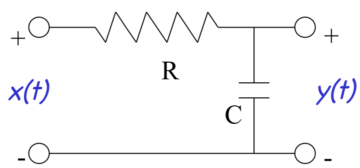
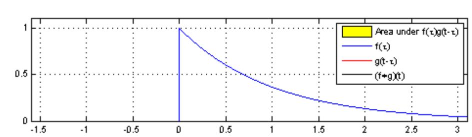
Network Calculus:

- Reference Material:
  J.-Y. LeBoudec and Patrick Thiran: "Network Calculus: A Theory of Deterministic Queuing Systems for the Internet", Springer Verlag Lecture Notes in Computer Science No. 2050.
- Network Calculus as system theory for computer networks.
- Some mathematical background
- Arrival Curves
- Service Curves
- Network Calculus Basics

Example Simple Electronic Circuit: RC Cell



- Output $y(t)$ of this circuit is **convolution** of input $x(t)$ and impulse response $h(t)$ of circuit.
- Impulse response: $h(t) = \frac{1}{RC} e^{(-t/RC)} \quad t \geq 0$
- Output: $y(t) = (h \otimes x)(t) = \int_0^t h(t-s)x(s)ds$

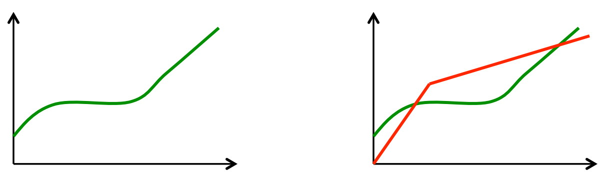


Goal: Apply System Theory to Networks
Example: Greedy Shaper

- A shaper forces an input traffic flow $x(I)$ to have an output $y(I)$ which adheres to an envelope σ.

- The output function $y(I)$ can be derived as follows:

$$y(I) = (\sigma \otimes x)(I) = \inf_{0 \le s \le t} \{\sigma(I - s) + x(s)\}$$



- Other analogies apply as well (commutativity and associativity), which allow to extend this analysis to large-scale systems.

- There are significant differences, though!

Min-Plus Calculus: Infimum vs. Minimum

- Let S be nonempty subset of R.

Definition [Infimum]

$\inf(S) = (M \text{ s.t. } s \ge M \ \forall \ s \in S)$

$\inf(\emptyset) = +\infty$

Definition [Minimum]

$\min(S) = (M \underline{\in S} \text{ s.t. } s \ge M \ \forall \ s \in S)$

- Notation: ^ denotes infimum (e.g. a ^ $b = \min\{a,b\}$)

The Dioid ($R \cup \{+\infty\}$, ^, +)

- Conventional ("plus-times") algebra operates on algebraic structure (R,+,*).
- Min-plus algebra replaces operations:
  - addition becomes computation of infimum
  - multiplication becomes addition
- Resulting algebraic structure becomes ($R \cup \{+\infty\}$, ^, +)
- Example:
  - Conventional algebra: (3+4) * 5 = (3*5) + (4*5) = 15 + 20
  - min-plus algebra: (3^4) + 5 = (3+5) ^ (4+5) = 8 ^ 9 = 8

Properties of ($R \cup \{+\infty\}$, ^, +)

- **(Closure of ^)** For all $a,b \in R \cup \{+\infty\}$, a ^ $b \in R \cup \{+\infty\}$
- **(Associativity of ^)** For all $a,b,c \in R \cup \{+\infty\}$, $(a$^$b)$^$c=a$^$(b$^$c)$
- **(Existence of a zero element of ^)** There is some $e \in R \cup \{+\infty\}$, such that for all $a \in R \cup \{+\infty\}$, a ^ $e = a$.
- **(Idempotency of ^)** For all $a \in R \cup \{+\infty\}$, a ^ $a = a$.
- **(Commutativity of ^)** For all $a,b \in R \cup \{+\infty\}$, a ^ $b = b$ ^ a.
- **(Closure of +)** For all $a,b \in R \cup \{+\infty\}$, $a + b \in R \cup \{+\infty\}$.
- **(Zero element of ^ is absorbing for +)** For all $a \in R \cup \{+\infty\}$, $a + e = e = e + a$.
- **(Existence of neutral element for +)** There is some $u \in R \cup \{+\infty\}$ such that for all $a \in R \cup \{+\infty\}$, $a + u = a = u + a$.
- **(Distributivity of + with respect to ^)** For all $a,b,c \in R \cup \{+\infty\}$, $(a$ ^ $b) + c = (a + c)$ ^ $(b + c) = c + (a$ ^ $b)$

Wide-Sense Increasing Functions

Definition [wide-sense increasing]

A function is wide-sense increasing iff $f(s) \leq f(t)$ for all $s \leq t$.

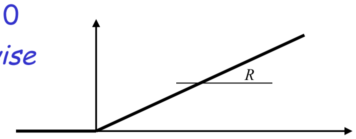
- Define G as the set of non-negative wide-sense increasing functions.
- Define F as the set of non-negative wide-sense increasing functions with $f(t) = 0$ for $t < 0$.

- Operations on functions:
  $(f + g)(t) = f(t) + g(t)$
  $(f \wedge g)(t) = f(t) \wedge g(t)$

Wide-Sense Increasing Functions

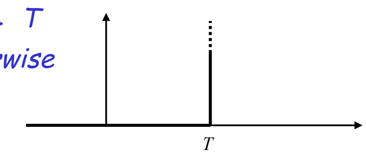
- Peak rate function λ_R:
  "Rate" R

$$\lambda_R(t) = \begin{cases} Rt & \text{if } t > 0 \\ 0 & \text{otherwise} \end{cases}$$



- Burst delay function δ_T:
  "Delay" T

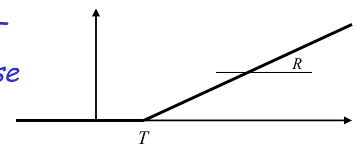
$$\delta_T(t) = \begin{cases} +\infty & \text{if } t > T \\ 0 & \text{otherwise} \end{cases}$$



Wide-Sense Increasing Functions (2)

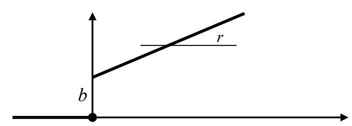
- Rate latency function $\beta_{R,T}$:
  "Rate" R , "Delay" T

$$\beta_{R,T}(t) = \begin{cases} R(t-T) & \text{if } t > T \\ 0 & \text{otherwise} \end{cases}$$



- Affine functions $\gamma_{r,b}$:
  "Rate" r, "Burst" b

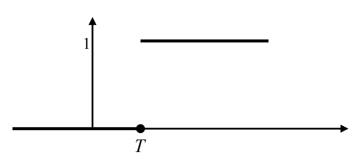
$$\gamma_{r,b}(t) = \begin{cases} rt+b & \text{if } t > 0 \\ 0 & \text{otherwise} \end{cases}$$



Wide-Sense Increasing Functions (3)

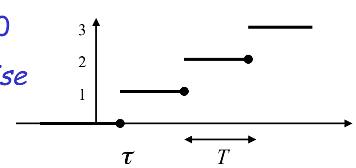
- Step function υ_T:

$$\upsilon_T(t) = \begin{cases} 1 & \text{if } t > T \\ 0 & \text{otherwise} \end{cases}$$



- Staircase function $u_{T,\tau}$:
  "Interval" T,
  "Tolerance" τ

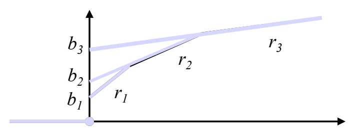
$$u_{T,\tau}(t) = \begin{cases} \left\lceil \frac{t+\tau}{T} \right\rceil & \text{if } t > 0 \\ 0 & \text{otherwise} \end{cases}$$



Wide-Sense Increasing Functions (4)

- More general functions in F can be constructed by combining basic functions.
- Example 1: $r_1 > r_2 > \ldots > r_I$ and $b_1 < b_2 < \ldots < b_I$

$$f_1 = \gamma_{r_1,b_1} \wedge \gamma_{r_2,b_2} \wedge \ldots \wedge \gamma_{r_I,b_I} = \min_{1 \le i \le I} \{\gamma_{r_i,b_i}\}$$



- Example 2:

$$f_2 = \lambda_R \wedge \{\beta_{R,2T} + RT\} \wedge \{\beta_{R,4T} + 2RT\} \wedge \ldots = \inf_{i \ge 0} \{\beta_{R,2iT} + iRT\}$$

Pseudo-Inverse of Wide-Sense Increasing Functions

Definition [Pseudo-inverse]

Let f be a function of F. The pseudo-inverse of f is the function

$$f^{-1}(x) = \inf\{t \text{ such that } f(t) \ge x\}.$$

- Examples:

$$\lambda_R^{-1} = \lambda_{1/R}$$
$$\delta_T^{-1} = \delta_0 \wedge T$$
$$\beta_{R,T}^{-1} = \gamma_{1/R,T}$$
$$\gamma_{r,b}^{-1} = \beta_{1/r,b}$$

Properties of Pseudo-Inverse

- (**Closure**)

$$f^{-1} \in F \quad and \quad f^{-1}(0) = 0$$

- (**Pseudo-inversion**) We have that

$$f(t) \geq x \quad \Rightarrow \quad f^{-1}(x) \leq t$$
$$f^{-1}(x) < t \quad \Rightarrow \quad f(t) \geq x$$

- **(Equivalent definition)**

$$f^{-1}(x) = \sup\{t \text{ such that } f(t) < x\}$$

Min-Plus Convolution

- Integral of function *f(t)* (*f(t)* = 0 for $t \leq 0$) in conventional algebra:

$$\int_0^t f(s)ds$$

- "Integral" for same function *f(t)* in min-plus algebra:

$$\inf_{s \in \Re \text{ such that } 0 \leq s \leq t} \{f(s)\}$$

- Convolution of two functions *f(t)* and *g(t)* that are zero for $t < 0$ in conventional algebra:

$$(f \otimes g)(t) = \int_0^t f(t-s) * g(s)ds$$

Definition **[Min-plus convolution]**

Let *f* and *g* be two functions of *F*. The min-plus convolution of *f* and *g* is the function

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

Min-Plus Convolution: Example 1

- Compute $(\gamma_{r,b} \otimes \beta_{R,T})(t)$
- Case 1: $0 \le t \le T$

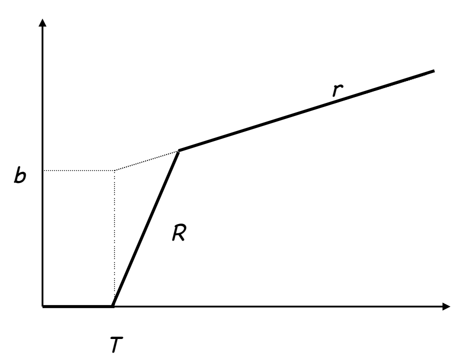
$$(\gamma_{r,b} \otimes \beta_{R,T})(t) = \inf_{0 \le s \le t} \{\gamma_{r,b}(t-s) + \beta_{R,T}(s)\}$$
$$= \inf_{0 \le s \le t} \{\gamma_{r,b}(t-s) + 0\} = \gamma_{r,b}(0) + 0 = 0$$

- Case 2: $t > T$

$$(\gamma_{r,b} \otimes \beta_{R,T})(t)$$
$$= \inf_{0 \le s \le t} \{\gamma_{r,b}(t-s) + \beta_{R,T}(s)\}$$
$$= \inf_{0 \le s \le T} \{\gamma_{r,b}(t-s) + \beta_{R,T}(s)\} \wedge \inf_{T \le s \le t} \{\gamma_{r,b}(t-s) + \beta_{R,T}(s)\} \wedge \inf_{s = t} \{\gamma_{r,b}(t-s) + \beta_{R,T}(s)\}$$
$$= \inf_{0 \le s \le T} \{b + r(t-s) + 0\} \wedge \inf_{T \le s \le t} \{b + r(t-s) + R(s-T)\} \wedge \{0 + R(t-T)\}$$
$$= \{b + r(t-T)\} \wedge \left\{b + rt - RT + \inf_{T \le s \le t} \{(R-r)s\}\right\} \wedge \{R(t-T)\}$$
$$= \{b + r(t-T)\} \wedge \{b + r(t-T)\} \wedge \{R(t-T)\}$$
$$= \{b + r(t-T)\} \wedge \{R(t-T)\}$$

Min-Plus Convolution: Example 1 (2)

$(\gamma_{r,b} \otimes \beta_{R,T})(t)$



Min-Plus Convolution: Example 2

$\delta_T \otimes \lambda_R = ?$

$$(\delta_T \otimes \lambda_R)(t) = \inf_{0 \le s \le t}\{\delta_T(t-s) + \lambda_R(s)\}$$

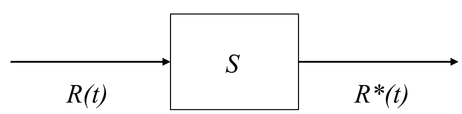
Case 1 $(0 \le t \le T)$:

$$(\delta_T \otimes \lambda_R)(t) = \inf_{0 \le s \le t}\{\delta_T(t-s) + \lambda_R(s)\} = \inf_{0 \le s \le t}\{0 + \lambda_R(s)\} = 0$$

Case 2 $(t > T)$:

$$\begin{aligned}(\delta_T \otimes \lambda_R)(t) &= (\lambda_R \otimes \delta_T)(t) \\ &= \inf_{0 \le s \le T}\{\lambda_R(t-s) + \delta_T(s)\} \\ &\wedge \inf_{T < s \le t}\{\lambda_R(t-s) + \delta_T(s)\} \\ &= \inf_{0 \le s \le T}\{\lambda_R(t-s) + 0\} \wedge \inf_{T < s \le t}\{\lambda_R(t-s) + \infty\} \\ &= \lambda_R(t-T) = \beta_{R,T}\end{aligned}$$

Models for Data Flow



- Consider system *S*: receives input data, and delivers data after a variable delay.
- *R(t)* is cumulative input function at time *t*.
- *R*(t)* is cumulative output function at time *t*.

Definition [Backlog]

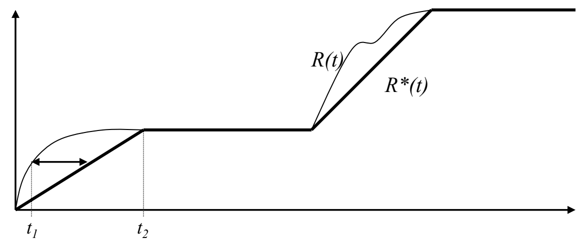
The backlog at time t is $R(t) - R^*(t)$.

Definition [Virtual Delay]

The virtual delay at time t is $d(t) = \inf\{\tau \ge 0 : R(t) \le R^*(t + \tau)\}$

Virtual Delay

$$d(t) = \inf\{\tau \geq 0 : R(t) \leq R^*(t + \tau)\}$$



- If input and output are continuous

$$R^*(t + d(t)) = R(t) \qquad (*)$$

$d(t)$ is smallest value satisfying (*)

Arrival Curves

Definition [Arrival Curve $\alpha(.)$]

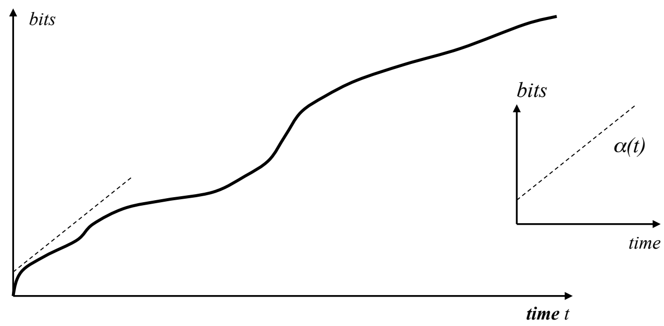
Given a wide-sense increasing function $\alpha(.)$ defined for $t \geq 0$ (i.e. $\alpha(.) \in F$) we say that a flow R is constrained by $\alpha(.)$ iff for all $s \leq t$:

$$R(t) - R(s) \leq \alpha(t - s).$$

- "R has $\alpha(.)$ as arrival curve."
- "R is bounded by $\alpha(.)$."
- "R is α-smooth."

- Note:
  - $\alpha(.)$ is in the interval-domain.
  - for all $s \geq 0$ and $I \geq 0$, $R(s + I) - R(s) \leq \alpha(I)$.

Arrival Curves (2)



Example: Affine Arrival Curve $\gamma_{r,b}$

- $\alpha(t) = rt$ Flow is peak-rate limited. For example when physical bit rate is limited.
- $\alpha(t) = b$ Maximum number of bits ever sent is at most b.
- $\alpha(t) = rt + b$ Leaky bucket with rate r and burst tolerance b.
- A leaky bucket constrains the arrival to the affine arrival curve $\gamma_{r,b} = rt + b$.

Example: Staircase Function $u_{T,\tau}$

Definition [**Generic Cell Rate Algorithm** ***GCRA(T,τ)***]

The Generic Cell Rate Algorithm (GCRA) with parameters (T,τ) is used with fixed size packets, called cells and defines conformant cells as follows: It takes as input a cell arrival time t and returns `result`. It has an internal (static) variable `tat` (theoretical arrival time).

- initially, `tat = 0`
- when a cell arrives at time t, then

```
if (t < tat - tau)
    result = NON-CONFORMANT
else {
    tat = max(t, tat) + T;
    result = CONFORMANT;
}
```

- For cells of size k, $GCRA(T,\tau)$ constrains flows to the staircase arrival function $k\ u_{T,\tau}(.)$.

Equivalence of Leaky Bucket and GCRA

For a flow with packets of constant size δ, satisfying the $GCRA(T,\tau)$ is equivalent to satisfying a leaky bucket controller with rate r and burst tolerance b given by:

$$b = (\tau/T + 1)\ \delta \qquad \text{and} \qquad r = \delta\ /\ T$$

Applications to ATM and Intserv:

- Constant Bit Rate (CBR) in ATM:
  - Single GCRA controller with parameters T (ideal cell interval) and τ (cell delay variation tolerance).
- Variable Bit Rate (VBR) in ATM:
  - Two GCRA controllers.
- Intserv: T-SPEC (p,M,r,b) with peak rate p, maximum packet size M, sustainable rate r, and burst tolerance b.

$$\alpha(t) = \min(M + pt,\ rt + b)$$

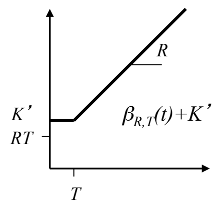
Sub-Additivity

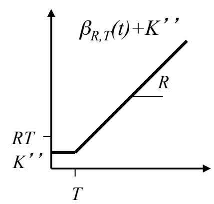
Definition [Sub-additive function]

Let f be a function of F. Then f is sub-additive iff

$$f(t + s) \leq f(t) + f(s) \text{ for all } s, t \geq 0 .$$

- Notes:
  - If $f(0) = 0$, this is equivalent to imposing that $f = f \otimes f$.
  - Concave functions passing through origin are sub-additive.
  - While concavity and convexity are simple to check visually, sub-additivity is not.





Sub-Additive Closure

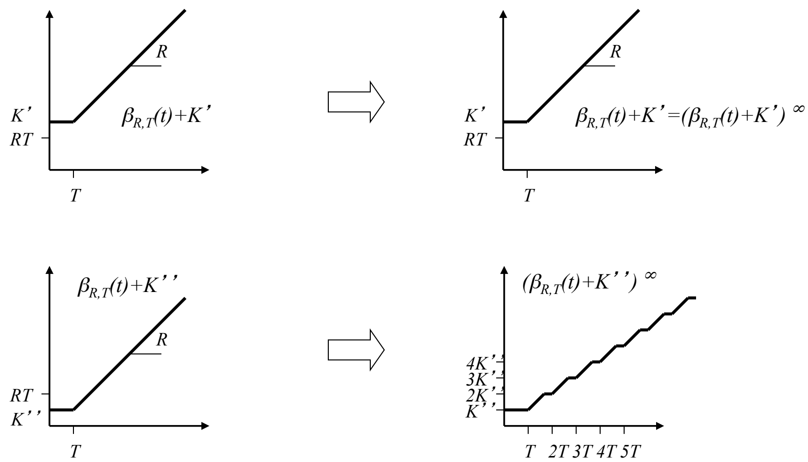
Definition [Sub-additive closure]

Let f be a function of F. Denote $f^{(n)}$ the function obtained by repeating $(n-1)$ convolutions of f with itself. By convention, $f^{(0)}=\delta_0$, so that $f(1) = f$, $f(2) = f \otimes f$, etc. Then the sub-additive closure of f, denoted by f^{∞}, is defined by

$$f^{\infty} = \delta_0 \wedge f \wedge (f \otimes f) \wedge (f \otimes f \otimes f) \wedge \ldots = \inf_{n \geq 0}\{f^{(n)}\}$$

- The sub-additive closure is the largest sub-additive function smaller than f and zero in $t = 0$.

Sub-Additive Closure: Example



Sub-Additivity and Arrival Curves

Theorem: [Reduction of Arrival Curve to a Sub-Additive One]
Saying that a flow is constrained by a wide-sense increasing funtion $\alpha(.)$ is equivalent to saying that it is constrained by the sub-additive closure $\alpha^{\infty}(.)$.

Lemma: A flow R is constrained by arrival curve α iff $R \leq R \otimes \alpha$.

Lemma: If α_1 and α_2 are arrival curves for a flow R, then so is $\alpha_1 \otimes \alpha_2$.

Min-Plus Deconvolution and Traffic Envelopes

Definition [Min-Plus Deconvolution]

Let f and g be two functions of F. The min-plus deconvolution of f by g is the function

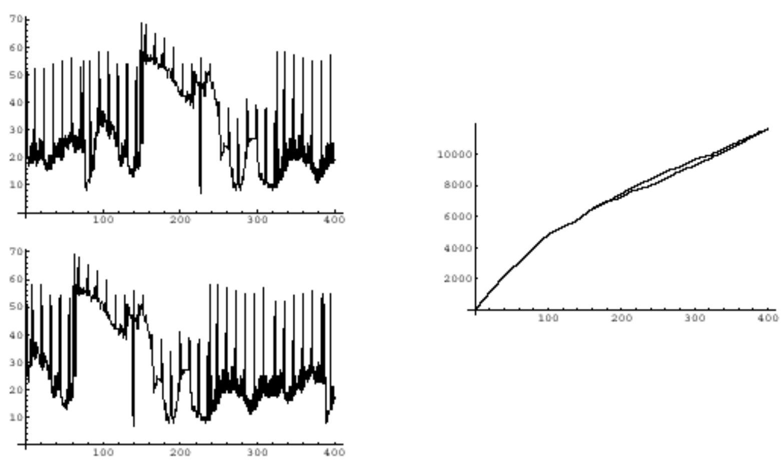
$$(f \oslash g)(t) = \sup_{u \geq 0}\{f(t + u) - g(u)\}.$$

Definition [Minimum Arrival Curve – or Envelope]

The envelope of a flow R is defined by $R \oslash R$.

By definition, we have $(R \oslash R)(t) = \sup_{v \geq 0}\{R(t + v) - R(v)\}$.

Envelopes: Examples



(Figures from J.-Y. LeBoudec and Patrick Thiran: "Network Calculus: A Theory of Deterministic Queuing Systems for the Internet", Springer Verlag Lecture Notes in Computer Science)

Service Curves

Example 1: **Generalized Processor Sharing** (GPS)

- During any busy period (flow is backlogged) of length t, flow receives at least rt amount of service.
- Input flow $R(t)$, output flow $R^*(t)$, with t_0 being the beginning of busy period for flow.

$$R^*(t) - R^*(t_0) \geq r(t - t_0)$$

- At time t_0, the backlog of flow is 0:

$$R(t_0) - R^*(t_0) = 0$$

- Therefore:

$$R^*(t) - R(t_0) \geq r(t - t_0)$$

- So:

$$R^*(t) \geq \inf_{0 \leq s \leq t}[R(s) + r(t - s)] \quad \Rightarrow \quad R^* \geq R \otimes \gamma_{r,0}$$

Service Curves

Example 2: **Guaranteed-Delay Server**

- Maximum delay for the bits of given flow R is bounded by some fixed value T, with bits of same flow served in FIFO order.

$$d(t) \leq T \quad \Leftrightarrow \quad R^*(t + T) \geq R(t)$$

- Can be re-written

$$R^*(s) \geq R(s - T) \quad \text{for all } s \geq T$$

- $R(s - T)$ can be re-written using "impulse function" δ_T:

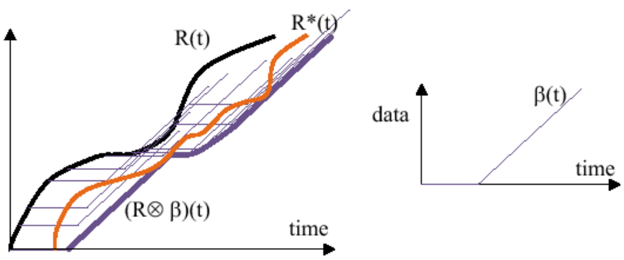
$$(R \otimes \delta_T)(t) = R(t - T)$$

- Maximum delay condition can be formulated as

$$R^* \geq R \otimes \delta_T$$

Service Curve: Definition

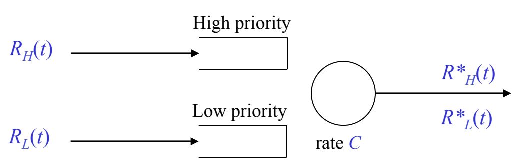
(Figures from J.-Y. LeBoudec and Patrick Thiran: "Network Calculus: A Theory of Deterministic Queuing Systems for the Internet", Springer Verlag Lecture Notes in Computer Science)



The output R^* must be above $R \otimes b$, which is the lower envelope of all curves $t \rightarrow R(t_0) + b(t - t_0)$.

Definition [Service Curve]
Consider a system S and a flow through S with input **and** output function R and R^*. We say that S offers to the flow a service curve b iff $b \in F$ and $R^* \geq R \otimes b$.

Service Curves: Non-Preemptive Priority Node



- Let s be the beginning of busy period for high-priority traffic.
- Let l_{max} be the maximum low-priority packet size.

- High-priority traffic:
- HP traffic can be blocked by a low-priority packet.

$$R_H^*(t) - R_H^*(s) \geq C(t - s) - l_{max}$$

- By definition of s: $R_H^*(s) = R_H(s)$

$$R_H^*(t) \geq R_H(s) + C(t - s) - l_{max}$$
$$R_H^*(t) \geq R_H(s) + \underbrace{\max\{0, C(t - s) - l_{max}\}}_{\text{rate-latency function with rate } C \text{ and latency } l_{max}/C}$$

Service Curves: Non-Preemptive Priority Node (2)

- Low-Priority Traffic:
- HP traffic is constrained by arrival function $\alpha_H(.)$.
- Let s' be beginning of server busy period (note that $s' \leq s$).
- At time s', backlogs for both flows are empty:

$$R_H^*(s') = R_H(s') \quad \text{and} \quad R_L^*(s') = R_L(s')$$

- Over $(s', t]$, the output is $C(t - s')$:

$$R_L^*(t) - R_L^*(s') = C(t - s') - [R_H^*(t) - R_H^*(s')]$$

$$\Rightarrow R_H^*(t) - R_H^*(s') = R_H^*(t) - R_H(s') \leq R_H(t) - R_H(s') \leq \alpha_H(t - s')$$

- $R_H^*(t) - R_H^*(s') \geq 0$

$$\Rightarrow R_L^*(t) - R_L(s') = R_L^*(t) - R_L^*(s') \geq \max\{0, C(t - s') - \alpha_H(t - s')\}$$

Network Calculus Basics: Backlog Bound

Theorem [**Backlog Bound**]

Assume a flow, constrained by arrival curve α, traverses a system that offers a service curve β. The backlog $R(t) - R^*(t)$ for all t satisfies:

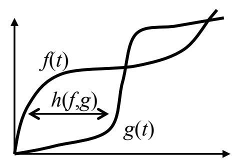
$$R(t) - R^*(t) \leq \sup_{s \geq 0}\{\alpha(s) - \beta(s)\} = (\alpha \oslash \beta)(0) .$$

Network Calculus Basics: Delay Bound

Definition [**Horizontal Deviation**]

Let f and g be two functions of F. The horizontal deviation is defined as

$h(f,g) = \sup_{t \geq 0}\{\inf\{d \geq 0 \text{ such that } f(t) \leq g(t+d)\}\}$.



Horizontal deviation can be computed using pseudo inverse:

$$g^{-1}(f(t)) = \inf\{D \text{ such that } g(D) \geq f(t)\}$$
$$= \inf\{d \geq 0 \text{ such that } g(t+d) \geq f(t)\} + t$$

$$\Rightarrow h(f,g) = \sup_{t \geq 0}\{g^{-1}(f(t)) - t\} = (g^{-1}(f) \varnothing I_1)(0).$$

Theorem [**Delay Bound**]

Assume a flow, constrained by arrival curve a, traverses a system that offers a service curve of b. The virtual delay $d(t)$ for all t satisfies: $d(t) \leq h(a, b)$.

Network Calculus Basic: Output Flow

Theorem [**Output Flow**]

Assume that a flow, constrained by arrival curve α, traverses a system that offers a service curve of β. The output flow is constrained by the arrival curve $\alpha^* = \alpha \varnothing \beta$.

Network Calculus Basics: Concatenation

Theorem [**Concatenation of Nodes**]

Assume a flow traverses systems S_1 and S_2 in sequence. Assume that S_i offers a service curve of β_i, i = 1,2 to the flow. Then the concatenation of the two systems offers a service curve of $\beta_1 \otimes \beta_2$ to the flow.

Proof:

- Call R_1 the output of node 1. This is also the input to node 2.

$$R_1 \geq R \otimes \beta_1$$

- and at node 2

$$R^* \geq R_1 \otimes \beta_2 \geq (R \otimes \beta_1) \otimes \beta_2 = R \otimes (\beta_1 \otimes \beta_2)$$

Example 1: $\beta_{R1,T1} \otimes \beta_{R2,T2} = \beta_{\min(R1,R2),T1+T2}$

Example 2: A rate-latency server can be described as

$\beta_{R,T} = (\delta_T \otimes \lambda_R)(t)$.

It can therefore be view as a concatenation of a guaranteed-delay node with delay T followed by a GPS node with rate R.