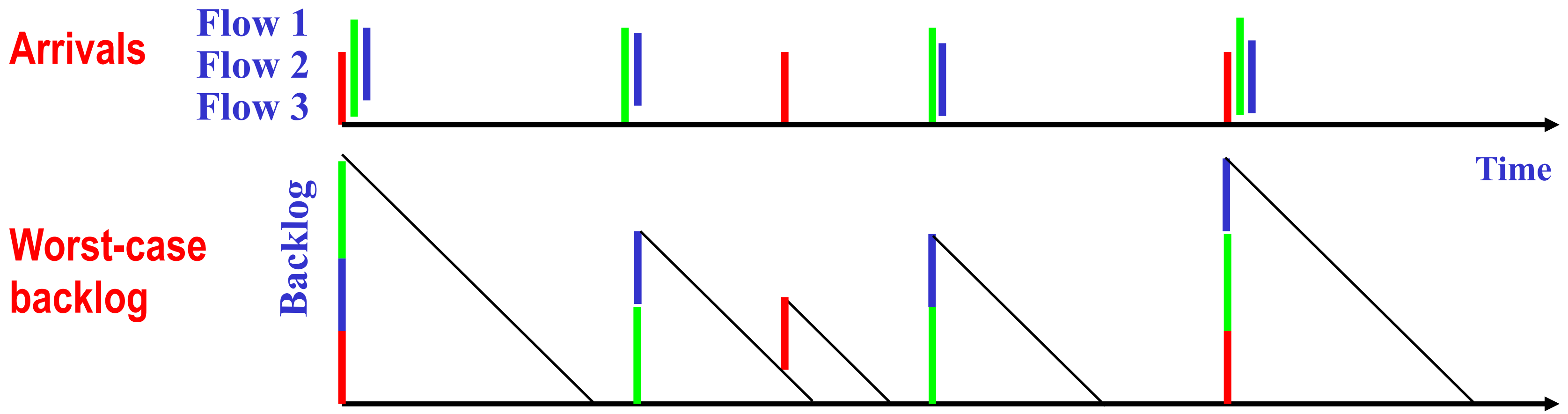
- **Not a Myth:** "Deterministic Network does not account for statistical multiplexing"
  - This motivates Stochastic Network Calculus

STOCHASTIC NETWORK CALCULUS


Expected case
Deterministic worst-case
Probable worst-case

Statistical Multiplexing Gain

Worst-case arrivals



Statistical Multiplexing Gain

$$\begin{pmatrix} \text{Resources needed} \\ \text{to support} \\ \text{guarantees} \\ \text{for N flows} \end{pmatrix} << N \cdot \begin{pmatrix} \text{Resources needed} \\ \text{to support} \\ \text{guarantees} \\ \text{for 1 flow} \end{pmatrix}$$

Statistical multiplexing gain is the raison d'être for packet networks.

- Source of statistical multiplexing is the stochastic independence of aggregated traffic sources

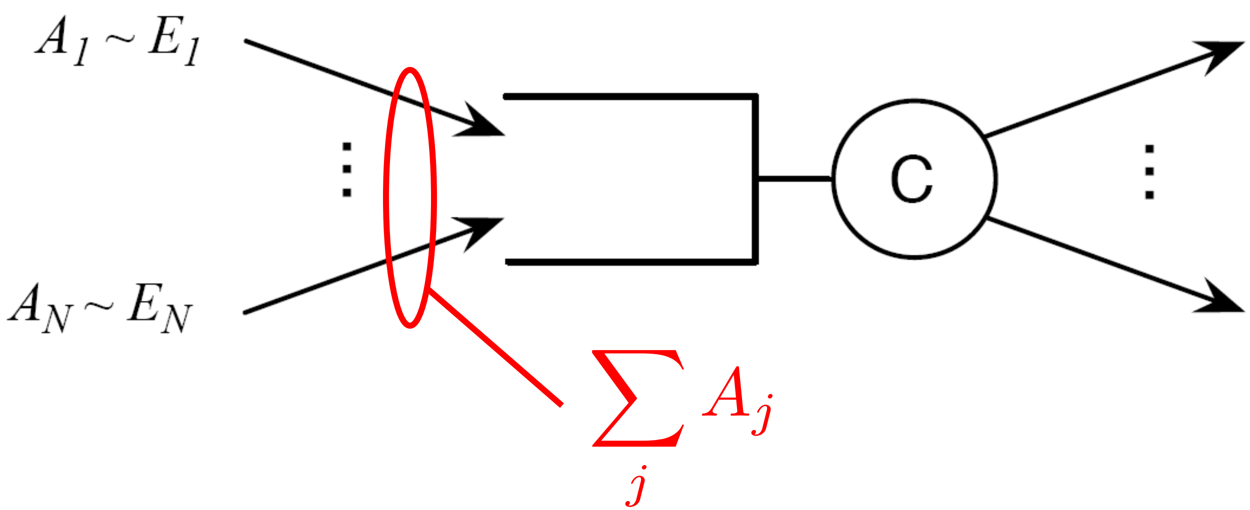
Motivation for Stochastic Network Calculus

Extend network calculus analysis so that it can account for

1. Statistical multiplexing
2. Random service models (wireless!)

Statistical Envelopes

Multiplexed arrivals



- Arrivals from group of flows: $\sum_{j=1}^{N} A_j$
- Deterministic envelope of aggregate: $\sum_{j=1}^{N} E_j$
- Statistical envelope: a traffic envelope $\mathcal{G}$ of the aggregate that may be violated with a small probability,
  and with $\mathcal{G} \ll \sum_j E_j$

Statistical Envelope

A statistical (traffic) envelope $\mathcal{G}$ for a random arrival process A with stationary increments satisfies for all $s \leq t$

$$Pr\left[A(s,t) > \mathcal{G}(t-s;\varepsilon)\right] < \varepsilon$$

- ε is violation probability
- Statistical envelope bounds arrival of a flow with high certainty
- Statistical envelopes are non-random functions
- Due to stationary increments, we can write

$$Pr\left[A(t) > \mathcal{G}(t;\varepsilon)\right] < \varepsilon$$

Constructing Statistical Envelopes

Given: N i.i.d. flows $A_1(t), \ldots, A_N(t)$

- Expected value $\mathrm{E}[A(t)]$
- Variance $\mathrm{Var}[A(t)]$
- Moment generating function $\mathrm{M}_A(\theta, t)$

Compute:

$$\boxed{Pr\left[\sum_{j=1}^{N} A_j(t) > \mathcal{G}(t;\varepsilon)\right] < \varepsilon}$$

Tools:

- Central Limit Theorem
- Chernoff Bound
- Hoeffding Bound

Statistical Envelopes: Central Limit Theorem

Central Limit Theorem:

$$Pr\left[\sum_{j=1}^{N} A_j(t) > x\right] \approx 1 - \Phi\left(\frac{x - N\mathsf{E}[A(t)]}{\sqrt{N\mathsf{Var}[A(t)]}}\right) .$$

$$\Longrightarrow \quad \mathcal{G}(t;\varepsilon) = N\mathsf{E}[A(t)] + \sqrt{N} z_{1-\varepsilon} \sqrt{\mathsf{Var}[A(t)]}$$

($\Phi(x)$: CDF of standard normal and $\Phi(z_{1-\varepsilon}) = \varepsilon$)

Statistical Envelopes: Chernoff Bound

Chernoff Bound:

$$Pr\left[\sum_{j=1}^{N} A_j(t) > x\right] \leq e^{-\theta x}\left(\mathsf{M}_A(\theta, t)\right)^N$$

$$\Longrightarrow \quad \mathcal{G}(t; \varepsilon) = \inf_{\theta > 0} \tfrac{1}{\theta}\left(N \ln \mathsf{M}_A(\theta, t) - \ln \varepsilon\right)$$

Statistical Envelopes: Hoeffding Bound

($A(t)$ must be bounded)

Hoeffding Bound :

$$Pr\left[\sum_{j=1}^{N} A_j(t) > N\,\mathsf{E}[A(t)]t + \alpha\right] \le e^{-\frac{2\alpha^2}{N(E(t))^2}} .$$

$$\Longrightarrow \quad \mathcal{G}(t;\varepsilon) = N\,\mathsf{E}[A(t)] + \sqrt{N}\,E(t)\,\sqrt{\tfrac{1}{2}\ln\tfrac{1}{\varepsilon}}$$

Example: Shaped Traffic

Traffic: $A_j \sim E_j$ (each flow is shaped) with $\rho_j = \lim_{t\to\infty} \frac{E_j(t)}{t}$

- $\mathsf{E}[A(t)] = \rho\, t$
- $\mathsf{Var}[A_j(t)] \leq \rho\ t\ \big(E(t) - \rho t\big)$
- $\mathsf{M}_A(\theta, t) \leq \underbrace{1 + \frac{\rho t}{E(t)}\left(e^{\theta E(t)} - 1\right)}_{\overline{M}_A(\theta,t)}$

Central Limit Theorem: $\mathcal{G}(t;\varepsilon) = N\rho t + \sqrt{N} z_{1-\varepsilon}\rho t\,\sqrt{\frac{E(t)}{\rho t} - 1}$

Chernoff bound: $\mathcal{G}(t;\varepsilon) = \inf_{\theta>0}\left\{\frac{1}{\theta}\left(N\ln\big(1 + \frac{\rho t}{E(t)}(e^{\theta E(t)} - 1)\big) - \ln\varepsilon\right)\right\}$

Hoeffding bound: $\mathcal{G}(t;\varepsilon) = N\,\rho\,t + \sqrt{N}\,E(t)\,\sqrt{\frac{1}{2}\ln\frac{1}{\varepsilon}}$

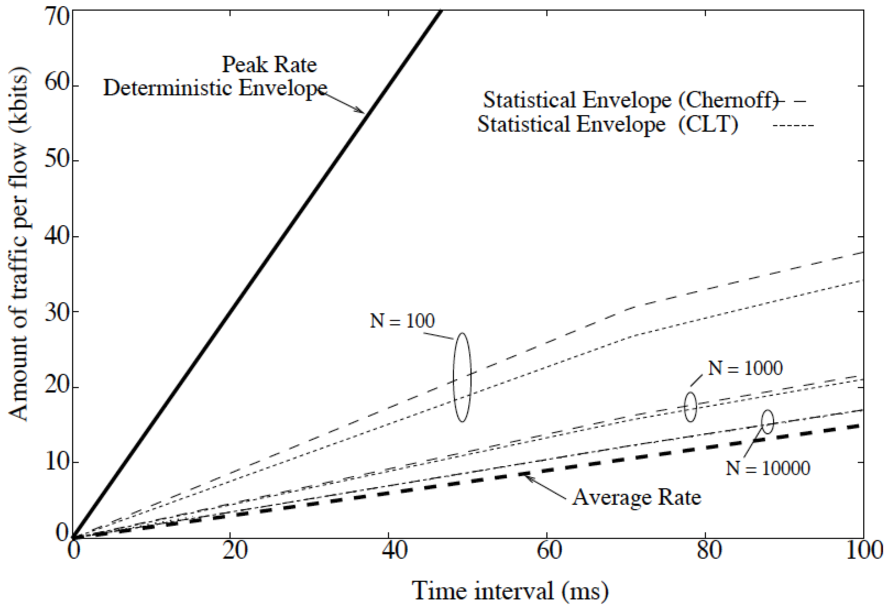
Example: Shaped Traffic

(BBLO00)

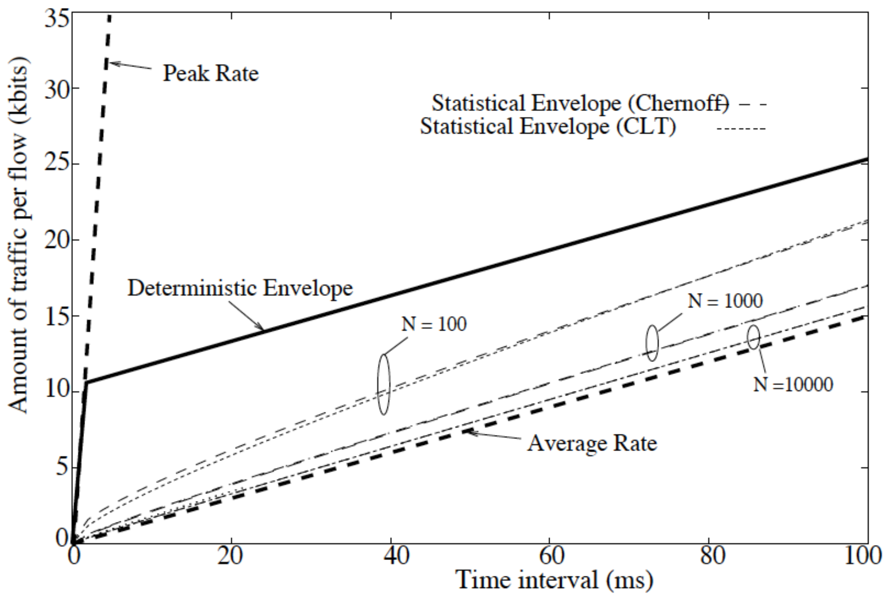
Dual token bucket: $E(t) = \min\{Pt\,, b + rt\} \quad (P > r).$

Type	Peak rate P (Mbps)	Mean rate r (Mbps)	Burst size b (bits)
1	1.5	0.15	95,400
2	6.0	0.15	10,345

$$\varepsilon = 10^{-6}$$



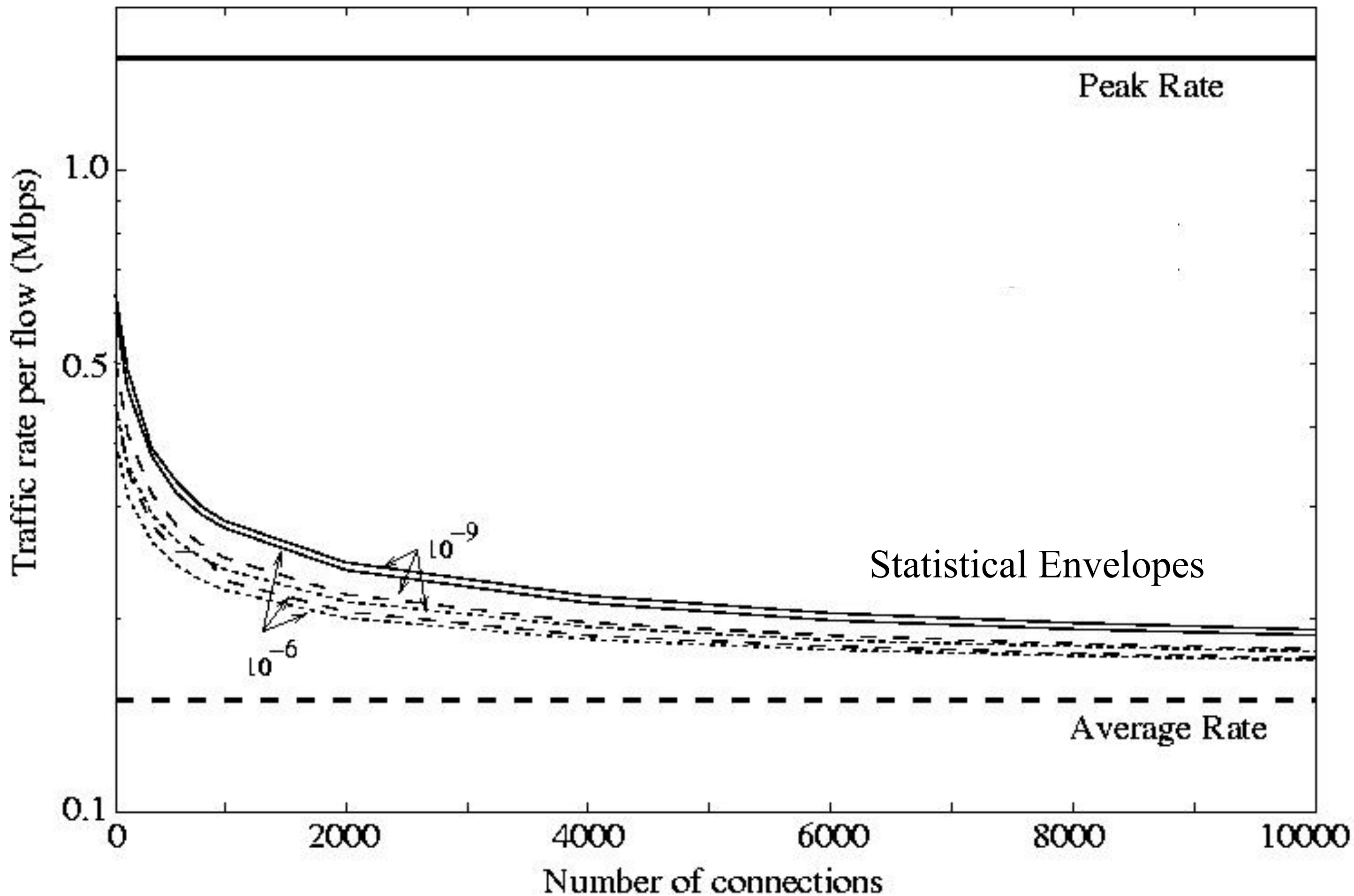
Type 1



Type 2

Statistical vs. Deterministic Envelopes *(BBLO00)*

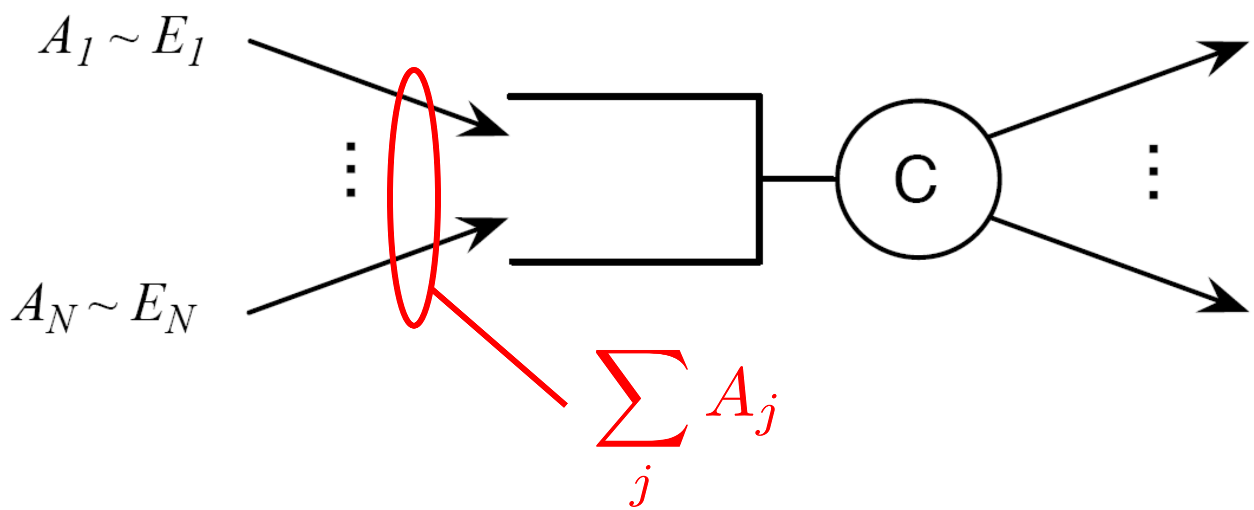
Traffic rate at t = 50 ms
Type 1 flows



Backlog Computation

(or: Why statistical envelopes are not enough)

Backlog Analysis



Next step: Use the statistical envelope for a probabilistic backlog analysis

Stochastic backlog bound: Encountering a problem

Recall Reich's backlog expression (for buffered link):

$$B(n) = \max_{0 \le k \le n} \{A(k,n) - C(n-k)\} \; .$$

Stochastic backlog bound:

$$\begin{aligned}
Pr\left[B(n) > x\right] &= Pr\left[\max_{0 \le k \le n} \{A(k,n) - C(n-k)\} > x\right] \\
&= Pr\left[\bigcup_{0 \le k \le n} \left(A(n-k) - C(n-k) > x\right)\right] \\
&\le \sum_{k=0}^{n} Pr\left[A(k) - Ck > x\right] \quad \textit{(union bound)} \\
&\le \sum_{k=0}^{n} \varepsilon(n)
\end{aligned}$$

- Good news: If $\mathcal{G}(k;\varepsilon) < Ck + x$, then $Pr\left(A(k) - Ck > x\right) < \varepsilon(k)$
- Bad news: $\sum_{k=0}^{n} \varepsilon(n)$ is unbounded (in general) for $k \to \infty$

Different ways to address the problem

- The problem (prev. slide) can be addressed differently, leading to a range of approaches:

$$
\begin{aligned}
Pr\left[B(n) > x\right] &= Pr\left[\max_{0\le k\le n}\{A(k,n) - C(n-k)\} > x\right] \\
&= Pr\left[\bigcup_{0\le k\le n}\left(A(n-k) - C(n-k) > x\right)\right] \\
&\le \sum_{k=0}^{n} Pr\left[A(k) - Ck > x\right] \qquad \textit{(union bound)} \\
&\le \sum_{k=0}^{n} \varepsilon(n)
\end{aligned}
$$

Abandon rigor: Principle of the largest term

Asymptotics only: effective bandwidth

Avoid union bound: martingale analysis

Use tail distribution of A: sample path envelope

Use distribution of A: moment generating function calculus

- Each approach presents a tradeoff (with pros and cons)

Principle of the largest term

$$\boxed{Pr\left[\max_{k\leq n} X(n) > x\right] \approx \max_{k\leq n} Pr\left[X(n) > x\right]}$$

- Rhs. is always lower bound for lhs.
- Rhs. is upper bound when it is justified that

$$\sum_{k\leq n} Pr\left[X(n) > x\right] \approx \max_{k\leq n} Pr\left[X(n) > x\right]$$

Then: Backlog bound $P(B(n) > b^{\varepsilon}) < \varepsilon$ with

$$b^{\varepsilon} = \max_{k\leq n}\left(\mathcal{G}(k;\varepsilon) - Ck\right)$$

Effective bandwidth (abbreviated)

Effective bandwidth: $\alpha(\theta) = \lim_{k\to\infty} \frac{1}{\theta k} \ln \mathsf{M}_A(\theta, k)$

Then:

$$\boxed{C > \alpha(\theta) \implies \exists K > 0 : \; Pr\left(B(n) > x\right) \leq K e^{-\theta x}}$$

- Provides exponential decay rate (but no constants)
- Avg. rate $\leq \alpha(\theta) \leq$ Peak rate
- Multiplexing of N independent flows $A_1, \ldots, A_N$ with effective bandwidths $\alpha_1(\theta), \ldots, \alpha_N(\theta)$:

$$\alpha(\theta) = \sum_{j=1}^{N} \alpha_j(n)\,.$$

Exponentially Bounded Burstiness (EBB)

Arrival process A has exponentially bounded burstiness (EBB) if

$$Pr\Big(A(s,t) > \rho(t-s) + \sigma\Big) < Me^{-\alpha\sigma} \ ,$$

for all $s \leq t$ and all $\sigma > 0$, with constants $M > 0$ and $\alpha > 0$.

- Interpretations:
  1. Probability of deviation from average rate ρ by more than σ has exponential decay
  2. Probabilistic token bucket
- EBB traffic is large class: Poisson, Markov-modulated On-Off, etc.
- EBB assumption is sufficient to obtain a stochastic backlog bound
- Other nice properties

Backlog Bound with EBB traffic

$$\begin{aligned}
Pr\left[B(n) > x\right] &= Pr\left(\max_{0\le k\le n}\{A(n) - A(k) - C(n-k)\} > x\right) \\
&= Pr\left(\bigcup_{0\le k\le n}\left(A(n-k) - C(n-k) > x\right)\right) \\
&\le \sum_{k=0}^{n} Pr\left(A(k) - Ck > x\right) \qquad \textit{(union bound)} \\
&= \sum_{k=0}^{n} Pr\left(A(k) > rk - (C-r)k + x\right) \\
&\le \sum_{k=0}^{\infty} Me^{-\alpha((C-r)k+x)} \\
&= \frac{Me^{-\alpha x}}{1 - e^{-\alpha(C-r)}}
\end{aligned}$$

Revisiting Formulation of Statistical Envelopes

- EBB definition: $\boxed{Pr(A(t) > \rho t + \sigma) < Me^{-\alpha\sigma}}$
- We write this as a statistical envelope:

Statistical envelope: $Pr\left[A(t) > \mathcal{G}(t,\varepsilon)\right] < \varepsilon$

In EBB: $\varepsilon = Me^{-\alpha\sigma} \Longleftrightarrow \sigma = \ln\left(M/\varepsilon\right)^{\alpha}$

We get: $\boxed{\mathcal{G}(t,\varepsilon) = \rho t + \ln\left(M/\varepsilon\right)^{\alpha}}$

- Thus there are two equivalent formulations:

(1) decay function

$$\mathcal{G}(t;\sigma) = \rho t + \sigma$$
$$\varepsilon(x) = Me^{-\alpha\sigma}$$

$\Leftrightarrow$

(2) fixed violation probability

$$\mathcal{G}(t,\varepsilon) = \rho t + \ln\left(M/\varepsilon\right)^{\alpha}$$

ε fixed

- We use both: (1) is more convenient for analysis, and (2) is convenient for illustrations

Statistical Envelope (Revised Formulation)

A statistical (traffic) envelope $\mathcal{G}$ for a random arrival process A with stationary increments satisfies for all $s \leq t$

$$Pr\left[A(t) > \mathcal{G}(t;x)\right] < \varepsilon(x)$$

where $\varepsilon(x)$ is decreasing in x.

Exponentially bounded burstiness (EBB)	$\mathcal{G}(t;x) = rt + x\,,$ $\varepsilon(x) = Me^{-\alpha x}$
Stochastically bounded burstiness (SBB)	$\mathcal{G}(t;\sigma) = rt + x\,,$ $\int_0^\infty \varepsilon(u)du < \infty$

Doob's inequality: Avoiding the union bound

Doob's inequality: If X_n is a supermartingale, then

$$Pr\left(\max_{0\le k\le n} X_k \ge x\right) \le \frac{\mathsf{E}[X_o]}{x}$$

$$\left.\begin{array}{c} A \text{ has independent increments} \\ C > 0 \text{ constant} \end{array}\right\} \Longrightarrow \begin{cases} Z(k) = A(n-k,n) - Ck \\ e^{\alpha Z(k)} \text{is a supermartingale} \end{cases}$$

$$\begin{aligned} Pr\left[B(n) > x\right] &= Pr\left[\max_{0\le k\le n} \{A(n-k,k) - C(k)\} > x\right] \\ &= Pr\left[\max_{0\le k\le n} e^{\alpha Z(k)} > e^{\alpha x}\right] \\ &\le \mathsf{E}[e^{\alpha Z(0)}]e^{-\alpha x} \qquad \textit{(Doob's inequality)} \\ &= e^{-\alpha x} \end{aligned}$$

The art lies in the construction of a suitable supermartingale (very difficult for general network elements)

Statistical Sample Path Envelopes (SSPEs)

(and Single Node Performance Bounds)

Need for statistical sample path envelope

- Deterministic envelope E satisfies

$$\forall 0 \leq s \leq t: \; A(s,t) \leq E(t-s)$$

- Statistical envelope $\mathcal{G}$ gives us

$$\forall 0 \leq s \leq t: \; Pr\left[A(s,t) \leq \mathcal{G}(t-s;\varepsilon)\right] \geq 1-\varepsilon$$

- Suppose we had an envelope $\overline{\mathcal{G}}$ with

$$Pr\left[\, \forall 0 \leq s \leq t: \; A(s,t) \leq \overline{\mathcal{G}}(t-s;\varepsilon)\right] \geq 1-\overline{\varepsilon}$$

A statistical sample path envelope (SSPE) of a random arrival process A is given by an envelope function $\overline{\mathcal{G}}$ and a decay function $\overline{\varepsilon}(x)$ such that for all $t \geq 0$

$$Pr\left[\sup_{0 \leq s \leq t} \left\{A(s,t) - \overline{\mathcal{G}}(t-s;\sigma)\right\} \geq 0\right] \leq \overline{\varepsilon}(\sigma)$$

Sample path envelopes make it easy!

Deterministic analysis:

$$B(n) = \max_{0 \le k \le n} \{A(k,n) - C(n-k)\} \le \sup_{k \le n} \{E(k) - Ck\}$$

Stochastic analysis:

$$\begin{aligned} B(n) &= \max_{0 \le k \le n} \{A(k,n) - C(n-k)\} \\ &\le \max_{0 \le k \le n} \{\overline{\mathcal{G}}(k;\sigma) - Ck\} \quad \text{(with probability) } 1 - \overline{\varepsilon}(x)) \end{aligned}$$

Hence: Backlog bound $P(B(n) > b^{\varepsilon}) < \varepsilon$ is given by

$$b^{\varepsilon} = \max_{k \le n} \left(\overline{\mathcal{G}}(k;\varepsilon) - Ck\right)$$

Benefits of sample bound benefits

- Sample path envelope allows us to keep probabilistic argument separate:
  1. Analyze backlog for a single sample path
  2. Bound the result with statistical sample path envelope
- Statistical sample path envelope is often pessimistic
- Note: A statistical sample path envelope is equivalent to a probabilistic backlog bound at a buffered link
- Statistical sample path envelope $\substack{\Rightarrow \\ \not\Leftarrow}$ Statistical envelope

Statistical Sample Path Envelope for EBB traffic

Statistical Envelope

$$\mathcal{G}(t;\sigma) = \rho t + \sigma$$
$$\varepsilon(\sigma) = Me^{-\alpha\sigma}$$

$\Rightarrow$

Statistical Sample Path Envelope

$$\overline{\mathcal{G}}(t;\sigma) = (\rho + \delta)t + \sigma$$
$$\overline{\varepsilon}(\sigma) = \frac{Me^{-\alpha\sigma}}{1-e^{-\alpha\delta}}$$

$\delta > 0$ arbitrary constant

$$
\begin{aligned}
& Pr\Big(\sup_{0\le s\le t} \{A(t-s,t) - \overline{\mathcal{G}}(s)\} > 0\Big) \\
& \qquad = Pr\Big(\sup_{0\le s\le t} \{A(t-s,t) - (\rho+\delta)s + \sigma\} > 0\Big) \\
& \qquad \le \sum_{s=0}^{\infty} Pr\big(A(t-s,t) \ge (\rho+\delta)s + \sigma\big) \quad \textit{(union bound)} \\
& \qquad \le \sum_{s=0}^{\infty} Me^{-\sigma\alpha}e^{-\alpha\delta s} = \frac{Me^{-\alpha\sigma}}{1-e^{-\alpha\delta}}.
\end{aligned}
$$

Different ways to obtain SSPE . . .

- . . . with the Principle of the Largest Term:

$$\boxed{Pr\left[\max_{0\leq s\leq t}\{A(s,t)\}\geq \mathcal{G}(t-s)\right]\approx \max_{0\leq s\leq t} Pr\left[A(s,t)\geq \mathcal{G}(t-s)\right]}$$

$$\Longrightarrow \qquad \overline{\mathcal{G}}(t;\varepsilon)=\mathcal{G}(t;\varepsilon)\,, \quad \overline{\varepsilon}=\varepsilon$$

- . . . using a Statistical Backlog Bound

Suppose we have $\boxed{Pr\left[B(t)\geq x\right]\leq \varepsilon(x) \quad , \forall x}$

Since $B(t)=\sup_{0\leq s\leq t}\{A(s,t)-C(t-s)\}$, we get

$$\Longrightarrow \quad \overline{\mathcal{G}}(t;x)=Ct+x\,, \quad \overline{\varepsilon}(x)=\varepsilon(x)$$

Different ways to obtain SSPE ...

- ... using Time Scale Bounds:

  Deterministic bound on busy period $\boxed{T = \max\{k \mid E(k) > Ck\}}$

$$
\begin{aligned}
Pr\left[\max_{0\le l\le n-1}\{A(l,n) - G(n-l;\varepsilon)\} \ge 0\right] &\le \sum_{j=0}^{n-1} Pr\Big[A(l,n) \ge G(n-l)\Big] \\
&\le \sum_{j=0}^{T} Pr\Big[A(k) \ge G(k)\Big] \\
&\le T\varepsilon
\end{aligned}
$$

$$
\Longrightarrow \quad \overline{\mathcal{G}}(k;\varepsilon) = G(k;\varepsilon)\,, \qquad \overline{\varepsilon} = T\varepsilon
$$

Different ways to obtain SSPE ...

- ... using Time-Decaying Violation Probabilities:

  Assume statistical envelope where violation probability decays with time: $\boxed{Pr\left[A(n-k,n) \geq G(k)\right] \leq \varepsilon(k)}$

  For example: $\boxed{\varepsilon(k) = \dfrac{\varepsilon}{1+k^2}}$

$$
\begin{aligned}
Pr\left[\max_{0 \leq k \leq n-1}\{A(n-k,n) - G(k;\varepsilon)\} \geq 0\right] &\leq \sum_{k=0}^{n-1} Pr\Big[A(n-k,n) \geq G(k
\\
&\leq \sum_{k=0}^{\infty} \frac{\varepsilon}{1+k^2} \\
&= \frac{\pi}{2}\varepsilon\,.
\end{aligned}
$$

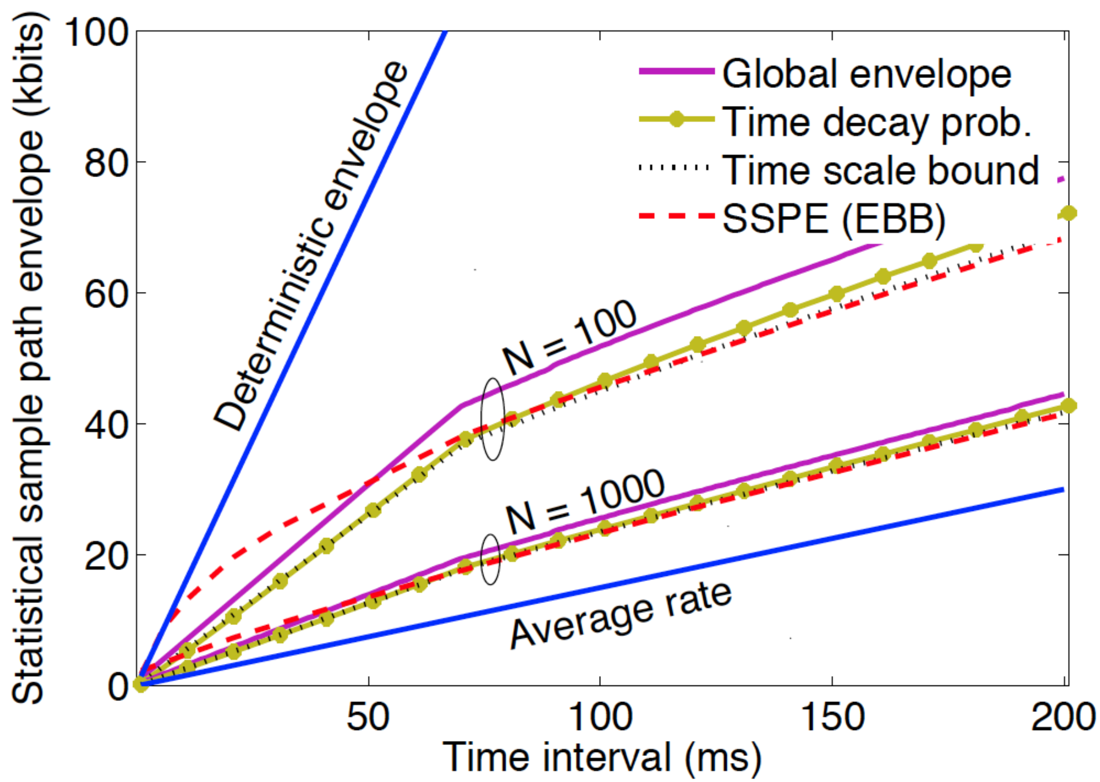
$$\Longrightarrow \quad \overline{\mathcal{G}}(k;\varepsilon) = G(k;\varepsilon)\,, \qquad \overline{\varepsilon} = \frac{\pi}{2}\varepsilon$$

Comparison: Statistical Sample Path Envelopes

Traffic: $A_j \sim E_j$ with $E(t) = \min\{Pt\,, \sigma + t\}$ with

$$P = 1.5 \text{ Mbps} \ \ \rho = 150 \text{ kbps} \ \ \sigma = 95,400 \text{ bits}$$

(There are additional parameters/choices !)



Different SSPEs ($\bar{\varepsilon} = 10^{-6}$)

Stochastic Bounds (Single Node)

Given:

- Statistical sample path envelope: $\overline{\mathcal{G}}(t;\sigma)\,,\ \varepsilon(\sigma)$
- Deterministic lower service curve: $S(t)$ with $D \geq A \otimes S$

1. Statistical envelope for departures:
   $\overline{\mathcal{G}} \oslash S(t;\sigma)$ with $\varepsilon(\sigma)$ is a statistical envelope of departures
2. Backlog bound: $b^*(\sigma) = \overline{\mathcal{G}} \oslash S(0;\sigma)$ with
   $$Pr\left(B(t) > b^*(\sigma)\right) < \varepsilon(\sigma)$$
3. Delay bound: $d^*(\sigma) = \inf\left\{d \geq 0 \,|\, \overline{\mathcal{G}} \oslash S(-d;\sigma) \leq 0\right\}$ with
   $$Pr\left(W(t) > d^*(\sigma)\right) < \varepsilon(\sigma)$$

Compare to deterministic bounds!

Proof of Backlog Bound

Pick: σ

Assume: $\forall s \le t : A(s,t) \le \overline{\mathcal{G}}(t-s;\sigma)$ (with prob. $\ge 1-\varepsilon(\sigma)$)

$$
\begin{aligned}
B(t) &= A(t) - D(t) \\
&= A(t) - \inf_{0\le s\le t} \{A(t-s) + \mathcal{S}(s)\} \\
&= \sup_{0\le s\le t} \{A(t-s,t) - \mathcal{S}(s)\} \\
&\le \sup_{s\ge 0} \{\overline{\mathcal{G}}(s;\sigma) - \mathcal{S}(s)\} \\
&= \overline{\mathcal{G}} \oslash \mathcal{S}(0;\sigma)
\end{aligned}
$$

This bound does not hold with probability $\le \varepsilon(\sigma)$

Note: Proof is essentially identical to deterministic case.

Statistical Service Curves

(and Single Node Performance Bounds)

Statistical Service Curve

Wanted: Probabilistic analogue to deterministic service curve

For arbitrary arrivals A and resulting departures D at a network system, a function $\mathcal{S}(\cdot,\varepsilon) \in \mathcal{F}_o$ is a *statistical service curve* with violation probability $\varepsilon > 0$ if for all t

$$P\left(D(t) < A \otimes \mathcal{S}(t,\varepsilon)\right) < \varepsilon$$

or

For arbitrary arrivals A and resulting departures D at a network system, a function $\mathcal{S}$ is a *statistical service curve* if for all t

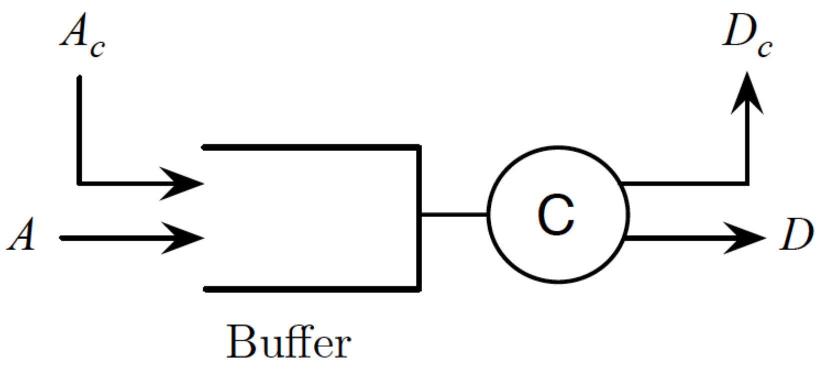
$$P\left(D(t) < A \otimes S(t,x)\right) < \varepsilon(x)$$

where $\varepsilon(x) > 0$ is a decreasing function.

Statistical service curve is a non-random stationary bound on available service

Stochastic Leftover Service Curve

Given: A fixed rate link with random cross traffic A_c.
Question: What is the available service to the flow with arrival A?



If cross traffic has arrival function A_c with SSPE $\overline{\mathcal{G}}(t, \varepsilon)$, a statistical service curve with violation probability ε for arrivals A is

$$\mathcal{S}(t; \varepsilon) = \left[Ct - \overline{\mathcal{G}}(t, \varepsilon)\right]^+$$

Note: Formulation involves a sample path argument.

Time-varying service

- Time-varying service: Available service in $[s, t)$ is $S(s, t)$
  - Service is not time-invariant: $S(s, t) \neq S(s + x, t + x)$
  - Service is additive: $S(s, t) = S(s, \tau) + S(\tau, t)$
  - Service can be deterministic or random
- Time-varying service requires $(\min, +)$ algebra on bivariate functions.

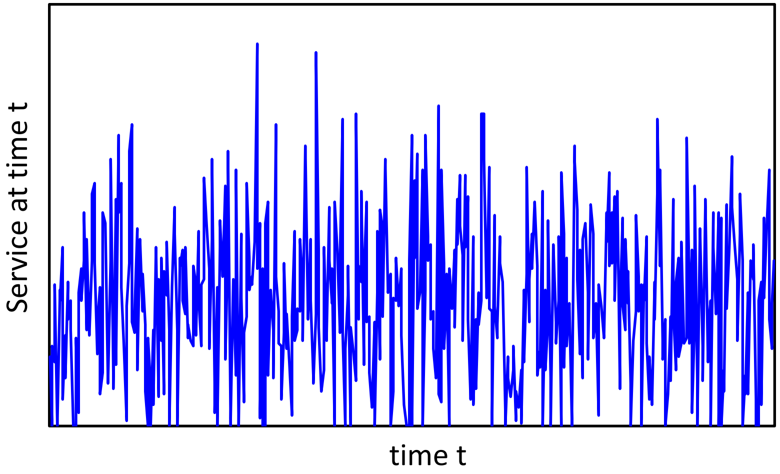
For arbitrary arrivals A and resulting departures D at a network system, a bivariate service process $S(s, t)$ is a *time-varying server*, if for all t,

$$D(t) \geq \inf_{0 \leq s \leq t} \{A(s) + S(s, t)\}$$

Random service

- Random Service: Available service $S(s,t)$ is a random variable such that:
    - Service is time-varying
- Consider a tail bound $\mathcal{S}(t)$ with decay function $\varepsilon(t)$ such that

$$Pr\Big(S(s,t) < \mathcal{S}(t-s)\Big) < \varepsilon(t-s)$$



Then: The tail bound $\mathcal{S}$ provides a statistical service curve with

$$P\Big(D(t) < A \otimes \mathcal{S}(t)\Big) < \sum_{k=1}^{t} \varepsilon(k)$$

Formulation involves a union bound.

Service with Exponentially Bounded Burstiness (EBB)

A bivariate service process S has exponentially bounded burstiness (EBB) if

$$Pr\Big(S(s,t) < \rho(t-s) - \sigma\Big) < Me^{-\alpha\sigma} \ ,$$

for all $s \leq t$ and all $\sigma > 0$, with constants $M > 0$ and $\alpha > 0$.

EBB service

$$\mathcal{S}(t;\sigma) = \rho t - \sigma$$
$$\varepsilon(x) = Me^{-\alpha\sigma}$$

$\Rightarrow$

Statistical Service Curve

$$\overline{\mathcal{S}}(t;\sigma) = (\rho - \delta)t + \sigma$$
$$\overline{\varepsilon}(\sigma) = \frac{Me^{-\alpha\sigma}}{1-e^{-\alpha\delta}}$$

$\delta > 0$ arbitrary constant

$$Pr\left(D(t) < \inf_{0\leq s\leq t}\{A(s) + \overline{\mathcal{S}}(t-s)\}\right) \leq \ldots\ldots$$

$$\leq \sum_{s=0}^{\infty} Pr(S(s,t) < (\rho-\delta)(t-s) + \sigma) \quad \textit{(union bound)}$$

$$\leq \sum_{s=0}^{\infty} Me^{-\sigma\alpha}e^{-\alpha\delta s} = \frac{Me^{-\alpha\sigma}}{1-e^{-\alpha\delta}}.$$

Stochastic Bounds (Single Node)

Given:

- Statistical sample path envelope: $\overline{\mathcal{G}}(t;\sigma_g)\,,\ \varepsilon_g(\sigma_g)$
- Statistical service curve: $\mathcal{S}(t;\sigma_s)\,,\ \varepsilon_s(\sigma_s)$
- Define: $\varepsilon(\underline{\sigma}) = \min_{\sigma_g+\sigma_s}\{\varepsilon(\sigma_g) + \ \varepsilon_s(\sigma_s)\}$

1. Statistical envelope for departures:
   $\overline{\mathcal{G}} \oslash \mathcal{S}(t;\underline{\sigma})$ with $\varepsilon(\underline{\sigma})$ is a statistical envelope of departures

2. Backlog bound: $b^*(\underline{\sigma}) = \overline{\mathcal{G}} \oslash \mathcal{S}(0;\underline{\sigma})$ with
   $$Pr\left(B(t) > b^*(\underline{\sigma})\right) < \varepsilon(\underline{\sigma})$$

3. Delay bound: $d^*(\underline{\sigma}) = \inf\left\{d \geq 0 \,|\, \overline{\mathcal{G}} \oslash \mathcal{S}(-d;\underline{\sigma}) \leq 0\right\}$ with
   $$Pr\left(W(t) > d^*(\underline{\sigma})\right) < \varepsilon(\underline{\sigma})$$

Compare to deterministic bounds!

Proof of Backlog Bound

Pick: σ_g, σ_s

Assume: $\forall s \le t : A(s,t) \le \overline{\mathcal{G}}(t-s;\sigma_g)$ (with prob. $\ge 1-\varepsilon_g(\sigma_g)$)

$D(t) \ge A \otimes \mathcal{S}(t;\sigma_s)$ (with prob. $\ge 1-\varepsilon_s(\sigma_s)$)

$$
\begin{aligned}
B(t) &= A(t) - D(t) \\
&= A(t) - \inf_{0\le s\le t}\{A(t-s) + \mathcal{S}(s;\sigma_s)\} \\
&= \sup_{0\le s\le t}\{A(t-s,t) - \mathcal{S}(s;\sigma_s)\} \\
&\le \sup_{s\ge 0}\{\overline{\mathcal{G}}(s;\sigma_g) - \mathcal{S}(s;\sigma_s)\} \\
&= \overline{\mathcal{G}} \oslash \mathcal{S}(0;\sigma_g,\sigma_s)
\end{aligned}
$$

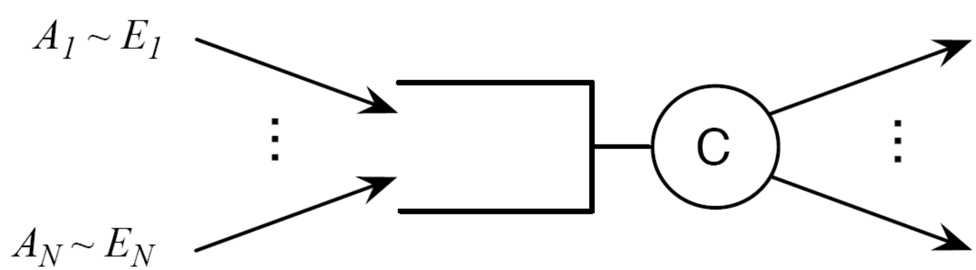
This bound does not hold with probability $\le \varepsilon_g(\sigma_g) + \varepsilon_s(\sigma_s)$

Note: Proof is essentially identical to deterministic case.

Example: Schedulability Analysis

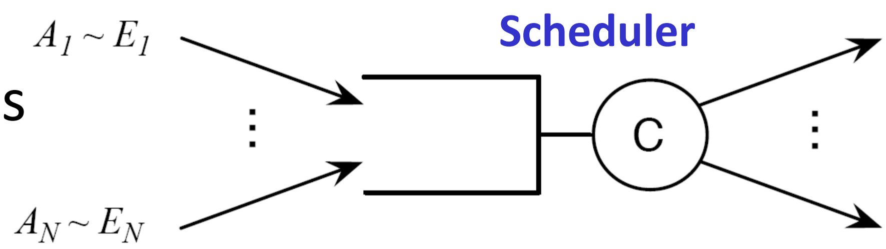
What is the maximum number of flows with delay requirements that can be put on a buffered link with and without considering statistical multiplexing?

- Link capacity C
- Each flows j has
  - arrival function A_j
  - envelope E_j
  - delay requirement d_j



Schedulability Analysis

- Work-conserving scheduler that serves flows from Q classes
- Class-q has delay bound d_q



Deterministic Service

Never a delay bound violation if:

$$d_q \geq \sup_s \frac{1}{C} \left\{ \sum_p E_p(s + \Delta_{qp}) - Cs \right\}$$

Statistical Service

Delay bound violation $\leq \varepsilon$ if:

$$d_q \geq \sup_s \frac{1}{C} \left\{ \sum_p \overline{\mathcal{G}}_p(s + \Delta_{qp}; \varepsilon) - Cs \right\}$$

Where

Δ_{qp} are scheduler-

E_p dependent constants (see [LGB11])

is det. envelope for all

$\overline{\mathcal{G}}_p(\cdot; \varepsilon)$ class-p flows

is SSPE of all class-p

Schedulability Analysis

(KWZL96)

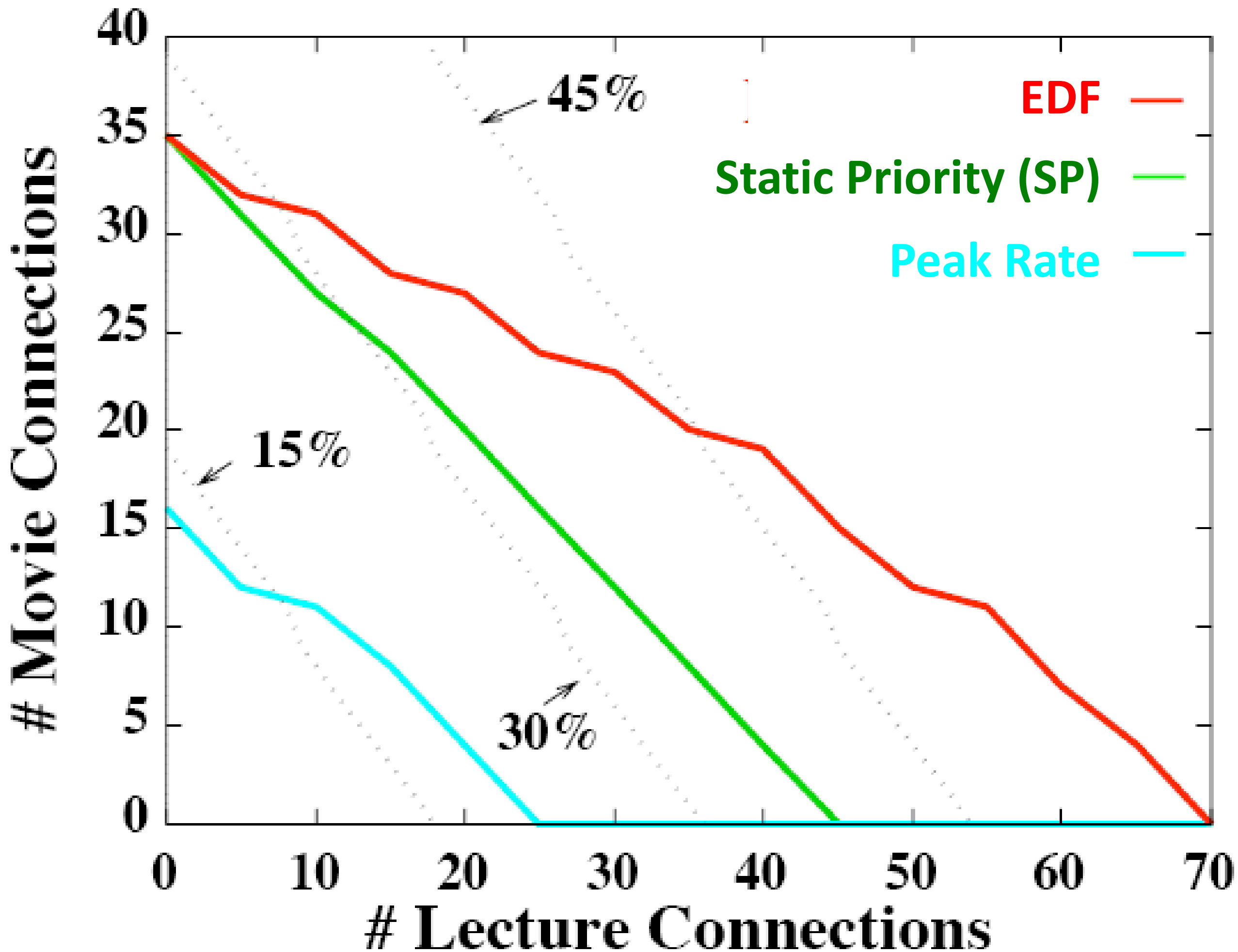
C = 45 Mbps

MPEG 1 traces:

Lecture:
d = 30 msec

Movie
(Jurassic Park):
d = 50 msec

Deterministic service

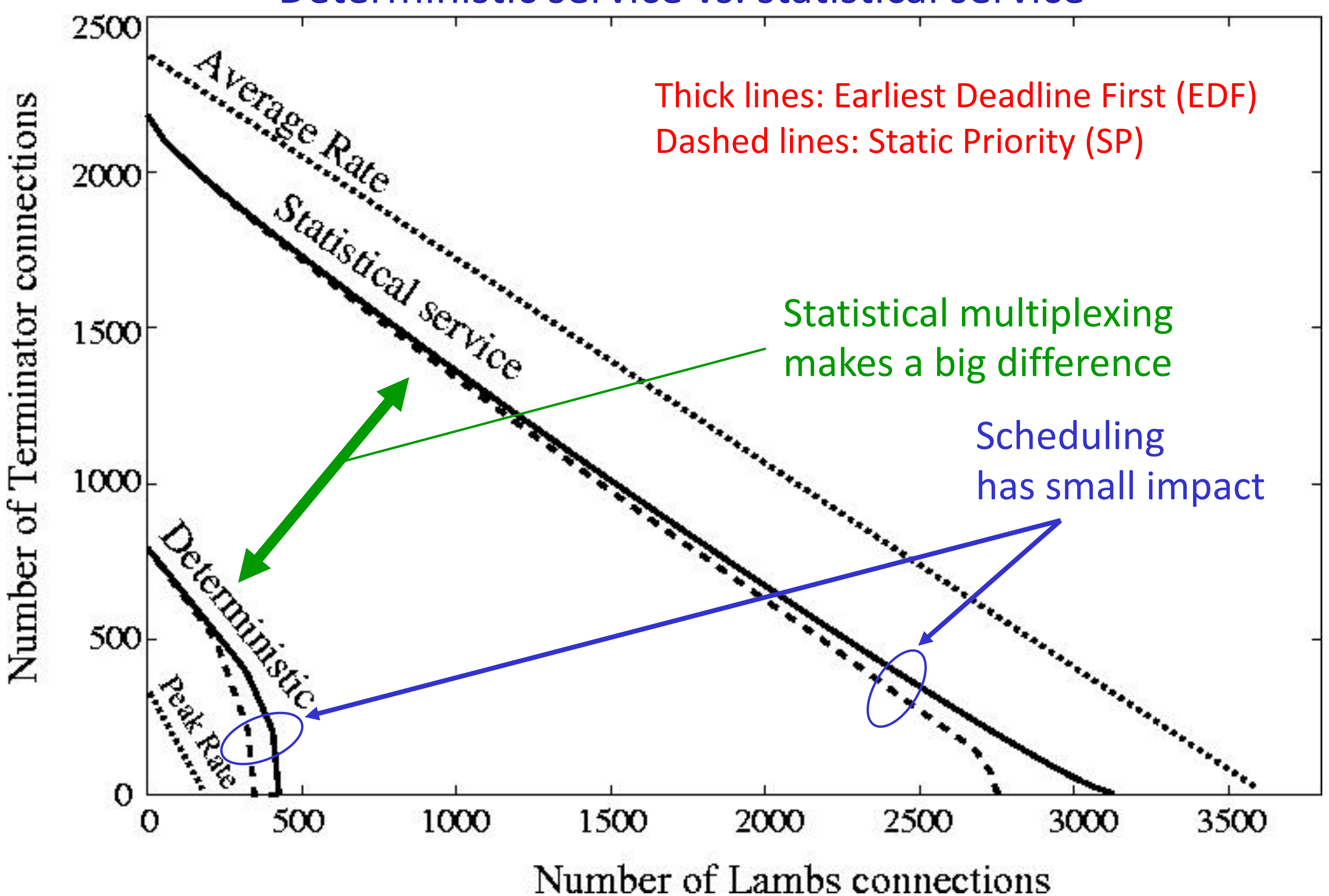


Schedulability Analysis

(BBLO00)

Example: MPEG videos with delay constraints at C= 622 Mbps

Deterministic service vs. statistical service



$d_{terminator}$=100 ms

d_{lamb}=10 ms

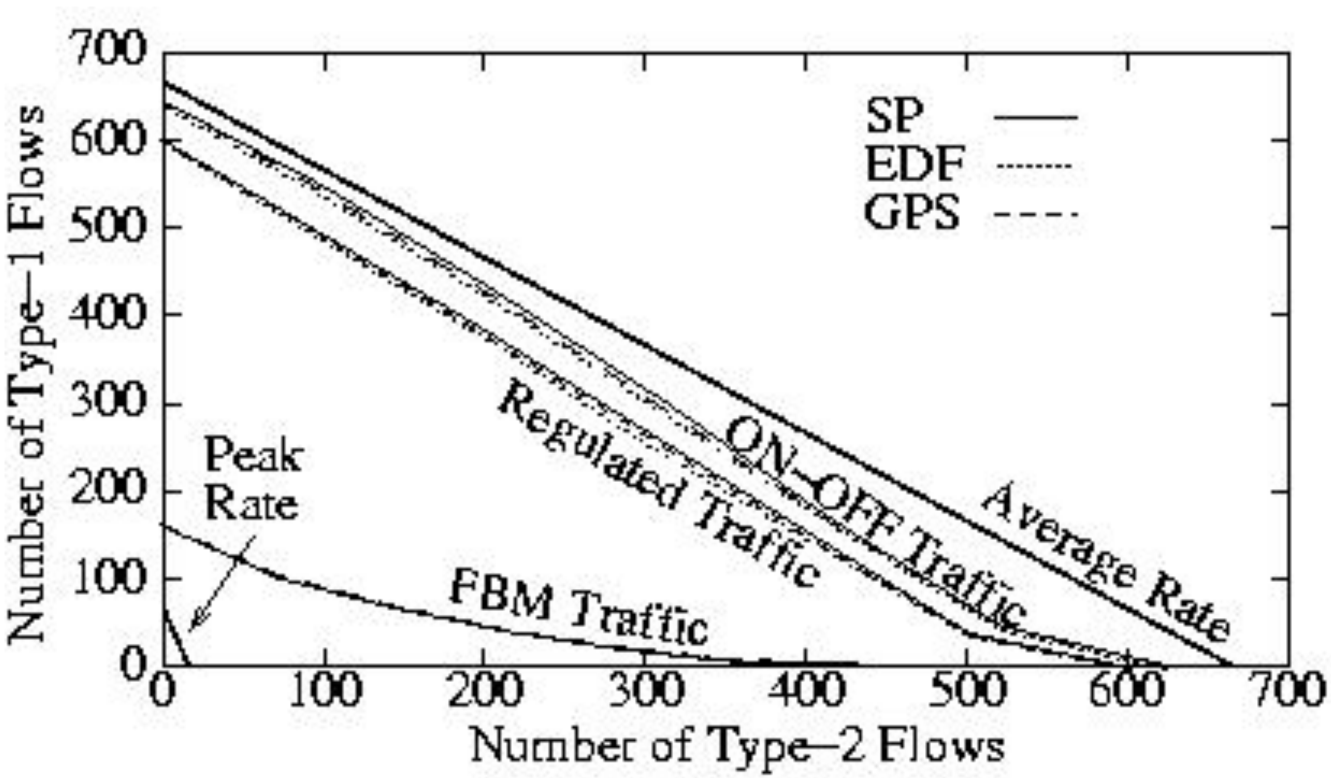
$\varepsilon = 10^{-6}$

Schedulability Analysis

(LBL07)

Different traffic types: - Regulated

- Memoryless On-Off
- Fractional Brownian Motion (FBM)



Schedulers:

SP- Static Priority

EDF – Earliest Deadline First

GPS – Generalized Processor Sharing

Traffic:

Regulated – 2-level token bucket

On-Off – On-off source

FBM – Fractional Brownian Motion

C= 100 Mbps, $\varepsilon = 10^{-6}$

Stochastic Network Calculus: End-to-End Performance Bounds

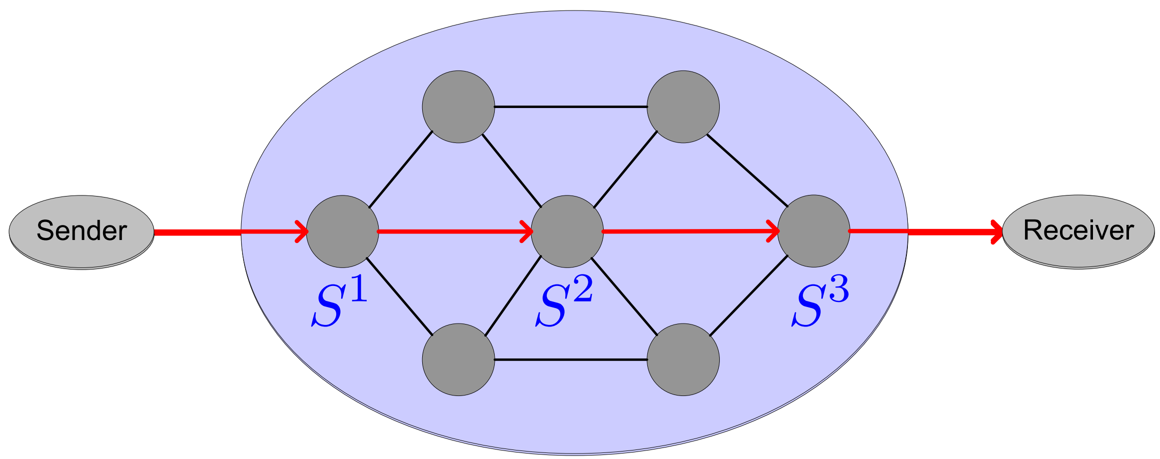
Stochastic Network Service Curve

Recall deterministic result:

If $S^1, S^2, \ldots, S^N$ describes the service at each node, then

$$S^{net} = S^1 \otimes S^2 \otimes \ldots \otimes S^N$$

describes the service given by the network as a whole.

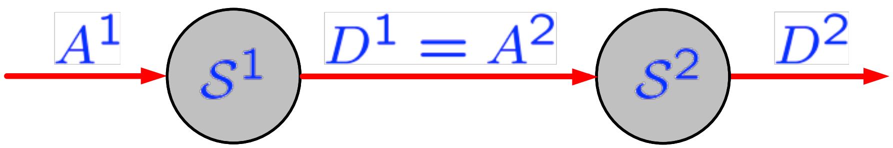


Stochastic Network Service Curve

- Problem: Given statistical service curves $\mathcal{S}^1, \mathcal{S}^2, \ldots, \mathcal{S}^N$ find statistical network service curve $\mathcal{S}^{net}$ so that

$$\mathcal{S}^{net} = \mathcal{S}^1 \otimes \mathcal{S}^2 \otimes \ldots \otimes \mathcal{S}^N$$

- Technical difficulty:



$$\begin{aligned} D^2(t) &\geq \inf_{0 \leq s \leq t} \{A^2(s) + \mathcal{S}^2(t-s)\} \quad (\text{with prob. } \varepsilon) \\ &= \{A^2(s_o) + \mathcal{S}^2(t-s_o)\} \\ &\leq A^1 \otimes \mathcal{S}^1(s_o) + \mathcal{S}^2(t-s_o)\} \quad (\cancel{\text{with prob.}}\ \varepsilon) \\ &= A^1 \otimes \mathcal{S}^1 \otimes \mathcal{S}^2(t) \end{aligned}$$

s_o is a random variable!

Stochastic Network Service Curve

- In early 2000s, finding a stochastic network service curve was an open problem for a few years:
  - Some incorrect solutions made it to publication
  - Some solutions were quite restricted (e.g., [BLP06])

- The following result - first presented at ACM Sigmetrics 2005 - was general enough to enable numerical computations under non-trivial probabilistic assumptions

Statistical Network Service Curve

(CBL06)

- Stat. service curve at node n: $\mathcal{S}^n(t,\sigma) = [\mathcal{S}^n(t) - \sigma]^+, \ \varepsilon^n(\sigma)$

  with $\sum_{j=1}^{\infty} \varepsilon^n(\sigma^n) < \infty$

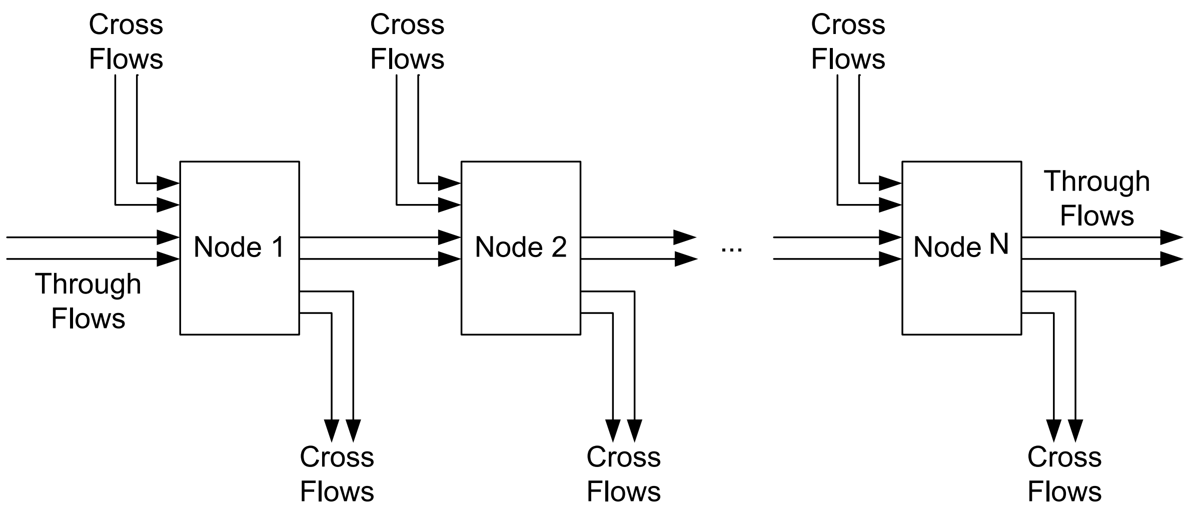
- <u>Define</u>: $\mathcal{S}_{-\delta}(t) = \mathcal{S}(t) - \delta t$

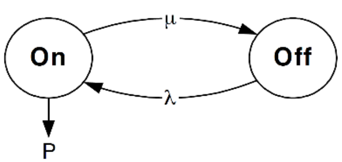
<u>Theorem</u>: If $\mathcal{S}^1, \mathcal{S}^2, \ldots, \mathcal{S}^H$ are statistical service curves, then for any $\gamma > 0$, $\mathcal{S}^{net}(t,\sigma) = [\mathcal{S}^{net}(t) - \sigma]^+$ is a statistical network service curve with

$$\mathcal{S}^{net}(t) = \mathcal{S}^1 \otimes \mathcal{S}^2_{\gamma} \otimes \ldots \otimes \mathcal{S}^N_{(N-1)\gamma}(t)$$

$$\varepsilon^{net}(\sigma) = \inf_{\sum_{n=1}^{N} \sigma^n = \sigma} \left\{ \varepsilon^N(\sigma^N) + \sum_{n=1}^{N} \sum_{j=1}^{\infty} \varepsilon^n(\sigma^n + j\gamma) \right\}$$

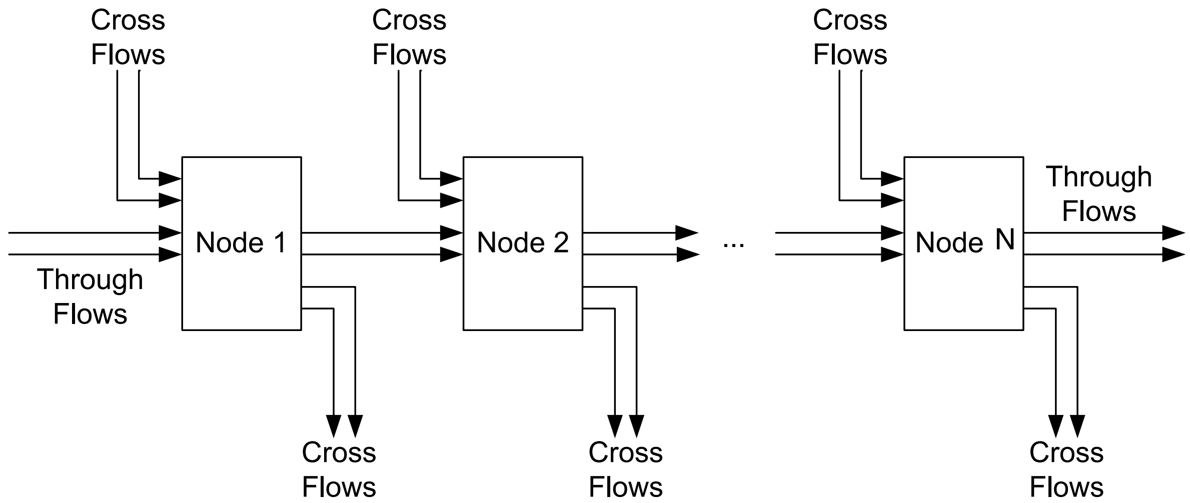
Example: Scaling of Delay Bounds





- Traffic: EBB model of Markov Modulated On-Off
- All links have capacity C
- Same cross-traffic **(not independent!)** at each node with SSPE $\overline{\mathcal{G}}_c(t,\sigma)$
- Service curve of through flows at node n: (through traffic has lower priority) $\mathcal{S}^n(t,\sigma) = [Ct - \overline{\mathcal{G}}_c(t,\sigma)]^+$

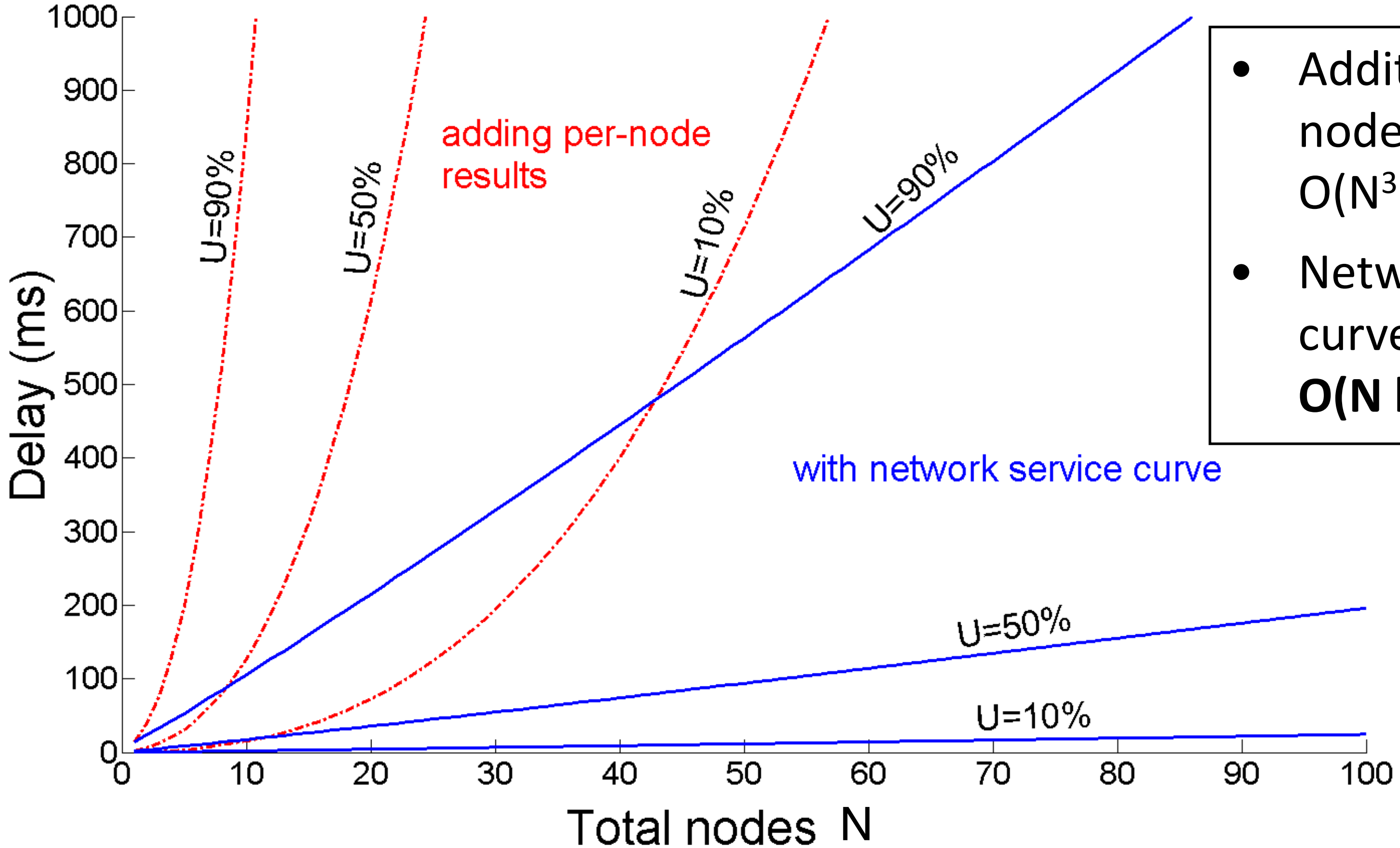
Example: Scaling of Delay Bounds



- Two methods to compute delay bounds:
  1. Network service curve:
     Compute single-node delay bound with statistical network service curve
  2. Add per-node bounds:
     Compute delay bounds at each node and sum up

Example: Scaling of Delay Bounds *(CBL06)*

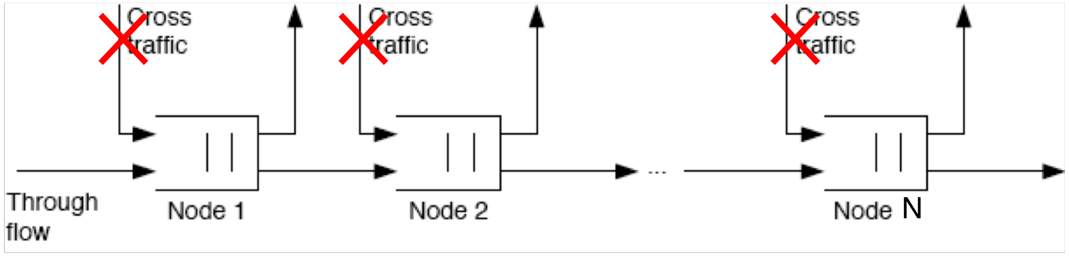
- Peak rate: P = 1.5 Mbps
  Average rate: r = 0.15 Mbps
- T= $1/\mu + 1/\lambda$ = 10 ms
- C = 100 Mbps
- Cross traffic = through traffic
- $\varepsilon = 10^{-9}$



- Addition of per-node bounds grows $O(N^3)$
- Network service curve bounds grow **O(N log N)**

Lower Bound on E2E Delay are Ω(N log N) *(BLC11)*

- M/M/1 queues with identical exponential service at each node



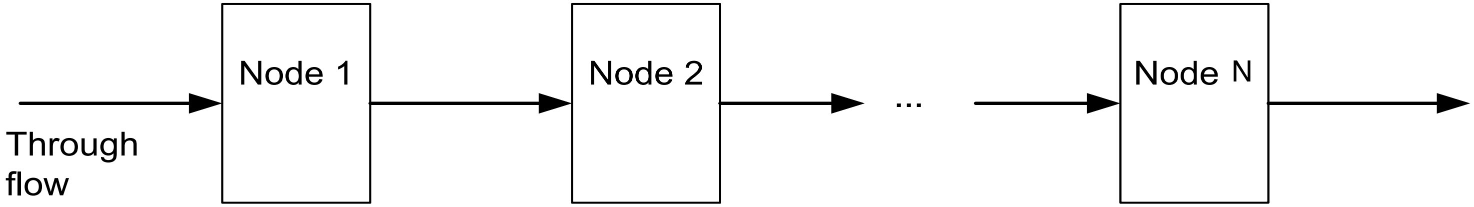
Theorem: E2E delay W^{net} satisfies for all $0 < z < 1$

$$Pr\left(W^{net} \leq \gamma_1 N \log(\gamma_2 N)\right) \leq z$$

Corollary: z-quantile $w_N(z)$ of W_H satisfies

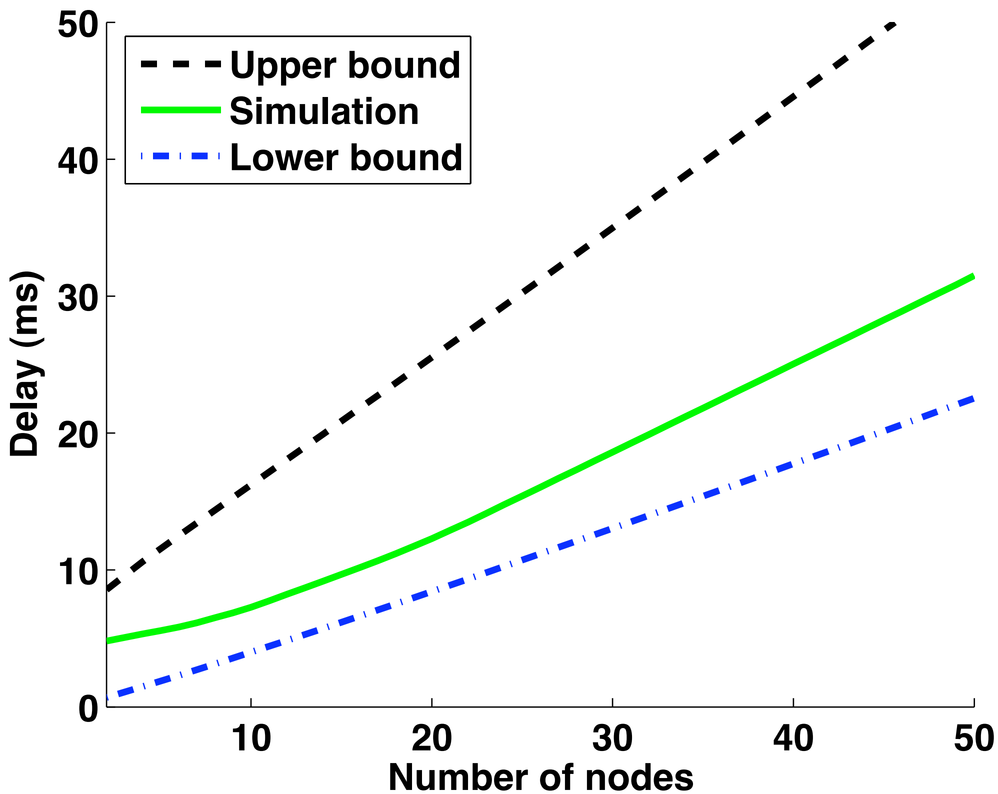
$$w_N(z) = \Omega(N \log N)$$

Upper and Lower Bounds on E2E Delays



- Tandem network without cross traffic
- Node capacity: C
- Arrivals are compound Poisson process
  - Packet arrival rate: λ
  - Packet size: $Y_i \sim exp(\mu)$
- Utilization: $\rho = \lambda/(\mu C)$

Upper and Lower Bounds on E2E Delays *(BLC11)*



Capacity

$C = 100\ Mbps$

Mean packet size

$\frac{1}{\mu} = 400$ Bytes

Load factor

$\rho = 90\%$

Violation probability

$\varepsilon = 10^{-6}$

End-to-End Performance Bounds for Heavy-tailed Traffic

(briefly)

Heavy-Tailed Self-Similar Traffic

- A heavy-tailed process X satisfies

$$\boxed{Pr(X(t) > x) \sim Kx^{-\alpha}}$$

$1 < \alpha < 2$

- A self-similar process satisfies

$$\boxed{X(t) \sim_{dist} a^{-H} X(at)}$$

$a > 0$

$H \in (0, 1)$ Hurst Parameter

htts Traffic Envelope

- Heavy-tailed self-similar (htss) envelope:

$$Pr(A(s,t) > \underbrace{r(t-s) + \sigma(t-s)^H}_{\mathcal{G}(t-s;\sigma)}) \leq \underbrace{K\sigma^{-\alpha}}_{\varepsilon(\sigma)}$$

- Main difficulty: Backlog and delay bounds require sample path envelopes of the form

$$Pr(\sup_{s \leq t}\{A(s,t) - \overline{\mathcal{G}}(t-s;\sigma)\} > 0) \leq \varepsilon(\sigma)$$

- Key contribution (not shown):
Derive sample path bound for htss traffic

Example: Node with Pareto Traffic

(LBC12)

Traffic parameters:

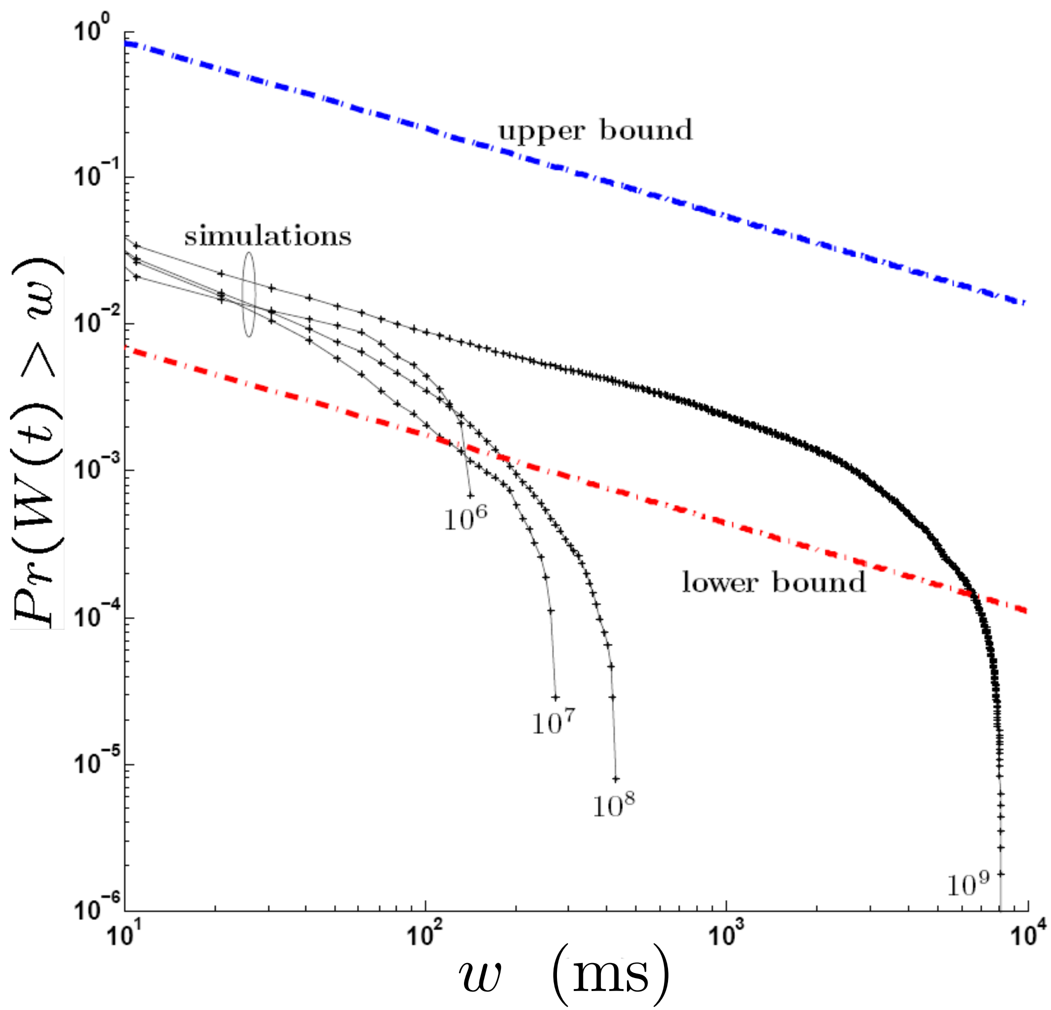
$$\alpha = 1.6$$
$$b = 150\ Byte$$
$$\lambda = 75\ Mbps$$

Node:

- Capacity C=100 Mbps with packetizer
- No cross traffic

Compare:

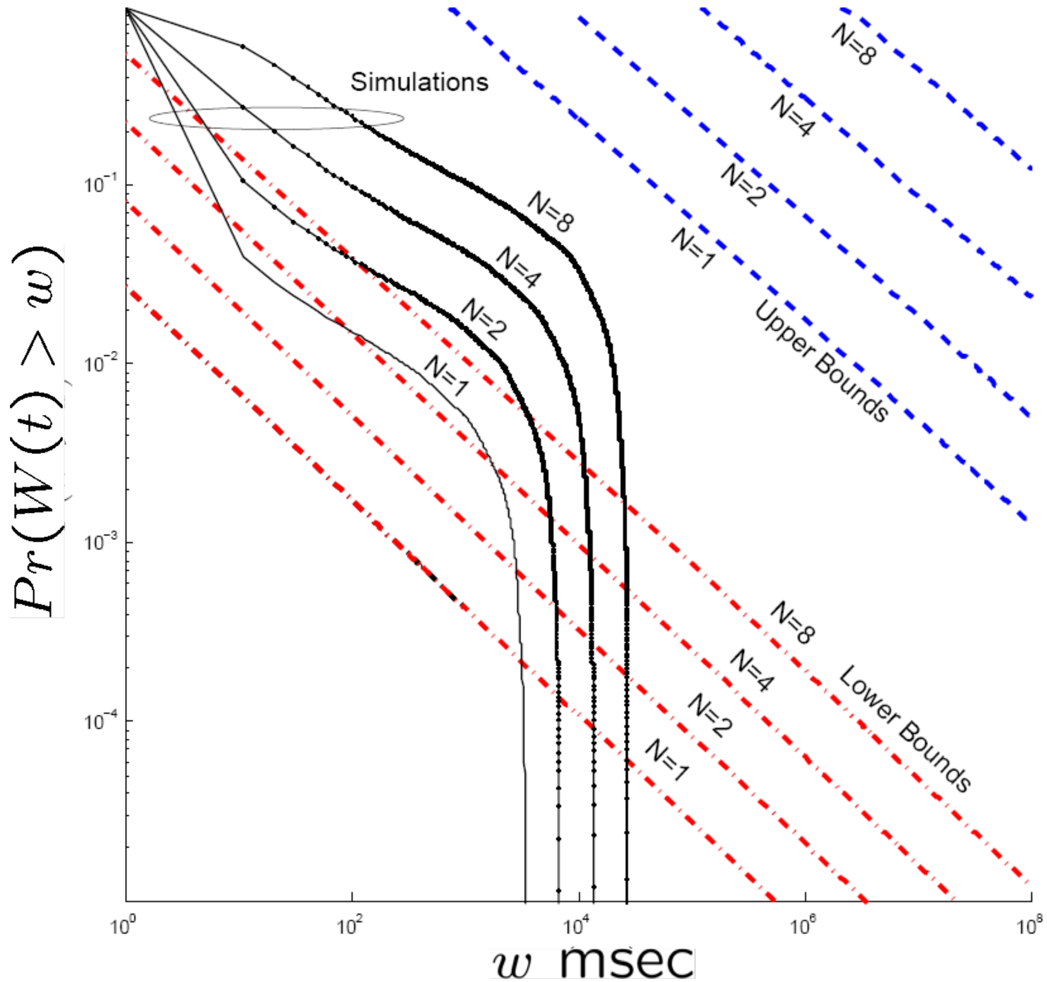
- Upper bound
- Lower bound
- Simulations of sample paths



Example: Nodes with Pareto Traffic (End-to-end) *(LBC12)*

Number of nodes:

$$N = 1, 2, 4, 8$$



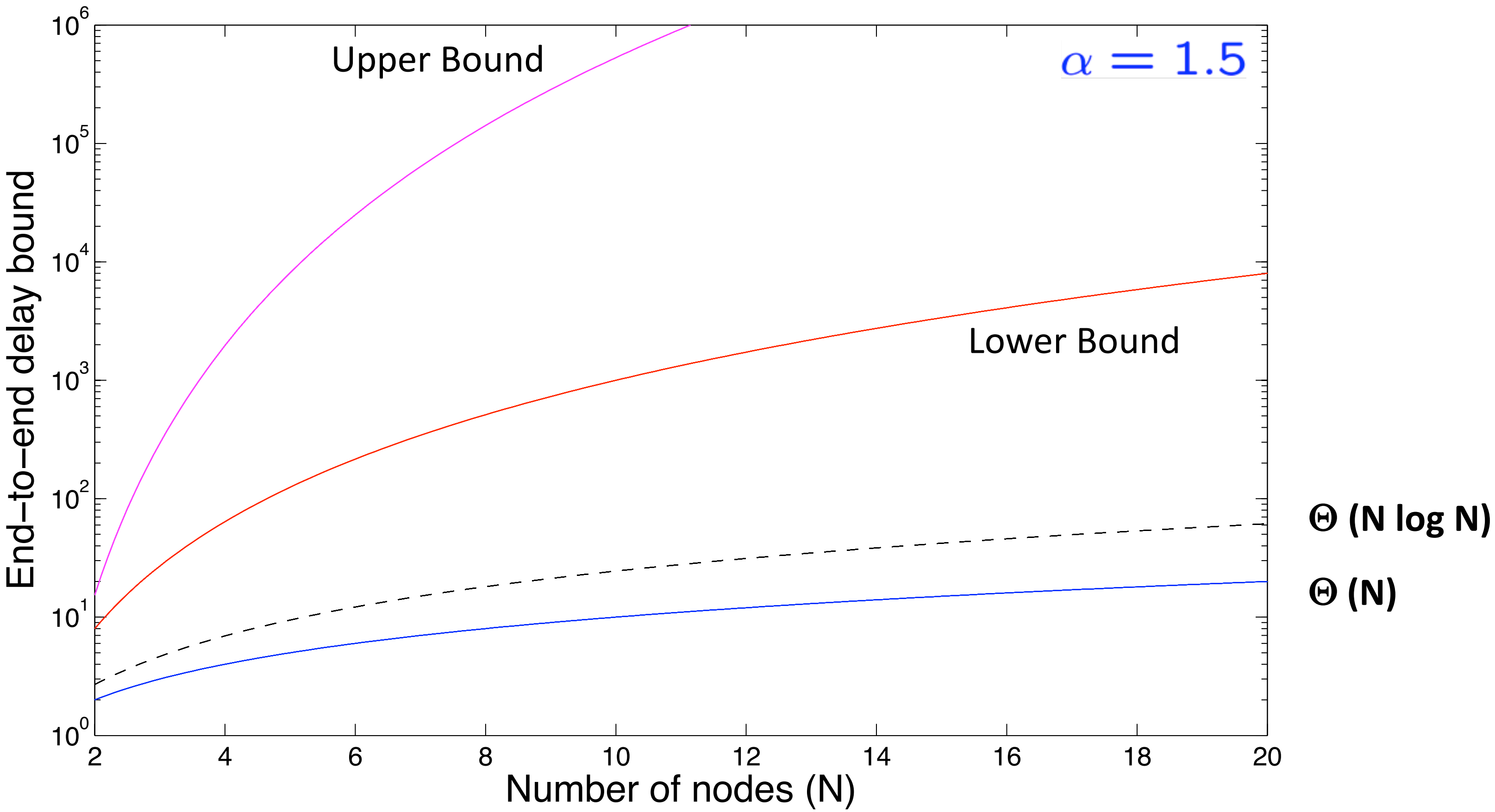
Compared with:

- Upper bound
- Lower bound
- Simulation of sample paths with 10^8 packets

Illustration of scaling bounds

End-to-end delays of htts traffic:

- Upper Bound: $O(N^{\frac{\alpha+1}{\alpha-1}}(\log N))^{\frac{1}{\alpha-1}})$
- Lower Bound: $\Omega(N^{\frac{\alpha}{\alpha-1}})$



Moment Generating Function Network Calculus

(briefly)

Moment Generating Function (MGF) Network Calculus

- An alternative method to derive stochastic bounds without constructing envelopes for traffic or service
- Assumes independence of arrivals and service
  - MGF of random variable (RV) X: $M_X(\theta) = E[e^{\theta X}]$
  - For two independent RVs X and Y: $M_{X+Y}(\theta) = M_X(\theta) M_Y(\theta)$
- Describes arrivals, departures, and service as bivariate functions
  - $A(s,t)$ are arrivals in time interval $[s,t)$
  - $D(s,t)$ are random departures in $[s,t)$
  - $S(s,t)$ is the random available service in $[s,t)$
- Derivations require network calculus for bivariate functions

Network Calculus for bivariate functions

Arrivals: $A(s,t)$
Departures: $D(s,t)$
Time-varying server: $D(0,t) \geq A \otimes S(0,t)$
Backlog: $B(t) = A(0,t) - D(0,t)$
Delay: $W(t) = \inf\{s > 0 \,|\, A(0,t) \leq D(0,t+s)\}$

$$f \otimes g\,(s,t) = \min_{s \leq \tau \leq t}\{f(s,\tau) + g(\tau,t)\}$$
$$f \oslash g(s,t) = \max_{0 \leq \tau \leq s}\{f(\tau,t) - g(\tau,s)\}$$

1. Envelope for departures:
   $D(s,t) \leq A \oslash S(s,t)$
2. Backlog bound:
   $B(t) \leq A \oslash S(t,t)$
3. Delay bound: $W(t) \leq \inf\{d \geq 0 \,|\, A \oslash S(t+d,t) \leq 0\}$

Network service curve:

$S_1 \otimes S_2 \otimes \ldots \otimes S_N(s,t)$ is the time-varying server of a sequence of time-varying servers with $S_1, S_2, \ldots, S_N$

Note: Network calculus for bivariate functions has weaker properties.

E.g., $f \otimes g(s,t) \neq g \otimes f(s,t)$

Moment Generating Function (MGF) Network Calculus

- For independent bivariate processes f and g

$$M_{f\otimes g}(-\theta,s,t) \le \sum_{\tau=s}^{t} M_f(-\theta,s,\tau)M_g(-\theta,\tau,t)$$

$$M_{f\oslash g}(\theta,s,t) \le \sum_{\tau=0}^{s} M_f(\theta,\tau,t)M_g(-\theta,\tau,s)\,.$$

With the Chernoff bound, this yields:

$$Pr\left(f\otimes g(s,t) < Y(s,t)\right) < \varepsilon$$

$$Pr\left(f\oslash g(s,t) < Z(s,t)\right) < \varepsilon$$

with

$$Y(s,t) = \min_{\theta>0}\frac{1}{\theta}\left\{\log\Big(\sum_{\tau=s}^{t} M_f(-\theta,s,\tau)M_g(-\theta,\tau,t)\Big) - \log\varepsilon\right\}$$

$$Z(s,t) = \min_{\theta>0}\frac{1}{\theta}\left\{\log\Big(\sum_{\tau=0}^{s} M_f(\theta,\tau,t)M_S(-\theta,\tau,s)\Big) - \log\varepsilon\right\}$$

Moment Generating Function (MGF) Network Calculus

$$\boxed{M_A(\theta, s, t) = E[e^{\theta A(s,t)}]\ ,}\ \boxed{M_S(-\theta, s, t) = E[e^{-\theta S(s,t)}]}.$$

$Pr(B(t) > b^*(t)) < \varepsilon$ with

$$b^*(t) = \min_{\theta > 0} \frac{1}{\theta} \left\{ \log \left(\sum_{\tau=0}^{t} M_A(\theta, \tau, t) M_S(-\theta, \tau, t) \right) - \log \varepsilon \right\}$$

$Pr(D(s,t) > D^*(s,t)) < \varepsilon$ with

$$D^*(s,t) = \min_{\theta > 0} \frac{1}{\theta} \left\{ \log \left(\sum_{\tau=0}^{s} M_A(\theta, \tau, t) M_S(-\theta, \tau, s) \right) - \log \varepsilon \right\}$$

$Pr(W(t) > w^*(t)) < \varepsilon$ with

$$w^*(t) = \min_{\theta > 0} \frac{1}{\theta} \left\{ \log \left(\sum_{\tau=0}^{t+d} M_A(\theta, \tau, t) M_S(-\theta, \tau, t+d) \right) - \log \varepsilon \right\}$$

Moment Generating Calculus *(F06)*

- Due to independence assumptions, the end-to-end bounds of backlog and delays in a tandem network increase linearly in the number of does → $\Theta(N)$

- Compare this this with the $\Theta(N \log N)$ scaling, when arrivals and service are correlated

A Few Remarks on Stochastic Network Calculus

Remarks: Stochastic Network Calculus vs. Queueing Theory

- Comparison is quite pointless, but often done
- Queueing theory …
  - … works with exact descriptions of arrival and service processes
  - … frequently seeks derivations of exact results
  - … frequently relies on independence of underlying processes
  - … frequently analyzes single node systems
  - … difficulty of analysis increases quickly beyond Poisson traffic and exp. service
- Stochastic Network Calculus …
  - … works with bounds on arrival and service processes
  - … seeks derivation of good bounds
  - … generally applies to correlated arrival and service processes
  - … analysis frequently extends to multi-node systems
  - … analysis of complex processes is often tractable

→ Stochastic Network Calculus extends the scope analytically tractable models, by giving up on achieving exact results

Remarks: Application of union bound

- Union bound is a crude tool for estimating unions of events
- Stochastic network calculus applies union bound widely
- Union bound wants to be used smartly:
  - As "bound of last resort", not to be used when other bounds are available
  - Bounds are often satisfactory within certain parameter ranges, but deteriorate outside such ranges
  - "Bad cases" are easily constructed, but sometimes due to inadequate parameter choices
- Alternatives to the union bound:
  - If analysis is supported by measurement statistics, union bound is not needed
  - Additional independence assumptions in the Martingale analysis allow to replace union bound by Doob's inequality (recently good progress, but only for single node systems)

Remarks: Lower bounds

- Most stochastic network calculus studies derive upper bounds on performance metrics
  - This is motivated by the desire to investigate a probabilistically relaxed worst-case scenario …
  - … but leaves open the accuracy of the derived bounds

- Deriving lower bounds complements the analysis and addresses questions about accuracy (see this slide set)

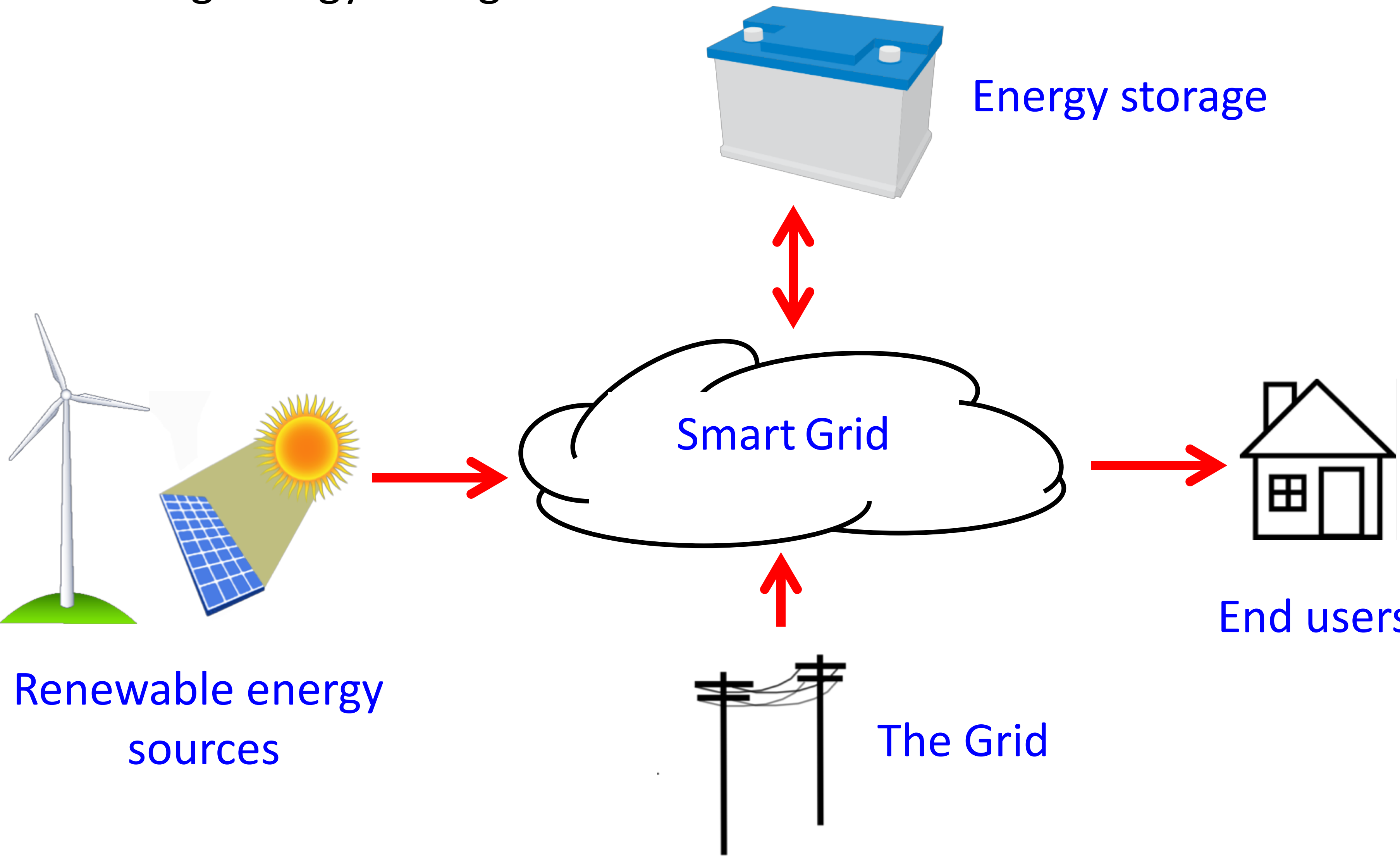
- Many opportunities exist in in studying lower bounds

APPLICATION AREAS

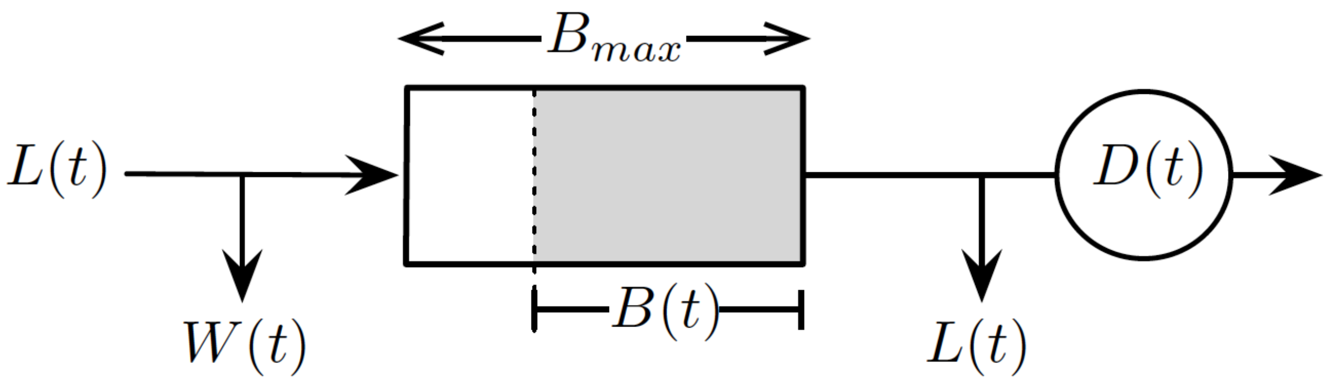
Smart Grids Storage Systems

Smart Grid Storage

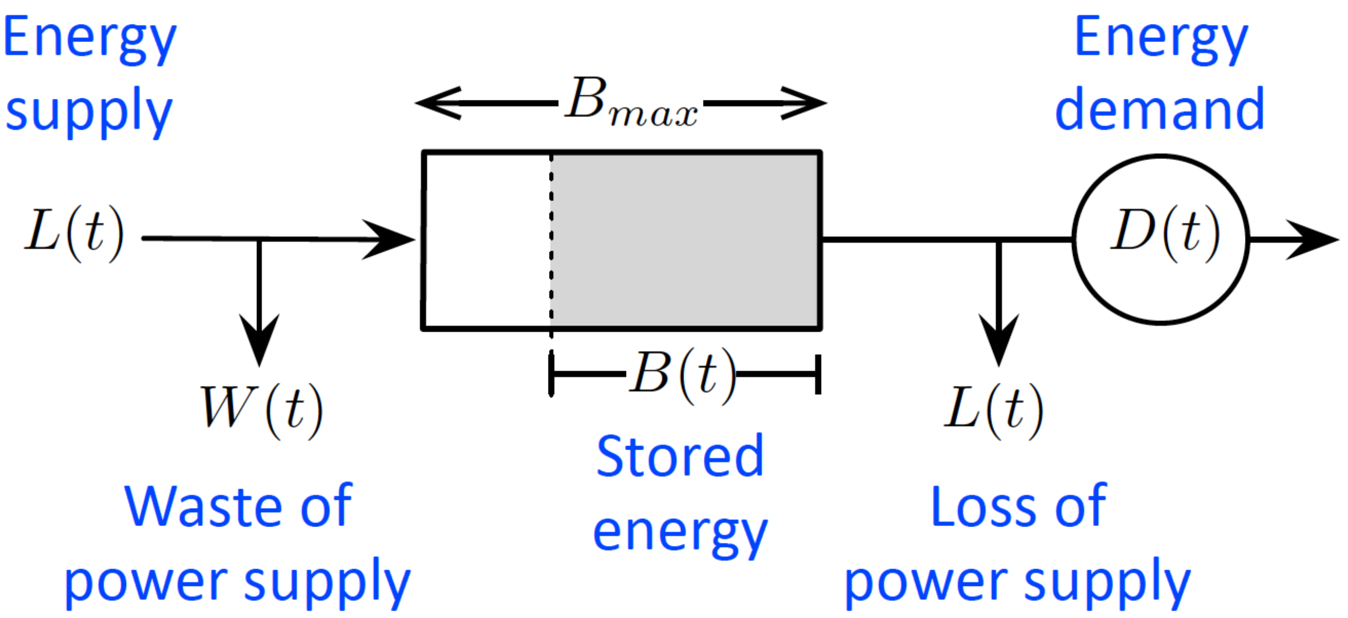
- **Problem:** Match time-varying energy demand and energy supply using energy storage



Smart Grid Storage



Smart Grid Storage



- Loss of power supply $B(t) < \text{Threshold } \underline{\text{and}} \ S(t) > 0$
- Waster of power supply: $B(t) = B_{max} \ \underline{\text{and}} \ S(t) > 0$

Smart Grid Storage

(WCLL12)

Use:

Backlog equation with loss by Cruz/Liu:

$$B(t) = \min_{0 \le s \le t} \left\{ \max_{s \le u \le t} \{ S(u,t) - D(u,t), S(s,t) - D(s,t) + B_{max}[t]^+ \} \right\}$$

To obtain:

Loss of Power Supply:

$$L(t) = \max_{0 \le s \le t-1} \left\{ \min_{s \le u \le t-1} \left[D(u,t) - S(u,t), D(s,t) - S(s,t) - B_{max}[t]^+ \right]^+ \right\}$$

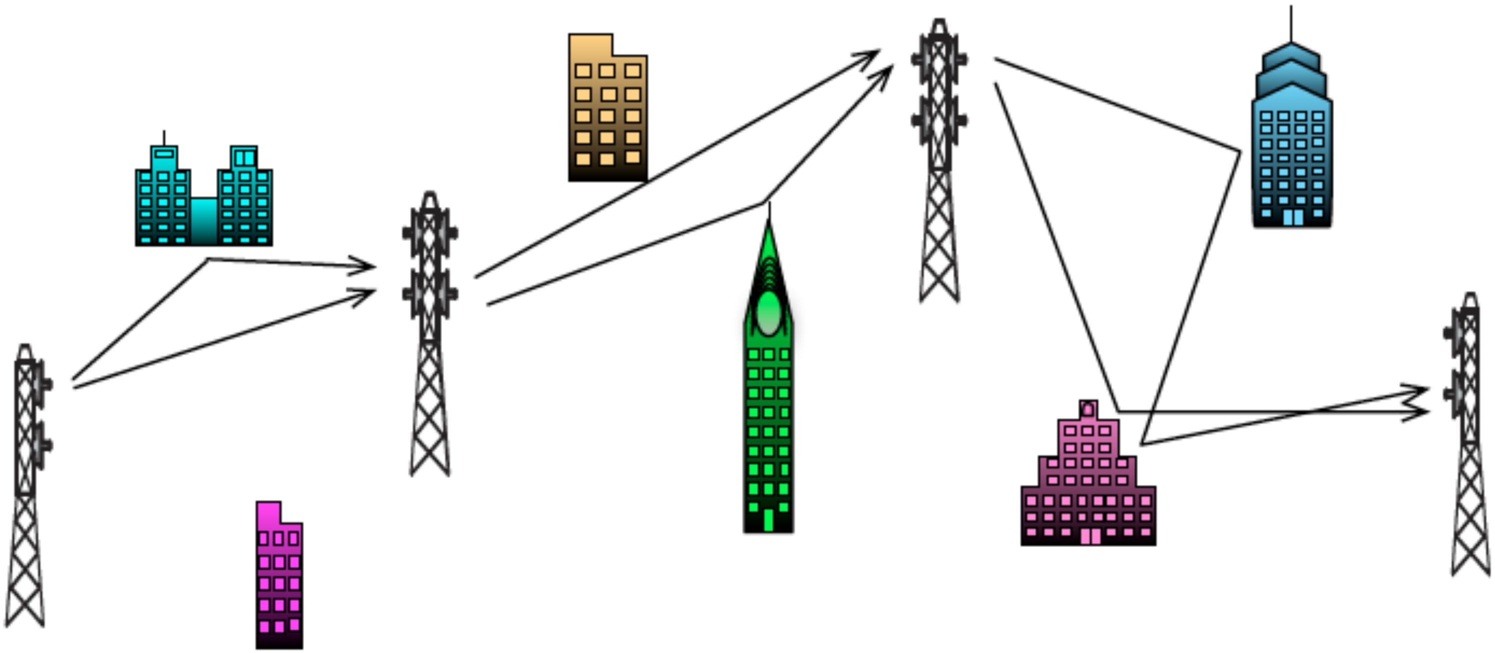
Waste of Power Supply:

$$W(t) = \min_{0 \le s \le t-1} \left\{ \max_{s \le u \le t-1} \{ S(u,t) - D(u,t) - B_{max}[t]^+, S(s,t) - D(s,t) + B_{max} - B_{max}[t]^+]^+ \right\}$$

Then create stochastic bounds with SSPE of S and D.

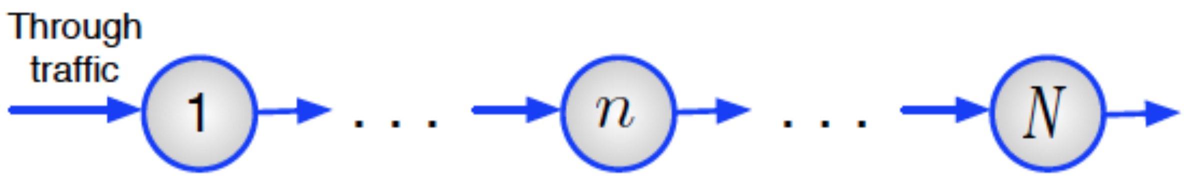
Wireless Networks
(Multihop multiaccess networks)

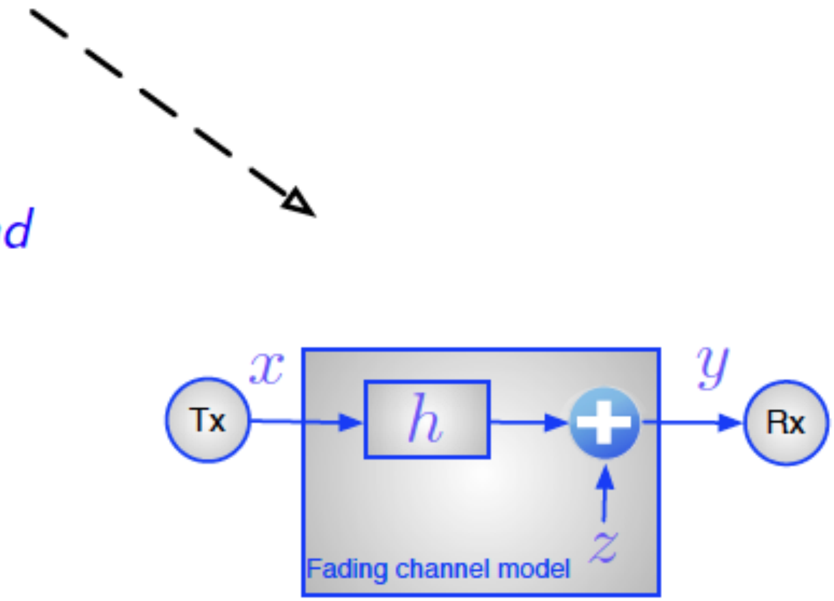
Multihop Fading Channels



- Transmission over a cascade of wireless channels with cross traffic/interference

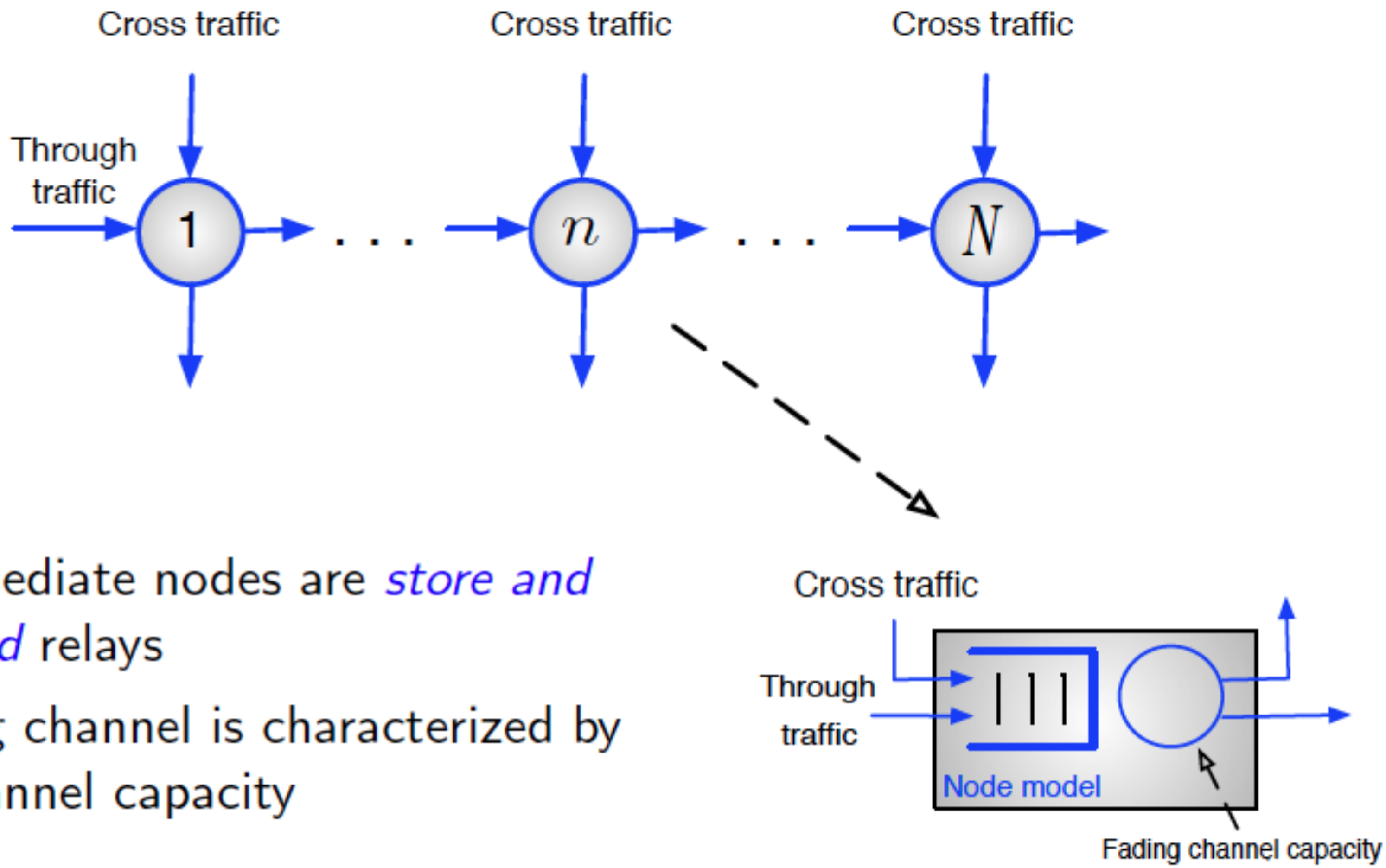
Multihop Wireless Network Model (Physical Layer)





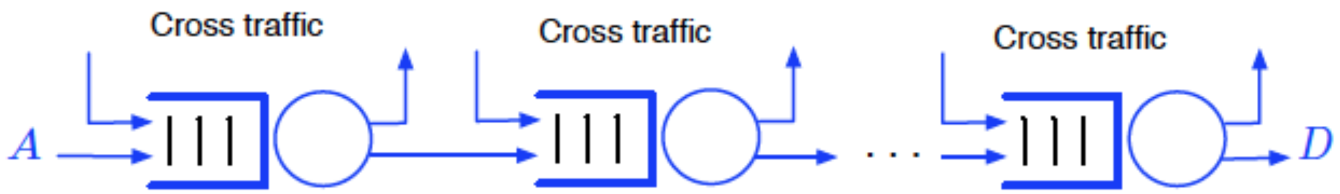
- Intermediate nodes are *amplify and forward* relays
- Fading channel model expresses variation of channel
  - Rician
  - Nakagami-m
  - Rayleigh

Multihop Wireless Network Model (Network Layer)



- Intermediate nodes are *store and forward* relays
- Fading channel is characterized by its channel capacity

Network Model



- Time-varying channel capacity:
  - Shannon: $\boxed{C(\gamma_t) = W\log(1+\gamma_t)}$
  - Fading channel: $\boxed{C(\gamma_t) = W\log(g(\gamma_t))}$
- W: channel bandwidth (in Hz)
- γ_t: signal-to-noise ratio (SNR) at time t
- I.i.d. cross traffic at each node

Service and Traffic Elements in the SNR Domain

- Service process in $[\tau, t)$:

$$S(\tau, t) = \sum_{i=\tau}^{t-1} \log g(\gamma_i) \qquad \Longrightarrow \text{ hard!}$$

- Exponentiated service process

$$\mathcal{S}(\tau, t) = e^{S(\tau,t)} = \prod_{i=\tau}^{t-1} g(\gamma_i) \qquad \Longrightarrow \text{ simpler}$$

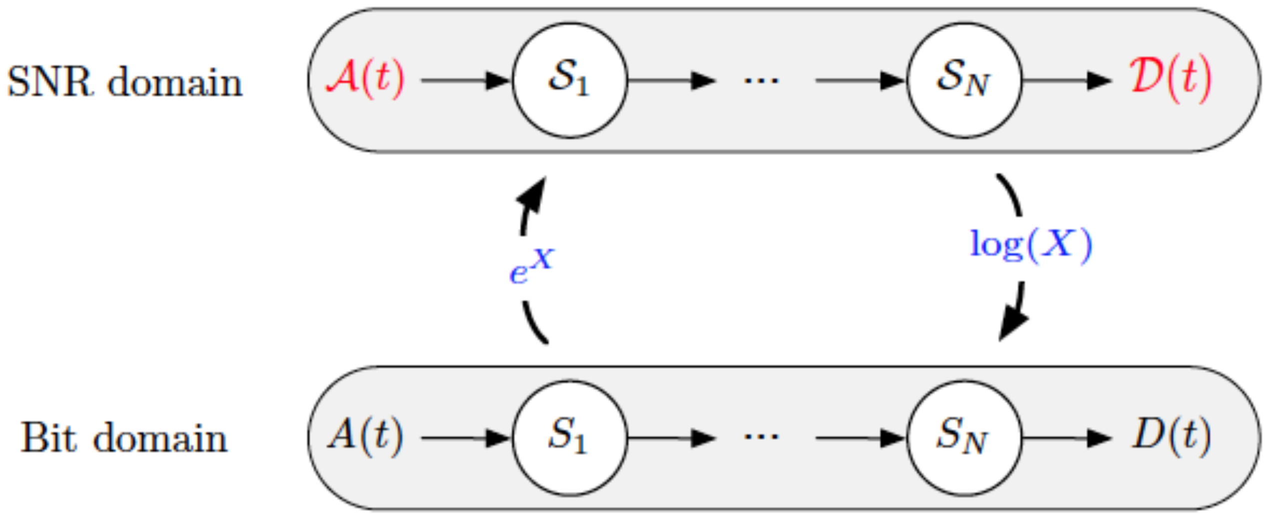
- Exponentiate arrivals and departures as well:

$$\mathcal{A}(\tau, t) = e^{A(\tau,t)} \quad \text{and} \quad \mathcal{D}(\tau, t) = e^{D(\tau,t)}$$

$\Longrightarrow$ Create a transfer domain where all processes are exponentiated

SNR Domain

(ZLB14)



- SNR service process: $\mathcal{S}(\tau, t) = \prod_{i=\tau}^{t-1} g(\gamma_i)$ $\rightarrow$ more tractable
- SNR arrival process: $\mathcal{A}(\tau, t) = e^{A(\tau,t)}$

$\Longrightarrow$ SNR domain is governed by $(\min, \times)$ dioid algebra

$\Longrightarrow$ Compute stochastic bounds in SNR domain, and transfer results back

Multiaccess Channels

R. Gallager (1985): A Perspective on Multiaccess Channels:
"For the last ten years there have been at least three bodies of research on multiaccess channels, each proceeding in virtual isolation from the others and each using totally different models."

Gallager's list:

1. collision resolution
2. multiaccess information theory
3. spread spectrum

Revised list:

1. Multiaccess information theory
2. Random Access
3. Dynamic scheduling

Service Characterization in Transfer Domain

(ZL14)

1. Multiaccess information theory
    - physical layer
    - permits concurrent transmissions with no coordination [Tse and Hanly 1998]

$$\boxed{\mathcal{S}_j^{IT}(\tau,t)=\prod_{u=\tau}^{t-1} g_j(\gamma_u)=\prod_{u=\tau}^{t-1}\left(1+\frac{1}{m(u)}\sum_{i=1}^{m(u)}\frac{|h_i|^2 p_i}{N_0 W}\right)}$$

2. Random access
    - MAC layer
    - collision when two or more users transmit simultaneously [Ciucu 2011]

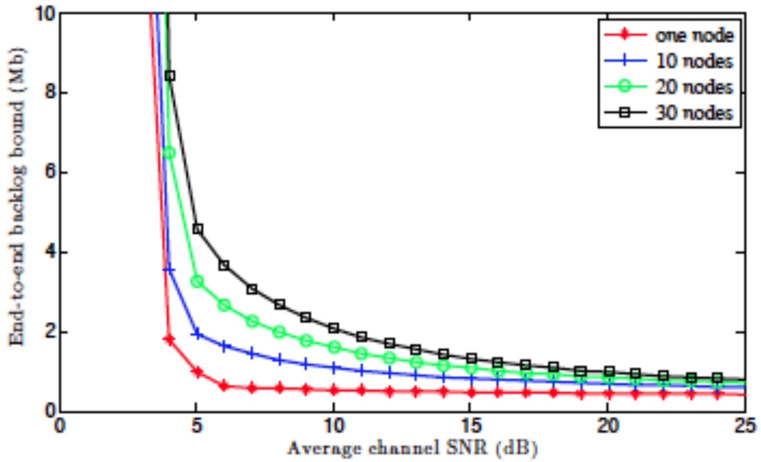
$$\boxed{\mathcal{S}_j^{RA}(\tau,t)=\frac{\prod_{u=\tau}^{t-1} g(\gamma_u)}{\prod_{u=\tau}^{t-1}[g(\gamma_u)]^{V_j(u)}}=\prod_{u=\tau}^{t-1}[g(\gamma_u)]^{1-V_j(u)}}$$

where $V_j(u)$ is virtual interference process.

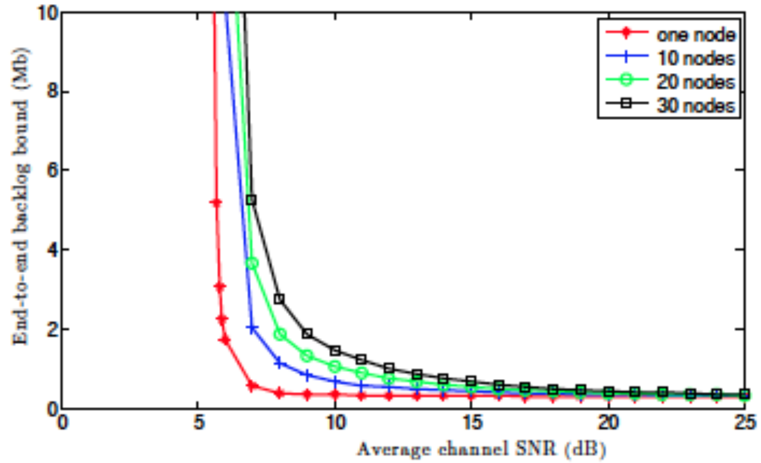
3. Dynamic scheduling (opportunistic scheduler)
    - link/network layer
    - centralized scheduler makes decision which user can transmit

$$\boxed{S_j^{OS}(\tau,t)=\prod_{u=\tau}^{t-1}[g(\gamma_u^{\max})]^{1/m}\triangleq\prod_{u=\tau}^{t-1} g_j(\gamma_u^{\max})}$$

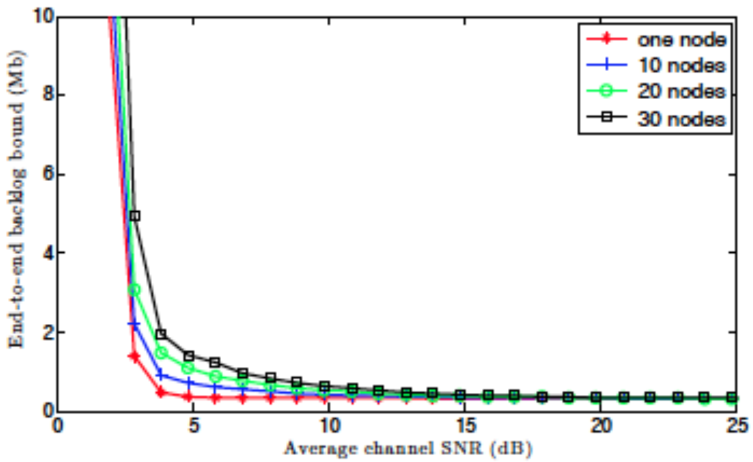
Backlog Bounds for N multiaccess Rayleigh Channels *(ZL14)*



Information–theoretic



Random access



Opportunistic scheduler

Literature and References

Literature

Deterministic Network Calculus

- C.-S. Chang. *Performance guarantees in communication networks.* Springer Verlag, 2000.
- J. Le Boudec and P. Thiran. *Network calculus: a theory of deterministic queuing systems for the Internet.* Lecture Notes in Computer Science 2050. Springer, 2001.
- J. Liebeherr, Draft chapters (upon request).

Stochastic Network Calculus

- Y. Jiang and Y. Liu. *Stochastic network calculus.* Springer, 2008.
- F. Ciucu, Scaling Properties in the Stochastic Network Calculus, Ph.D. Thesis, University of Virginia, 2007.
- Y. Ghiassi-Farrokhfal, Network Performance Analysis of Packet Scheduling Algorithms, Ph.D. Thesis, University of Toronto, 2012.
- M. Fidler. Survey of deterministic and stochastic service curve models in the network calculus. *IEEE Communications Surveys & Tutorials*, 2010.
- M. Fidler and A. Rizk. A guide to the stochastic network calculus. *IEEE Communications Surveys & Tutorials*, 2015.
- J. Liebeherr, Draft chapters (upon request).

Referenced Papers

Statistical Envelopes, Sample Path Envelopes

BBLO00 R. Boorstyn, A. Burchard, J. Liebeherr, C. Oottamakorn. Statistical Service Assurances for Packet Scheduling Algorithms, IEEE JSAC, Dec. 2000.

G12 Y. Ghiassi-Farrokhfal, Network Performance Analysis of Packet Scheduling Algorithms, Ph.D. Thesis, University of Toronto, 2012.

Stochastic Performance Bounds

BLP06 A. Burchard, J. Liebeherr, S. D. Patek, A Min-Plus Calculus for End-to-end Statistical Service Guarantees, IEEE Trans. on Inf. Theory, Sep. 2006.

CBL06 F. Ciucu, A. Burchard, J. Liebeherr, Scaling Properties of Statistical End-to-end Bounds in the Network Calculus," IEEE Transaction on Inf. Theory, June 2006.

LBL07 C. Li, A. Burchard, J. Liebeherr, A Network Calculus with Effective Bandwidth, ACM/IEEE Trans. on Networking, Dec. 2007.

BLC11 A. Burchard, J. Liebeherr, F. Ciucu, On Superlinear Scaling of Network Delays, ACM/IEEE Trans. Netw., August 2011.

LGB11 J. Liebeherr, Y. Ghiassi-Farrokhfal, A. Burchard, On the Impact of Link Scheduling on End-to-End Delays in Large Networks, IEEE JSAC, May 2011.

LBC12 J. Liebeherr, A. Burchard, F. Ciucu, Delay Bounds in Communication Networks with Heavy-Tailed and Self-Similar Traffic , IEEE Trans. on Inf. Theory, Feb. 2012.

Referenced Papers

MGF Network Calculus

F06 M. Fidler. An end-to-end probabilistic network calculus with moment generating functions. In *IEEE 14th International Workshop on Quality of Service (IWQoS)*, pages 261–270, June 2006.

Network Calculus on Martingales

C07 F. Ciucu, Scaling Properties in the Stochastic Network Calculus, Ph.D. Thesis, University of Virginia, 2007.

PC15 F. Poloczek and F. Ciucu,Service Martingales: Theory and Applications to the Delay Analysis of Random Access Protocols, IEEE Infocom, April 2015.

Deterministic Schedulability

KWZL96 E. W. Knightly, D. E. Wrege, H. Zhang, and J. Liebeherr, Deterministic Delay Bounds for VBR Video in Packet-Switching Networks: Fundamental Limits and Practical Tradeoffs, IEEE/ACM Trans. on Networking, June 1996.

Bandwidth estimation

LFV10 J. Liebeherr, M. Fidler, and S. Valaee, A System-Theoretic Approach to Bandwidth Estimation, ACM/IEEE Trans. on Networking, August 2010.

Application Areas

WCLL12 K. Wang, F. Ciucu, C. Lin, and S. H. Low, A stochastic power network calculus for integrating renewable energy sources into the power grid, IEEE JSAC, July 2012.

ZLB14 H. Al-Zubaidy, J. Liebeherr, and A. Burchard, Network-Layer Performance Analysis of Multi-Hop Fading Channels, IEEE/ACM Transactions on Networking, to appear (DOI: 10.1109/TNET.2014.2360675).

ZL14 H. Al-Zubaidy and J. Liebeherr, Service Characterizations for Multi-Hop Multiaccess Wireless Networks, Sixth IEEE International Workshop on Network Science for Communication Networks, May 2014.