Getting a Grip on Delays in Packet Networks

Jorg Liebeherr

Department of Electrical and Computer Engineering
University of Toronto

Collaborators

- Almut Burchard
- Robert Boorstyn
- Chaiwat Oottamakorn
- Stephen Patek
- Chengzhi Li
- Florin Ciucu
- Yashar Ghiassi-Farrokhfal

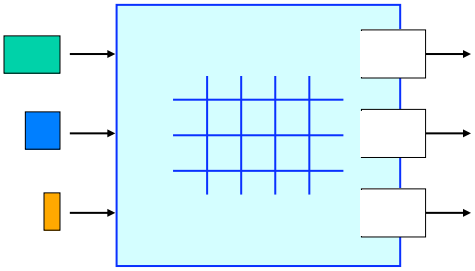
Disclaimer

- This talk makes a few simplifications
- Please see papers for complete results

Papers

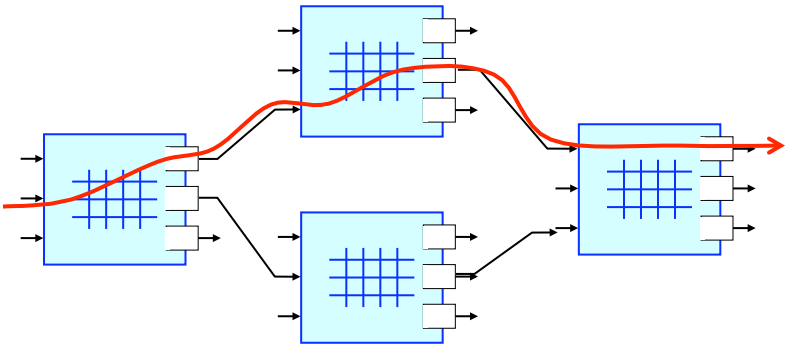
1. *Jörg Liebeherr, Dallas E. Wrege, Domenico Ferrari, "Exact admission control for networks with a bounded delay service," ACM/IEEE Trans. Netw. 4(6), 1996.*
2. *E. W. Knightly, D. E. Wrege, H. Zhang, J. Liebeherr, "Fundamental Limits and Tradeoffs of Providing Deterministic Guarantees to VBR Video Traffic," ACM Sigmetrics, 1995.*
3. R. Boorstyn, A. Burchard, J. Liebeherr, C. Oottamakorn. "Statistical Service Assurances for Packet Scheduling Algorithms", IEEE JSAC, December 2000.
4. C. Li, A. Burchard, J. Liebeherr, "A Network Calculus with Effective Bandwidth," ACM/IEEE Trans. on Networking, 15(6), 2007.
5. A. Burchard, J. Liebeherr, S. D. Patek, "A Min-Plus Calculus for End-to-end Statistical Service Guarantees," IEEE Trans. on Information Theory, 52(9), Sep. 2006.
6. F. Ciucu, A. Burchard, J. Liebeherr, "A Network Service Curve Approach for the Stochastic Analysis of Networks", ACM Sigmetrics 2005.
7. Almut Burchard, Jörg Liebeherr, Florin Ciucu, "On O(H log H) Scaling of Network Delays, INFOCOM 2007.
8. Jörg Liebeherr, Almut Burchard, Florin Ciucu, "Non-asymptotic Delay Bounds for Networks with Heavy-Tailed Traffic," INFOCOM 2010.
9. Jörg Liebeherr, Yashar Ghiassi-Farrokhfal, Almut Burchard, "Does Link Scheduling Matter on Long Paths?," ICDCS 2010.

Packet Switch

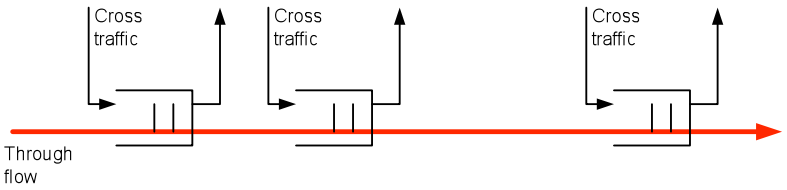


- Fixed-capacity links
- Variable delay due to waiting time in buffers
- Delay depends on
  1. Traffic
  2. Scheduling

Network

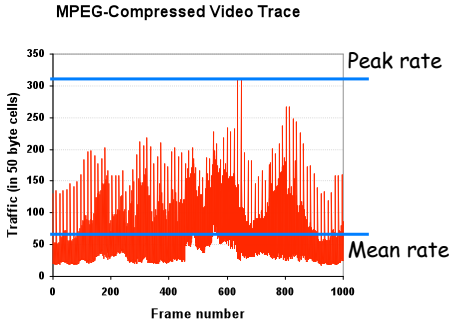


Simplified Network



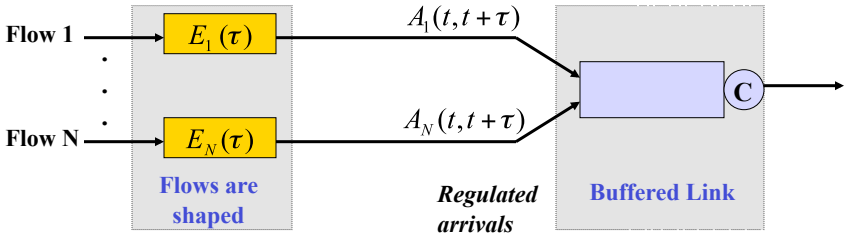
- Sequence of buffered links with fixed capacity

Traffic Arrivals



- Traffic arrivals in time interval (s,t] is $A(s,t)$

Regulated Arrivals

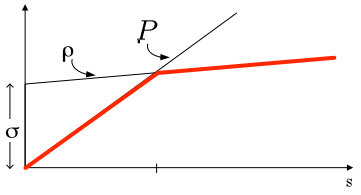


Traffic A_j is constrained by an envelope E_j such that:

$$E(t-s) \geq \sup_{s \leq t} \{A(s,t)\}$$

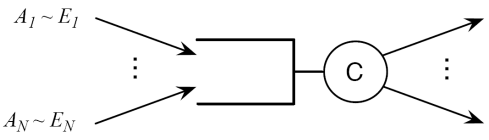
Popular envelope: "leaky bucket"

$$E(s) = \min(P\ s,\ \sigma + \rho\ s)$$

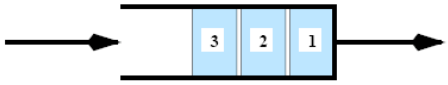


What is the maximum number of regulated flows with delay requirements that can be put on a single buffered link?

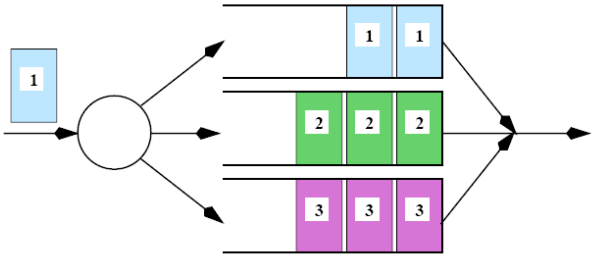
- Link capacity C
- Each flows j has
  - arrival function A_j
  - envelope E_j
  - delay requirement d_j



First-In-First-Out

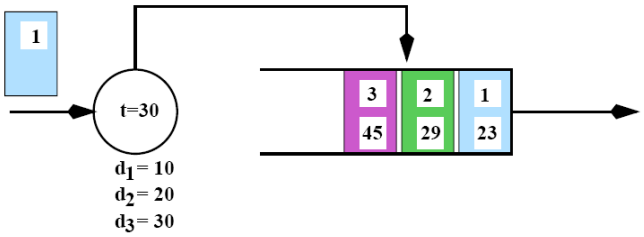


Static Priority (SP)



- **Blind Multiplexing (BMux):**
  All "other traffic" has higher priority

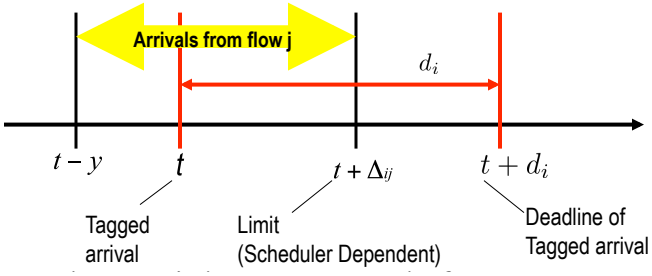
Earliest Deadline First (EDF)



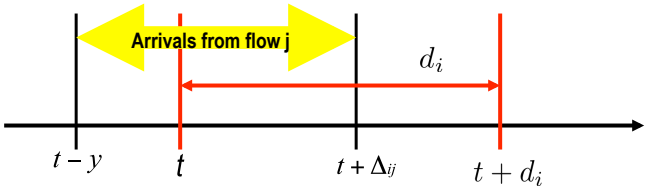
Optimal scheduler with respect to posed question.

Scheduling Algorithms

- Consider a work-conserving scheduler with rate C
- Consider arrival from flow i at t with $t+d_i$:



Scheduling Algorithms



$$\sup_y \left\{ \sum_j A_j(t-y, t+\Delta_{ij}) - C(y+d_i) \right\} \le 0$$

with

FIFO: $\Delta_{ij} = 0.$

Static Priority: $\Delta_{ij} = -\infty$ (higher) , 0 (higher) , d_q (higher).

EDF: $\Delta_{ij} = d_i - d_j$

Condition for meeting a delay bound

We have: $A_j(t, t+\tau) \le E_j(\tau) \qquad \forall t, \forall \tau.$

Therefore:

An arrival from class i **never** has a delay bound violation if

$$\sup_y \left\{ \sum_j E_j\ (y+\Delta_{ij}) - Cy \right\} \le Cd_i$$

Condition is tight, when E_j is concave

Numerical Result

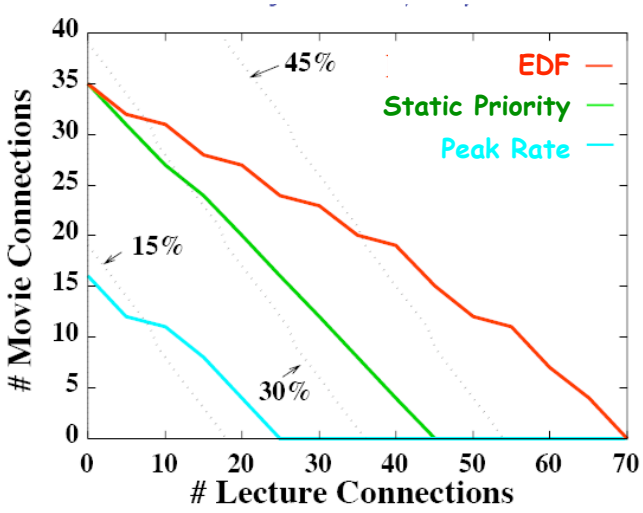
(Sigmetrics 1995)

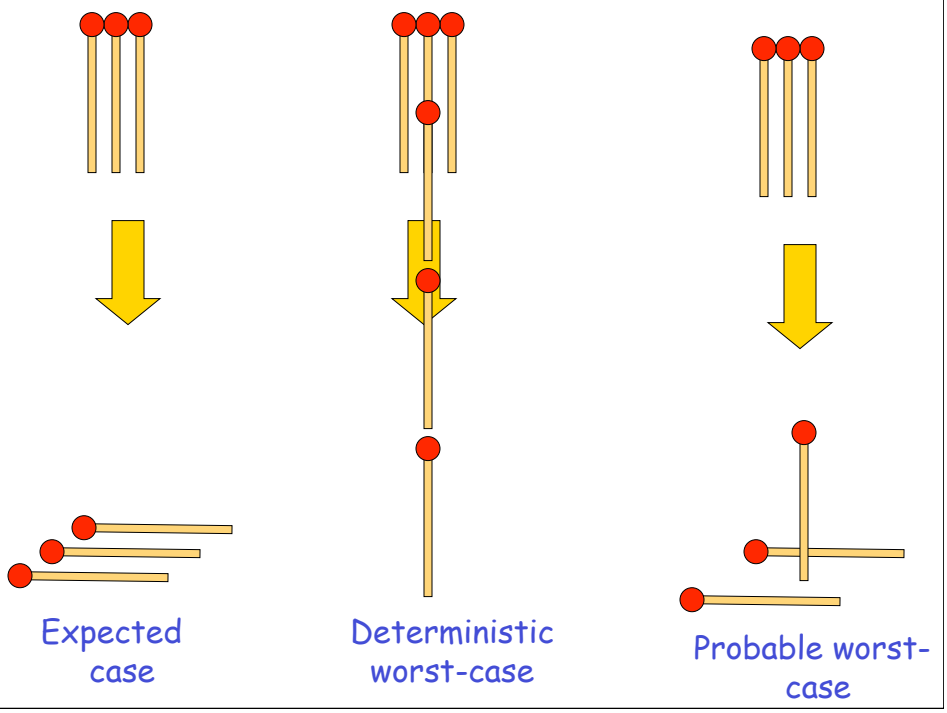
C = 45 Mbps

MPEG 1 traces:

Lecture:
d = 30 msec

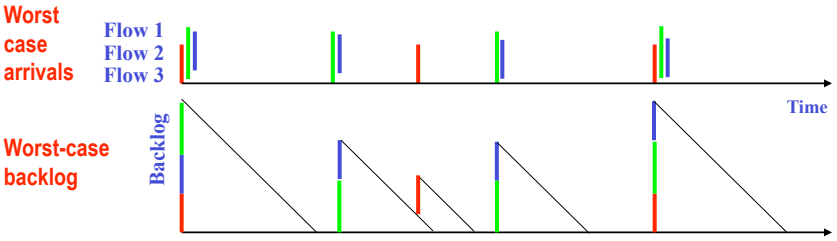
Movie
(Jurassic Park):
d = 50 msec





Statistical Multiplexing Gain

Without statistical multiplexing



What is the maximum number of flows with delay requirements that can be put on a buffered link **and considering statistical multiplexing**?

Arrivals $A_j(t, t+\tau)$ from a flow j are a random process

- Stationarity: The A_j are stationary random processes
- Independence: The A_i and $A_j (i \neq j)$ are stochastically independent

Envelopes for random arrivals

Statistical envelope bounds arrival from flow j with high certainty

- **Statistical envelope $\mathcal{G}$:**

$$Pr\{A(s,t) > \mathcal{G}(t-s) + \sigma\} < \varepsilon(\sigma) \quad \forall s,t$$

- **Statistical sample path envelope $\mathcal{H}$:**

$$Pr\{\sup_{s \le t} \{A(s,t) - \mathcal{H}(t-s)\} > \sigma\} < \varepsilon(\sigma)$$

Statistical envelopes are non-random functions

Statistical Multiplexing Gain

(Resources needed to support guarantees for N flows) << N · (Resources needed to support guarantees for 1 flow)

Multiplexing gain is the raison d'être for packet networks.

Arrivals from group of flows: $A_{\mathcal{C}} = \sum_j A_j$

Deterministic envelope: $E_{\mathcal{C}} = \sum_j E_j$

Statistical envelope: $\mathcal{G}_{\mathcal{C}} \ll \sum_j \mathcal{G}_j \lll E_{\mathcal{C}}$

Statistical envelope for group of indepenent (shaped) flows

- Exploit independence and extract statistical multiplexing gain when calculating $\mathcal{G}_{\mathcal{C}}$
- For example, using the Chernoff Bound, we can obtain

$$\mathcal{G}_{\mathcal{C}}(t) = \inf_{s>0} \frac{1}{s}\left(\sum_{j\in\mathcal{C}} \log \overline{M}_j(s,t) - \log \varepsilon\right)$$

$$\overline{M}_j(s,t) = 1 + \frac{\rho_j\, t}{E_j(t)}\left(e^{sE_j(t)} - 1\right)$$

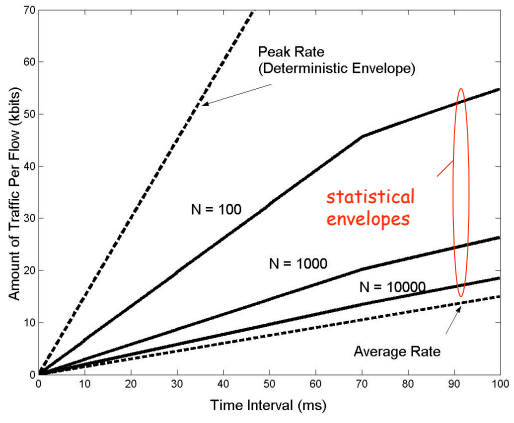
$$\rho_j = \lim_{\tau\to\infty} E_j(\tau)/\tau$$

Statistical Envelope vs. Deterministic Envelopes *(JSAC 2000)*

E(t) =min (Pt, σ+ρt)

Type 1 flows:
P =1.5 Mbps
ρ = .15 Mbps
σ =95400 bits

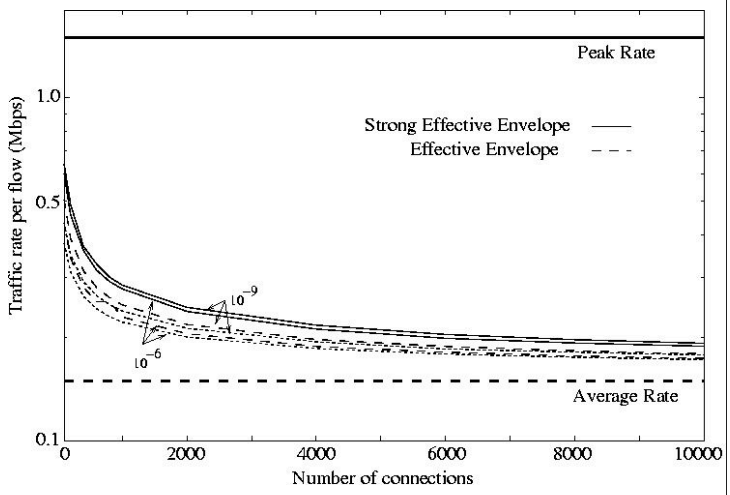
Type 2 flows:
P = 6 Mbps
ρ = .15 Mbps
σ = 10345 bits



Type 1 flows

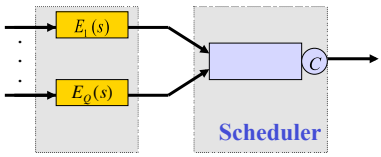
Statistical vs. Deterministic Envelope Envelopes *(JSAC 2000)*

Traffic rate at t = 50 ms
Type 1 flows



Scheduling Algorithms

- Work-conserving scheduler with unit rate that serves Q classes
- Class-q traffic has delay bound d_q
- Scheduling algorithm



Deterministic Service
Never a delay bound violation if:

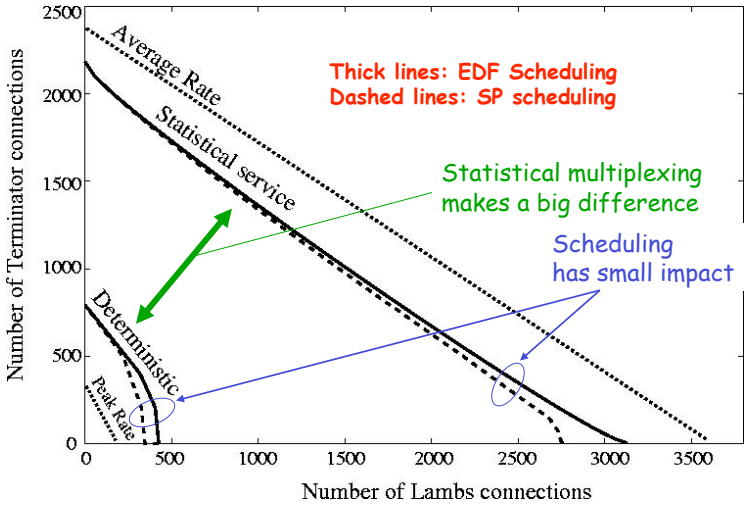
$$\sup_s \left\{ \sum_p E_{\mathcal{C}_p}(\Delta_{qp} + Cs) - s \right\} \leq Cd_q$$

Statistical Service
Delay bound violation with ε if:

$$\sup_s \left\{ \sum_p \mathcal{H}_{\mathcal{C}_p}(\Delta_{qp} + Cs) - s \right\} \leq Cd_q$$

Statistical Multiplexing vs. Scheduling *(JSAC 2000)*

Example: MPEG videos with delay constraints at C= 622 Mbps
Deterministic service vs. statistical service ($\varepsilon = 10^{-6}$)



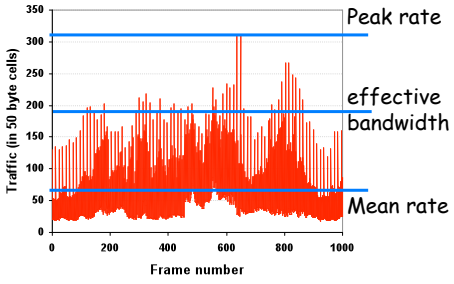
$d_{terminator}$=100 ms
d_{lamb}=10 ms

More interesting traffic types

- **So far:** Traffic of each flow was regulated
- **Next:**
  - On-Off traffic
  - Fraction Brownian Motion (FBM) traffic

Approach:

- Exploit literature on Effective Bandwidth
- Derived for many traffic types



Statistical Envelopes and Effective Bandwidth

Effective Bandwidth (Kelly 1996)

$$\alpha(s,\tau) = \sup_{t\geq 0}\left\{\frac{1}{s\tau}\log E[e^{s(A(t+\tau)-A(t))}]\right\}$$

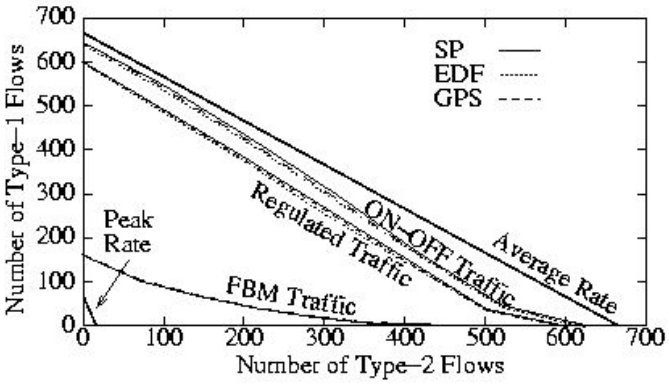
$$s,\tau \in (0,\infty)$$

Given $\alpha(s,\tau)$, an effective envelope is given by

$$\mathcal{G}^{\varepsilon}(\tau) = \inf_{s>0}\{\tau\alpha(s,\tau) - \frac{\log\varepsilon}{s}\}$$

Statistical Envelopes and Effective Bandwidth *(ToN 2007)*

Comparisons of statistical service guarantees for different schedulers and traffic types



Schedulers:

SP- Static Priority
EDF – Earliest Deadline First
GPS – Generalized Processor Sharing

Traffic:

Regulated – leaky bucket
On-Off – On-off source
FBM – Fractional Brownian Motion

C= 100 Mbps, ε = 10^{-6}

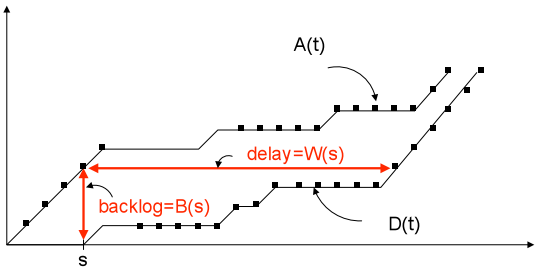
Delays on a long path with multiple nodes:

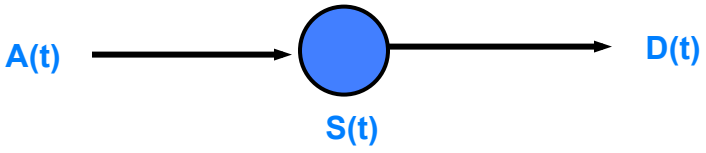
- Role of Scheduling
- Impact of Statistical Multiplexing

- How do delays scale?
- Does scheduling still matter in a large network?

Deterministic Network Calculus (1/3)

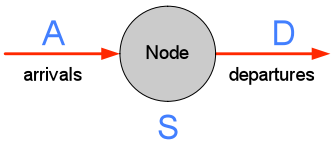
- Cruz, Chang, LeBoudec (90's)
- Worst case delay and backlog bounds for fluid flow traffic





Deterministic Network Calculus (2/3)

- Worst-case view of
  - arrivals: $A(s,t) \leq \mathcal{G}(t-s)$
  - service : $D(t) \geq A * S(t)$

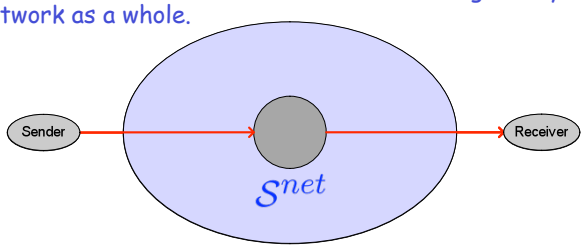


- Implies worst-case bounds
  - backlog: $B(t) \leq \mathcal{G} \oslash \mathcal{S}(0)$
  - delay : $W(t) \leq \inf\{d \mid \mathcal{G}(s) \leq \mathcal{S}(s+d)\ \forall s \geq 0\}$

- (min,+) formulation with
  - convolution operator:
    $$f * g(t) = \inf_{0 \leq s \leq t} (f(s) + g(t-s))$$
  - deconvolution operator:
    $$f \oslash g(t) = \sup_{s \geq 0} (f(t+s) - g(s))$$

Deterministic Network Calculus (3/3)

- Main result:
  - If S^1, S^2 and S^3 describes the service at each node, then $S^{net} = S^1 * S^2 * S^3$ describes the service given by the network as a whole.



Stochastic Network Calculus

- Probabilistic view on arrivals and service
  - Statistical Sample Path Envelope

$$Pr(\forall s \leq t : A(s,t) > \mathcal{H}(t-s)) \leq \varepsilon$$

  - Statistical Service Curve

$$Pr(D(t) < A * \mathcal{S}(t)) \leq \varepsilon$$

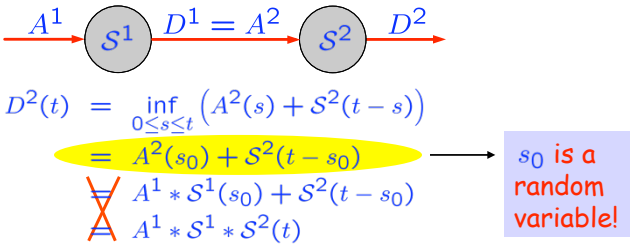
- Results on performance bounds carry over, e.g.:
  - Backlog Bound

$$Pr(B(t) > \mathcal{H} \oslash \mathcal{S}(0)) \leq \varepsilon$$

Stochastic Network Calculus

- Hard problem: Find $\mathcal{S}^{net}$ so that $\mathcal{S}^{net} = \mathcal{S}^1 * \mathcal{S}^2 * \ldots * \mathcal{S}^H$

- Technical difficulty:



Statistical Network Service Curve (Sigmetrics 2005)

- Notation: $\mathcal{S}_{-\delta}(t) = \mathcal{S}(t) - \delta t$

- **Theorem**: If $\mathcal{S}^1, \mathcal{S}^2, \ldots, \mathcal{S}^H$ are statistical service curves, then for any $\delta > 0$:

$$\mathcal{S}^{net} = \mathcal{S}^1 * \mathcal{S}^2_{-\delta} * \cdots * \mathcal{S}^H_{-(H-1)\delta}$$

is a statistical network service curve with some finite violation probability.

EBB model

- Traffic with **Exponentially Bounded Burstiness (EBB)**

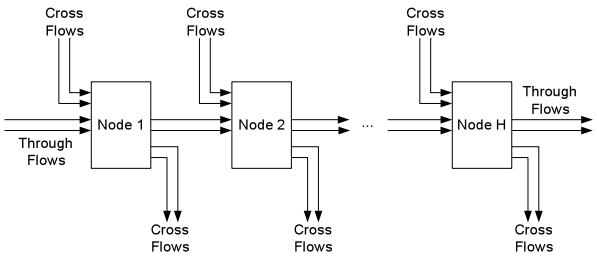
$$P\big(A(s,t) - \rho(t-s) > \sigma\big) \leq Me^{-\alpha\sigma}$$

for constants M, α

- Envelope: $\mathcal{G}(t;\sigma) = \rho t\ , \quad \varepsilon(\sigma) = Me^{-\alpha\sigma}$

- EBB model gives closed bounds for E2E delays.

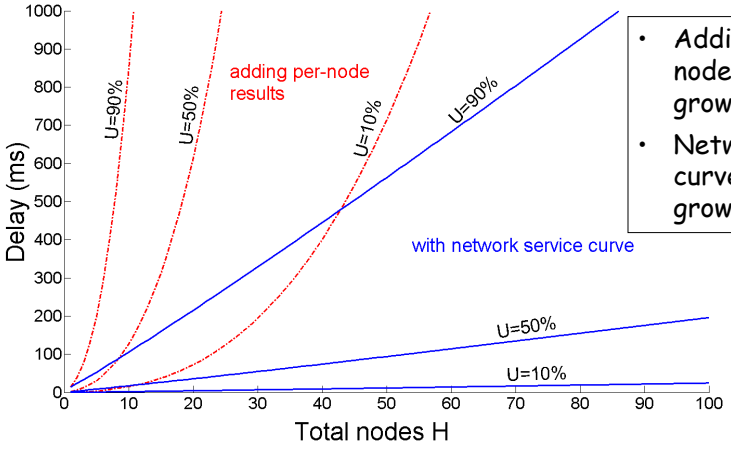
Example: Scaling of Delay Bounds



- Traffic is Markov Modulated On-Off Traffic (EBB model)
- Fixed capacity link
- Through flow has lower priority
- Compare delay with network service curve to a summation of per-node bounds

Example: Scaling of Delay Bounds *(Sigmetrics 2005)*

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



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

Result: Lower Bound on E2E Delay *(Infocom 2007)*

- Tandem network of M/M/1 queues with identical service times
- W_H is the E2E delay for H queues

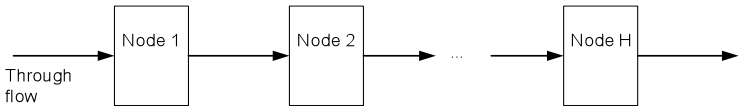
Theorem: E2E delay W_H satisfies for all $0 < z < 1$

$$Pr\left(W_H \leq \gamma_1 H \log(\gamma_2 H)\right) \leq z$$

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

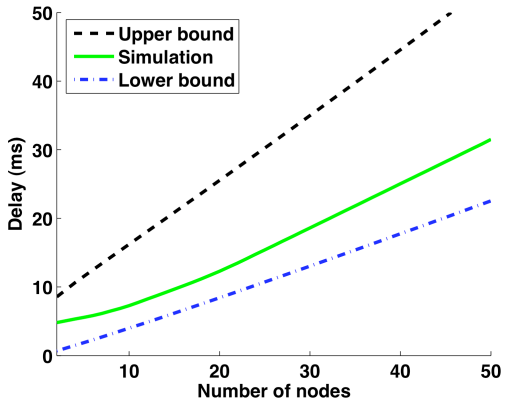
$$w_H(z) = \Omega(H \log H)$$

Numerical examples



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

Upper and Lower Bounds on E2E Delays *(Infocom 2007)*



Capacity

$C = 100\ Mbps$

Mean packet size

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

Load factor

$\rho = 90\%$

Violation probability

$\varepsilon = 10^{-6}$

→ Delays in Networks scale as $\Omega(H \log H)$

Superlinear Scaling of Network Delays

- For traffic satisfying "Exponential Bounded Burstiness", E2E delays follow a scaling law of $\Theta(H \log H)$

- E2E delays indeed scale differently
  - … than worst-case delays
  - … than delays with independent cross traffic and service times

Can we compute scaling of delays for very difficult traffic ?

Heavy-Tailed Self-Similar Traffic

- A **heavy-tailed** process X satisfies

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

with $1 < \alpha < 2$

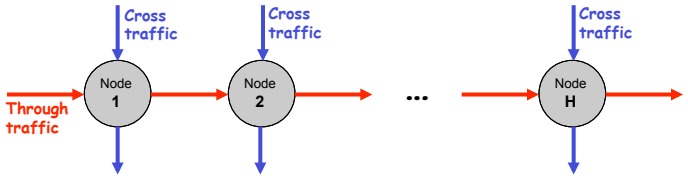
- A **self-similar** process satisfies

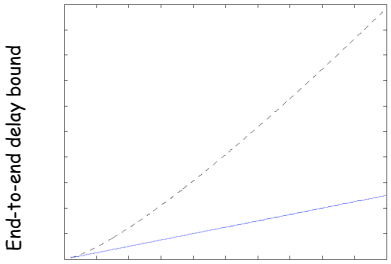
$$X(t) \sim_{dist} a^{-H} X(at)$$

$a > 0$

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

End-to-End Delays





Exponentially bounded traffic
Θ (N log N)
(Sigmetrics 2005, Infocom 2007)

Worst-case delays
Θ (N)
(e.g., LeBoudec and Thiran 2000)

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: Pareto Traffic

- Size of i-th arrival: $Pr(X_i > x) = \left(\frac{x}{b}\right)^{-\alpha}$ $x \geq b$, $1 < \alpha < 2$
- Arrivals are evenly spaced with gap λ: $A(t) = \sum_{i=1}^{N(t)} X_i$
- With Generalized Central Limit Theorem … $A(t) \approx \lambda t E[X] + c_\alpha (\lambda t)^{1/\alpha} S_\alpha$
  … and tail bound $Pr(S_\alpha > \sigma) \sim (c_\alpha \sigma)^{-\alpha}$

  α-stable distribution
- … we get htss envelope

$$\mathcal{G}(t;\sigma) = \lambda E[X] t + \sigma t^{1/\alpha}$$
$$\varepsilon(\sigma) = \lambda \sigma^{-\alpha}$$

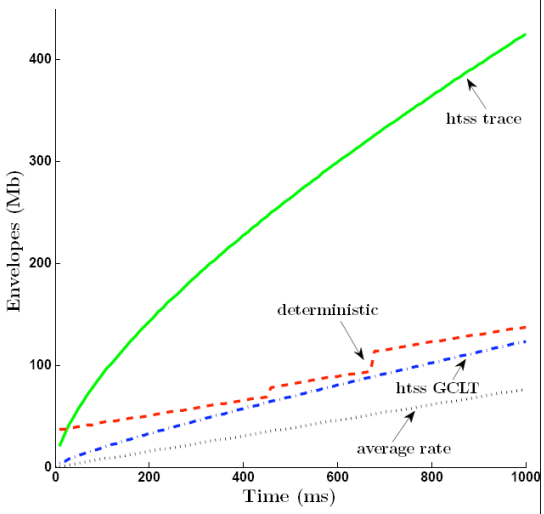
Example: Envelopes for Pareto Traffic (Infocom 2010)

Parameters:

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

Comparison of envelopes:

- htss GCLT envelope
- Average rate
- Trace-based
  - deterministic envelope
  - htts trace envelope



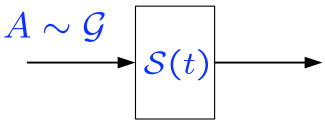
Single Node Delay Bound

- htss envelope:

$$\mathcal{G}(t;\sigma) = rt + \sigma t^H$$
$$\varepsilon(\sigma) = K\sigma^{-\alpha}$$

- ht service curve:

$$\mathcal{S}(t;\sigma) = [Rt - \sigma]_+$$
$$\varepsilon(\sigma) = L\sigma^{-\beta}$$



- Delay bound:

$$Pr(W(t) > w) \leq M(Rw)^{-\min\{\alpha(1-H),\beta\}}$$

Example: Node with Pareto Traffic

(Infocom 2010)

Traffic parameters:

$\alpha = 1.6$

$b = 150\ Byte$

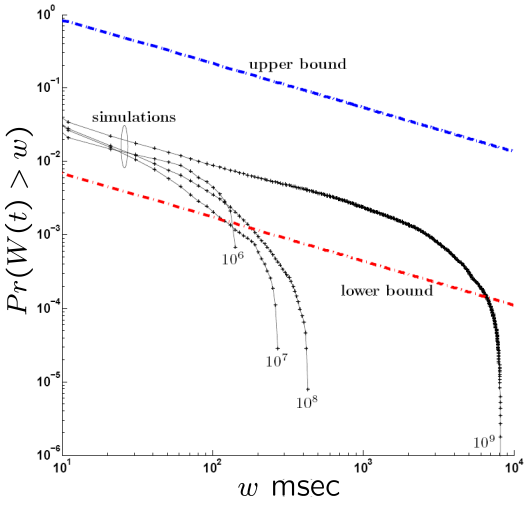
$\lambda = 75\ Mbps$

Node:

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

Compared with:

- Lower bound from Infocom 2007
- Simulations



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

Parameters:

$N = 1, 2, 4, 8$

Compared with:

- Lower bound from Infocom 2007
- Simulation traces of 10^8 packets

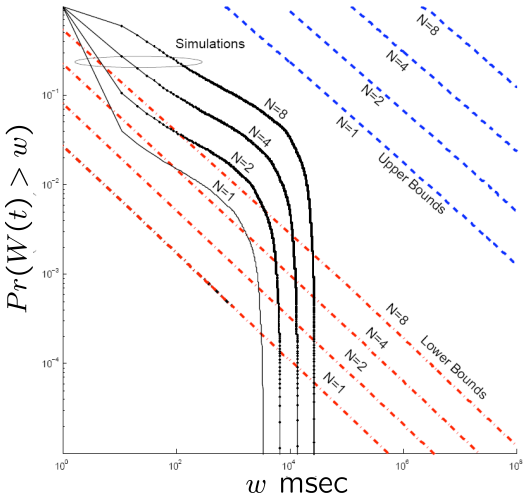
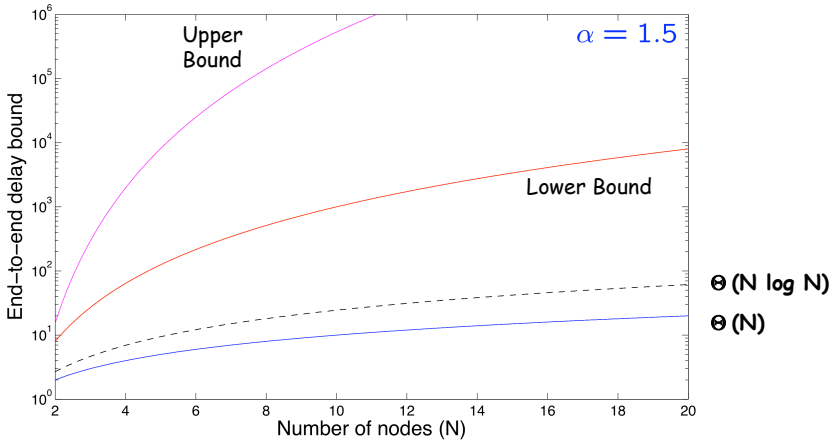


Illustration of scaling bounds *(Infocom 2010)*

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

Lower Bound: $\Theta(N^{\frac{\alpha}{\alpha-1}})$

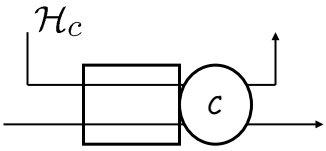


Bring back scheduling

So far:
Through traffic has lowest priority and gets leftover capacity

→ **Leftover Service**

or **Blind Multiplexing**



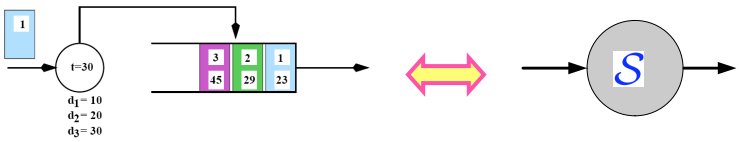
$$\mathcal{S}_j = [Ct - \mathcal{H}_c(t)]_+$$

How do end-to-end delay bounds look like for different schedulers?

Does link scheduling matter on long paths?

Service curves vs. schedulers

- How well can a service curve describe a scheduler?

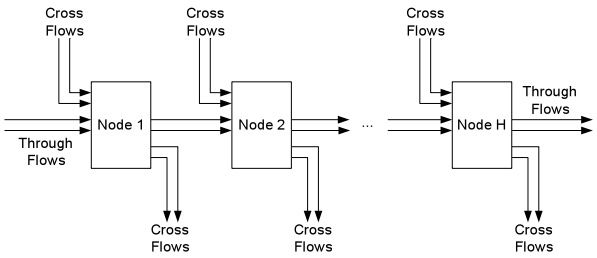


- For schedulers considered earlier, the following is ideal:

$$\mathcal{S}_j(t;\theta) = [Ct - \mathcal{H}_c(t - \theta + \Delta_{j,k}(\theta))]_+ I(t > \theta)$$

with indicator function $I(\text{expr})$ and parameter $\theta \geq 0$

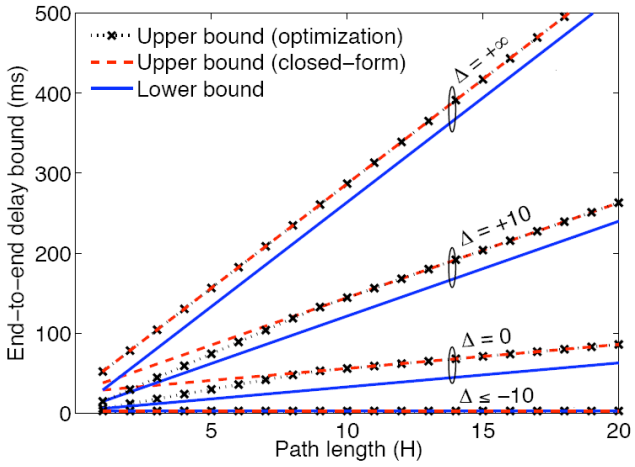
Example: End-to-End Bounds



- Traffic is Markov Modulated On-Off Traffic (EBB model)
- Fixed capacity link

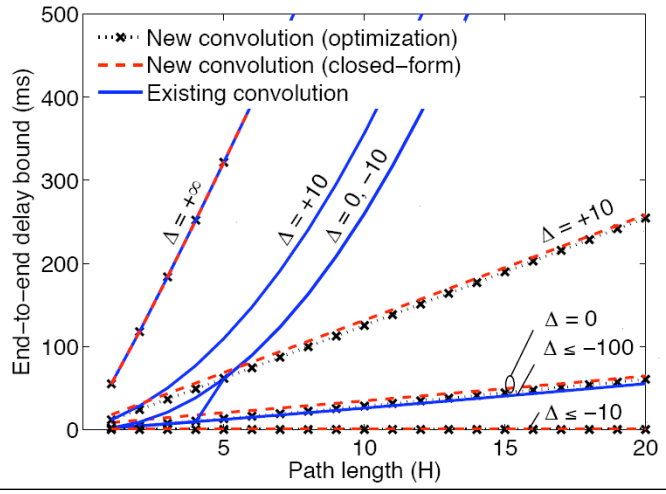
Example: Deterministic E2E Delays *(ICDCS 2010)*

- Peak rate: P = 1.5 Mbps
  Average rate: ρ = 0.15 Mbps
- C = 100 Mbps



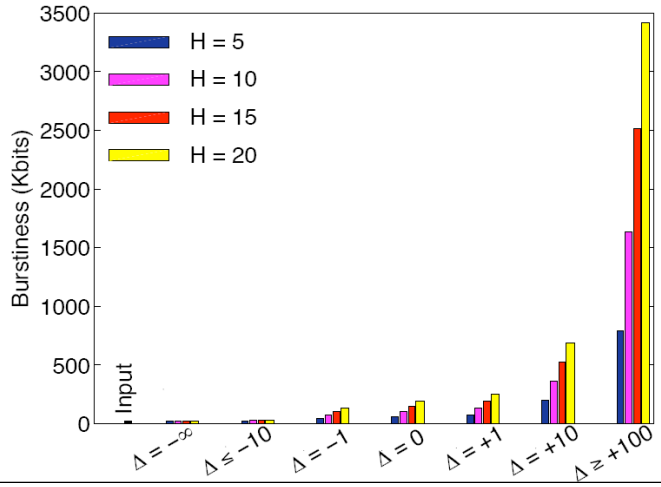
Example: Statistical E2E Delays *(ICDCS 2010)*

- Peak rate: P = 1.5 Mbps
  Average rate: ρ = 0.15 Mbps
- C = 100 Mbps
- $\varepsilon = 10^{-9}$



Example: Statistical Output Burstiness *(ICDCS 2010)*

- Peak rate: P = 1.5 Mbps
  Average rate: ρ = 0.15 Mbps
- C = 100 Mbps
- $\varepsilon = 10^{-9}$



Conclusions

Requirements	Queueing networks	Effective bandwidth	Network calculus	Stochastic network calculus
Traffic classes (incl. self-similar, heavy-tailed)	Limited	Broad	Broad (but loose)	Broad
Scheduling	Limited	No	Yes	Yes
QoS (bounds on loss, throughput delay)	Very limited	Loss, throughput	Deterministic	Yes
Statistical Multiplexing	Some	Yes	No	Yes