Network Calculus:

- Reference Material:
  

- Network Calculus as system theory for computer networks.

- Some mathematical background

- Arrival Curves

- Service Curves

- Network Calculus Basics

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:
  
  $y(t) = (h \ast x)(t) = \int_0^t h(t-s)x(s)ds$
Greedy Shaper

- A shaper forces an input traffic flow \( x(t) \) to have an output \( y(t) \) which adheres to an envelope \( \sigma \).

- The output function \( y(t) \) can be derived as follows:

\[
y(t) = (\sigma \otimes x)(t) = \inf_{0 \leq s \leq t} \{ \sigma(t - s) + x(s) \}
\]

- 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 \( \mathbb{R} \).

**Definition [Infimum]**

\[
\inf(S) = (M \ s.t. \ s \geq M \ \forall \ s \in S)
\]

\[
\inf(\emptyset) = +\infty
\]

**Definition [Minimum]**

\[
\min(S) = (M \in S \ s.t. \ s \geq M \ \forall \ s \in S)
\]

- Notation: \( ^\wedge \) denotes infimum (e.g. \( a ^\wedge b = \min\{a, b\} \))
The Diod \((\mathbb{R} \cup \{+\infty\}, \land, +)\)

- Conventional ("plus-times") algebra operates on algebraic structure \((\mathbb{R}, +, \ast)\).
- Min-plus algebra replaces operations:
  - "addition" becomes "computation of infimum"
  - "multiplication" becomes "addition"
- Resulting algebraic structure becomes \((\mathbb{R} \cup \{+\infty\}, \land, +)\)

Example:
- Conventional algebra: 
  \((3+4) \ast 5 = (3\ast5) + (4\ast5) = 15 + 20\)
- Min-plus algebra: 
  \((3 \land 4) + 5 = (3 + 5) \lor (4 + 5) = 8 \lor 9 = 8\)

Properties of \((\mathbb{R} \cup \{+\infty\}, \land, +)\)

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

**Definition**
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** \( \lambda_R \): “Rate” \( R \)
  \[
  \lambda_R(t) = \begin{cases} 
  Rt & \text{if } t > 0 \\
  0 & \text{otherwise}
  \end{cases}
  \]

- **Burst delay function** \( \delta_T \): “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 $\nu_{T}$:
  \[
  \nu_{T}(t) = \begin{cases} 
  1 & \text{if } t > T \\
  0 & \text{otherwise}
  \end{cases}
  \]

- Staircase function $u_{T,\tau}$:
  “Interval” $T$, “Tolerance” $\tau$
  \[
  u_{T,\tau}(t) = \begin{cases} 
  \left\lfloor \frac{t+\tau}{T} \right\rfloor & \text{if } t > 0 \\
  0 & \text{otherwise}
  \end{cases}
  \]
Wide-Sense Increasing Functions (4)

- More general functions in $F$ can be constructed by combining basic functions.

- Example 1: $r_1 > r_2 > \ldots > r_l$ and $b_1 < b_2 < \ldots < b_l$
  
  $f_1 = \gamma_{r_1,b_1} \wedge \gamma_{r_2,b_2} \wedge \ldots \wedge \gamma_{r_l,b_l} = \min\{\gamma_{r_i,b_i}\}$

- Example 2:
  
  $f_2 = \lambda_\infty \land (\beta_{R,2T} + RT) \land (\beta_{R,2T} + 2RT) \land \ldots$
  
  $= \inf\{\beta_{R,2T} + iRT\}$

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 

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

- Examples:
  
  $\lambda_R^{-1} = \lambda_{1/R}$
  
  $\delta_t^{-1} = \delta_0 \land T$
  
  $\beta_{R,T}^{-1} = \gamma_{R,RT}$
  
  $\gamma_{r,b}^{-1} = \beta_{1/r,b}$
### Properties of Pseudo-Inverse

- **(Closure)**

  
  \[ f^{-1} \in F \text{ and } f^{-1}(0) = 0 \]

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

  \[
  \begin{align*}
  f(t) \geq x & \implies f^{-1}(x) \leq t \\
  f^{-1}(x) < t & \implies f(t) \geq x
  \end{align*}
  \]

- **(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:

  \[ \int_0^\infty f(s) ds \]

- “Integral” for same function \( f(t) \) in min-plus algebra:

  \[ \inf_{s \in \mathbb{R}, s + t \geq 0} \{ f(s) \} \]

- 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: \(0 \leq t \leq T\)
  \[ (\gamma_{r,b} \otimes \beta_{R,T})(t) = \inf_{0 \leq s \leq t} \{ \gamma_{r,b}(t-s) + \beta_{R,T}(s) \} \]
  \[ = \inf_{0 \leq s \leq t} \{ \gamma_{r,b}(t-s) + 0 \} = \gamma_{r,b}(0) + 0 = 0 \]

Case 2: \(t > T\)
\[
(\gamma_{r,b} \otimes \beta_{R,T})(t) \\
= \inf_{0 \leq s \leq t} \{ \gamma_{r,b}(t-s) + \beta_{R,T}(s) \} \\
= \inf_{0 \leq s \leq t} \{ \gamma_{r,b}(t-s) + \beta_{R,T}(s) \} \wedge \inf_{0 \leq s \leq t} \{ \gamma_{r,b}(t-s) + \beta_{R,T}(s) \} \\
= \inf \{ b + r(t-s) + 0 \} \wedge \inf \{ b + r(t-s) + R(s-T) \} \wedge \{ 0 + R(t-T) \} \\
= \{ b + r(t-T) \} \wedge \{ b + r(t-T) \} \wedge \{ R(t-T) \} \\
= \{ b + r(t-T) \} \wedge \{ R(t-T) \} \\
= \{ b + r(t-T) \} \wedge \{ R(t-T) \} \\
= \{ b + r(t-T) \} \wedge \{ R(t-T) \}
\]

Min-Plus Convolution: Example 1 (2)

![Diagram of Min-Plus Convolution Example 1](image-url)
Min-Plus Convolution: Example 2

\[ \delta(t) \otimes \lambda_n = ? \]

\[ (\delta(t) \otimes \lambda_n)(t) = \inf_{s \in [0,t]} \{ \delta(t-s) + \lambda_n(s) \} \]

Case 1 \((0 \leq t \leq T)\) :

\[ (\delta(t) \otimes \lambda_n)(t) = \inf_{s \in [0,t]} \{ \delta(t-s) + \lambda_n(s) \} \]

\[ = \inf_{s \in [0,\infty)} \{ 0 + \lambda_n(s) \} = 0 \]

Case 2 \((t > T)\) :

\[ (\delta(t) \otimes \lambda_n)(t) = (\lambda_n \otimes \delta(t))(t) \]

\[ = \inf_{s \in [0,t]} \{ \lambda_n(t-s) + \delta(s) \} \]

\[ \wedge \inf_{s \in [t,\infty)} \{ \lambda_n(t-s) + \delta(s) \} \]

\[ = \inf_{s \in [0,\infty)} \{ \lambda_n(t-s) + 0 \} \wedge \inf_{s \in [t,\infty)} \{ \lambda_n(t-s) + \infty \} \]

\[ = \lambda_n(t-T) = \beta_{n,\infty} \]

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

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

- If input and output are continuous
  \[ R^*(t + d(t)) = R(t) \]  (*)
  \( d(t) \) 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 \( \alpha \)-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($T, \tau$)]**

The Generic Cell Rate Algorithm (GCRA) with parameters ($T, \tau$) 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 $k u_{T,\tau}(\cdot)$.

**Equivalence of Leaky Bucket and GCRA**

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

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

Applications to ATM and Intserv:

- **Constant Bit Rate (CBR) in ATM:**
  - Single GCRA controller with parameters $T$ (ideal cell interval) and $\tau$ (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 $f^\infty$, is defined by

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

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

Sub-Additive Closure: Example

\[ R_{T}(t)+K' = (\beta_{R,T}(t)+K')^{\infty} \]

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 function \( \alpha(.) \) is equivalent to saying that it is constrained by the sub-additive closure \( \alpha^{\infty}(.) \).

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

Lemma: If \( \alpha_1 \) and \( \alpha_2 \) 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

\[
(f \triangledown g)(t) = \sup_{u \geq 0} \{f(t + u) - g(u)\}.
\]

**Definition [Minimum Arrival Curve – or Envelope]**

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

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

**Envelopes: Examples**

(Figures from J.-Y. Le Boudec 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:
  \[ R^*(t) \geq \inf_{0 \leq s \leq t} [R(s) + r(t - s)] \implies R^* \geq R \otimes \gamma_{t,0} \]

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.
  \[ d(t) \leq T \iff R^*(t + T) \geq R(t) \]
- 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” $\delta_T$:
  \[ (R \otimes \delta_T)(t) = R(t - T) \]
- Maximum delay condition can be formulated as
  \[ R^* \geq R \otimes \delta_T \]
Service Curve: Definition

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

**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 $\beta$ iff $\beta \in F$ and $R^* \geq R \otimes \beta$.

Service Curves: Non-Preemptive Priority Node

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

- **High-priority traffic:**
- HP traffic can be blocked by a low-priority packet.
  \[ R^*_H(t) - R^*_H(s) \geq C(t - s) - \ell_{\text{max}} \]
- By definition of $s$: $R^*_H(s) = R_H(s)$
  \[ R^*_H(t) \geq R_H(s) + C(t - s) - \ell_{\text{max}} \]
  \[ R^*_H(t) \geq R_H(s) + \max\{0, C(t - s) - \ell_{\text{max}}\} \]

  rate-latency function with rate $C$ and latency $\ell_{\text{max}}/C$
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 $\alpha$ traverses a system that offers a service curve $\beta$. The backlog $R(t) - R^*(t)$ for all $t$ satisfies:

\[ R(t) - R^*(t) \leq \sup_{s \geq 0} \{ \alpha(s) - \beta(s) \} = (\alpha \ominus \beta)(0). \]
Network Calculus Basics: Delay Bound

**Definition [Horizontal Deviation]**

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

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

Horizontal deviation can be computed using pseudo inverse:

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

$$\Rightarrow h(f, g) = \sup_{t \geq 0} \{ g^{-1}(f(t)) - t \} = (g^{-1}(f) \ominus \lambda_1)(0).$$

**Theorem [Delay Bound]**

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

Network Calculus Basic: Output Flow

**Theorem [Output Flow]**

Assume that a flow, constrained by arrival curve $\alpha$, traverses a system that offers a service curve of $\beta$. The output flow is constrained by the arrival curve $\alpha^* = \alpha \ominus \beta$. 
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 $\beta_i$, $i=1,2$ 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 viewed as a concatenation of a guaranteed-delay node with delay $T$ followed by a GPS node with rate $R$. 
