Distributed Computer Systems Lab
http://disco.informatik.uni-kl.de

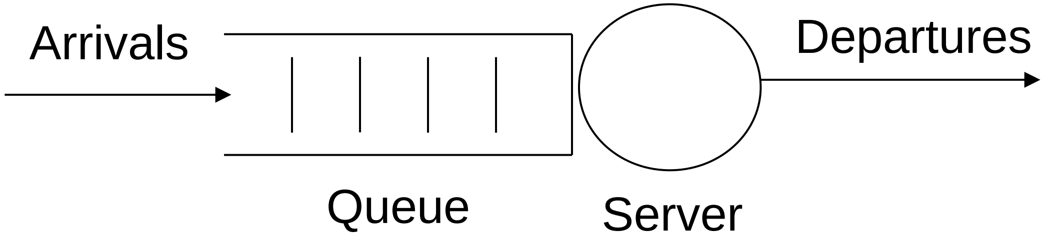
Performance Modelling of Distributed Systems

4. Modelling of the Service Processes

Prof. Dr.-Ing. Jens B. Schmitt
(jschmitt@cs.uni-kl.de)

The „Canonical“ Problem – Still ...

Goal: Compute delay for a single flow at a single server



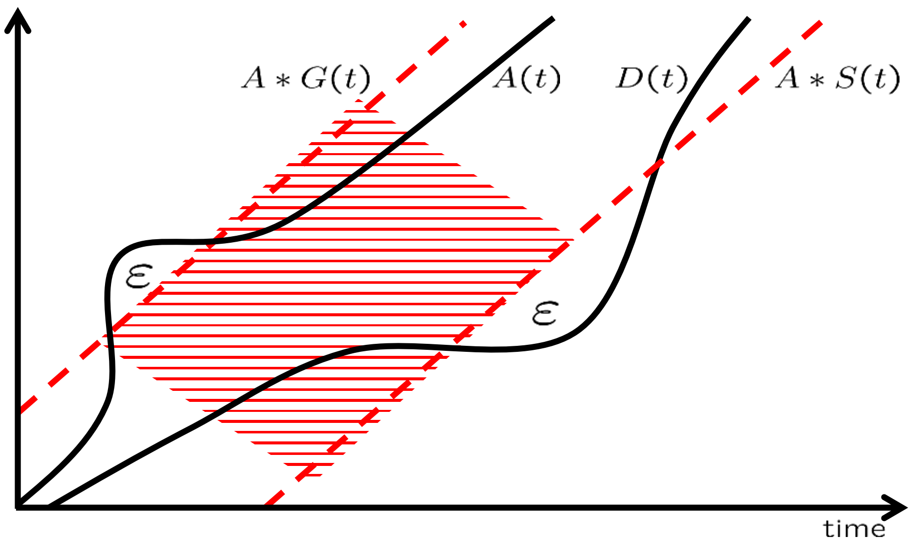
- QT approach: knowledge about arrival and service distributions

- DNC approach: deterministic bounds on arrivals and service

- SNC approach: **probabilistic** bounds on arrivals and **service**

Deterministic Bounds on Service (1)

- Lower bound on service necessary to control the system behaviour



- Definition: Strict Service Curve

A service element offers a strict service curve $\beta : \mathbb{N}_0 \to \mathbb{R}_0^+$, if for all backlogged periods $[m, n]$ and all input flows A it holds that:

$$D(n) - D(m) \geq \beta(n - m).$$

Take Step Back: The „Natural“ Laws of Queueing

- Lindley‘s equation

A service element fulfills Lindley's equation if for all $q(n) := A(n) - D(n)$ and all input flows A it holds that:

$$q(n+1) = [q(n) + a(n+1) - s(n+1)]^+,$$

where s(n) is the amount of data the service element can process at time n.

- Reich‘s equation

$$D(n) = \min_{0 \leq k \leq n} \{A(k) + S(k,n)\} =: A \otimes S(n)$$

$$\text{with } S(k,n) := \sum_{i=k+1}^{n} s(i)$$

Deterministic Bounds on Service (2)

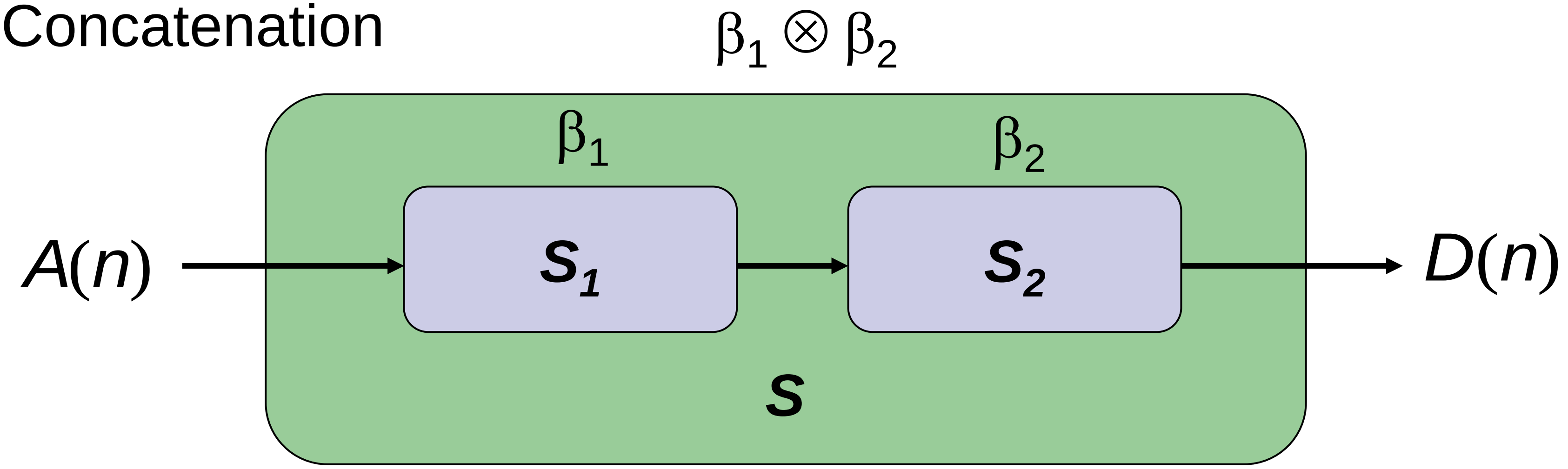
- Definition: Service Curve

A service element offers a service curve $\beta : \mathbb{N}_0 \rightarrow \mathbb{R}_0^+$

if for all $n \geq m \in \mathbb{N}_0$ and all input flows A it holds that:

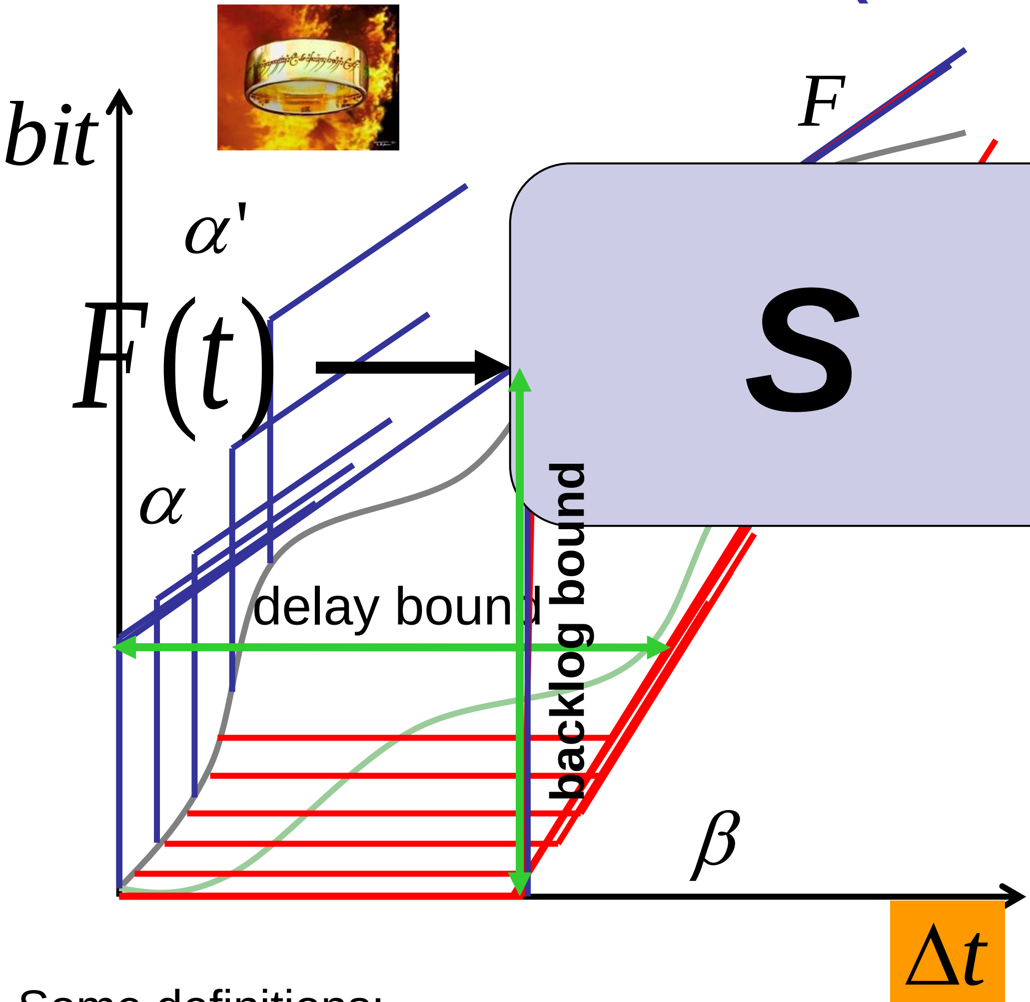
$$D(n) \geq \min_{0 \leq k \leq n} \{A(k) + \beta(n-k)\} = A \otimes \beta(n).$$

- Concatenation



- A strict service curve is a service curve but not vice versa
  - If the server is work-conserving than they are equivalent

Network Calculus – One (Semi-)Ring to Rule them All



arrival curve α

$\Leftrightarrow F \leq F \otimes \alpha$

service curve β

$\Leftrightarrow F' \geq F \otimes \beta$

$backlog \leq v(\alpha, \beta)$

$delay \leq h(\alpha, \beta)$

output bound

$\alpha' = \alpha \oslash \beta$

Some definitions:

$$(f \otimes g)(t) := \inf_{0 \leq s \leq t} \{ f(t-s) + g(s) \}$$

$$(f \oslash g)(t) := \sup_{u \geq 0} \{ f(t+u) - g(u) \}$$

Stochastic Bounds on Service

- The (recurring) difficulties of deterministic bounds on service
  - Assume two flows sharing a server, one of them with exponentially i.i.d. increments, (no queueing yet)

- Notation
  - triangular array $\Lambda(\mathbb{N}_0)$
  - doubly indexed stochastic process $\mathcal{S}_2$
  - bivariate flow $\mathcal{F}_2$

- System-theoretic operators
  - Remember univariate min-plus convolution

$$A \otimes B(n) = \min_{0 \le k \le n} \{A(k) + B(n-k)\}$$

  - Bivariate convolution

$$A \otimes B(m,n) := \min_{m \le k \le n} \{A(m,k) + B(k,n)\}$$

Dynamic S-Server

- Definition

  Assume a service element has a flow A as input and the output is denoted by D. Let $S \in \mathcal{S}_2$ be a doubly indexed stochastic process with:

  $$S(m,n) \le S(m,n') \qquad \forall\, n \le n' \in \mathbb{N}_0 \text{ a.s.}$$

  The service element is a dynamic S-server if for all $n \ge 0$ it holds that: $D(0,n) \ge A \otimes S(0,n) \quad$ a.s.

- Examples
  - Deterministic rate-latency service curve
  - Left-over service curve

Stochastic Bounds on Service: MGF and Tail

- Definition (MGF Bound on Service)

  A dynamic S-server is $(\sigma(\theta), \rho(\theta))$-bounded for some $\theta > 0$, if $\phi_{S(m,n)}(-\theta)$ exists and

  $$\phi_{S(m,n)}(-\theta) \leq e^{\theta\rho(\theta)(n-m)+\theta\sigma(\theta)} \qquad \forall (m,n) \in \Lambda(\mathbb{N}_0).$$

  A dynamic S -server is $f(\theta, n)$-bounded for some $\theta > 0$, if $\phi_{S(m,n)}(-\theta)$ exists and

  $$\phi_{S(m,n)}(-\theta) \leq f(\theta, n-m) \qquad \forall (m,n) \in \Lambda(\mathbb{N}_0).$$

- Definition (Tail Bound on Service)

  A dynamic S-server is tail-bounded by envelope α with error η if for all arrival flows A and all $n \in \mathbb{N}_0, \varepsilon \in \mathbb{R}^+$

  $$\mathbb{P}(D(0,n) < A \otimes (S - \alpha(\varepsilon))(0,n)) < \eta(n, \varepsilon).$$