Applications of the duality of min-plus and max-plus network calculus

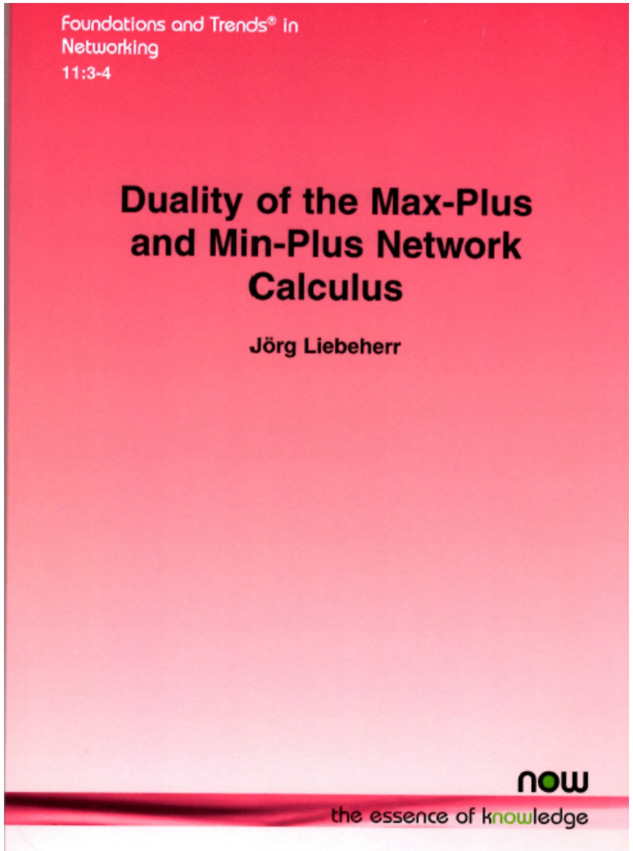
Jörg Liebeherr

University of Toronto

(WoNeCa 2018)

Background for this talk

- J. Liebeherr, "Duality of the Max-Plus and Min-Plus Network Calculus," Foundations and Trends in Networking, Vol. 11, No. 3-4, pp. 139–282, 2017.



Available from my home page (see: Publications)

Originally ….

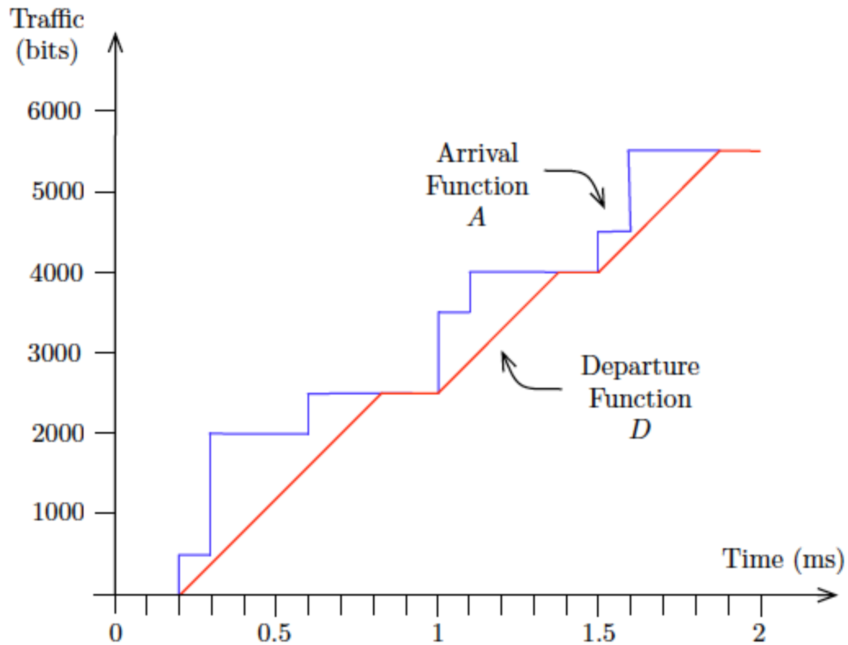
… I wanted to write an elementary introduction to max-plus network calculus for a course,

- relate it to the min-plus version, and
- discuss applications to scheduling with rate guarantees

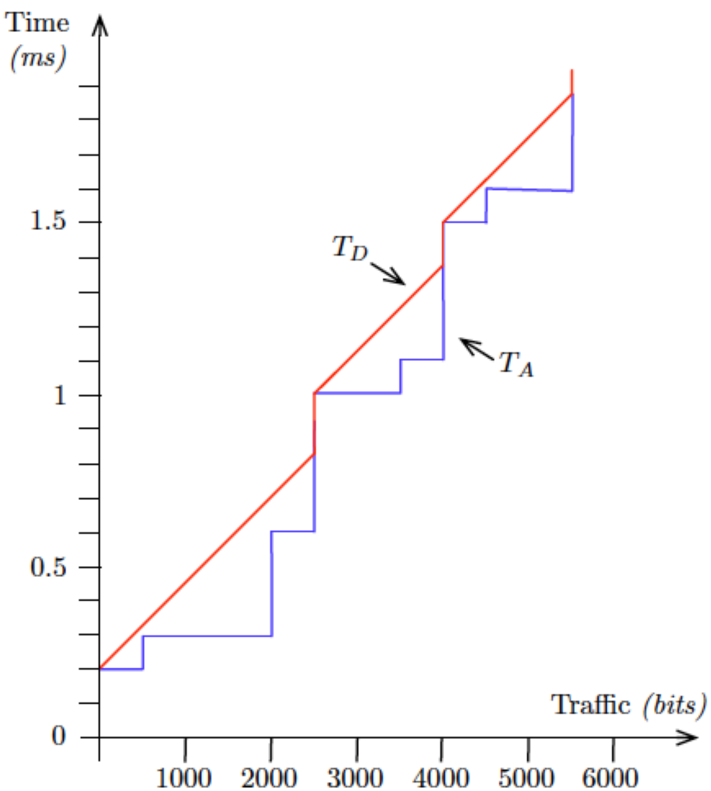
- This was supposed to be easy since:
  - The (min, +)- and (max, +)-dioids are isomorphic
  - Operations of the min-plus and max-plus network calculus are well-understood
  - Many have worked with concepts in both algebras

Min-Plus and Max-Plus Network Calculus

- Min-plus: Arrival, departures, service are functions of time.
- Max-plus: Arrival, departures, service are functions of space.
- Functions are related by a reflection at the diagonal!

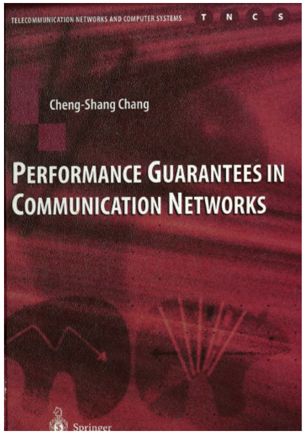


Time Domain.

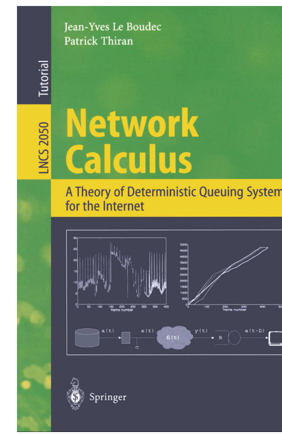


Space Domain.

But then …



Remark 6.2.7. We note that not every result in the min-plus algebra can be extended here. For example, a concatenation of the minimal g_1-regulator and the minimal g_2-regulator is not the minimal $g_1 \odot g_2$-regulator in general.



More specifically, there is not an exact correspondence between the set of flows that are g-regular on one hand, and that are σ-smooth on the other. We explain why with an example.

IEEE COMMUNICATIONS
SURVEYS & TUTORIALS
A PUBLICATION OF THE IEEE COMMUNICATIONS SOCIETY

IEEE ComSoc IEEE Communications Society

M. Fidler, First Quarter, 2010.

In the sequel we will restrict our exposition to min-plus systems theory and only use the max-plus approach where it is particularly useful. We note, however, that many concepts can be mirrored in the max-plus algebra

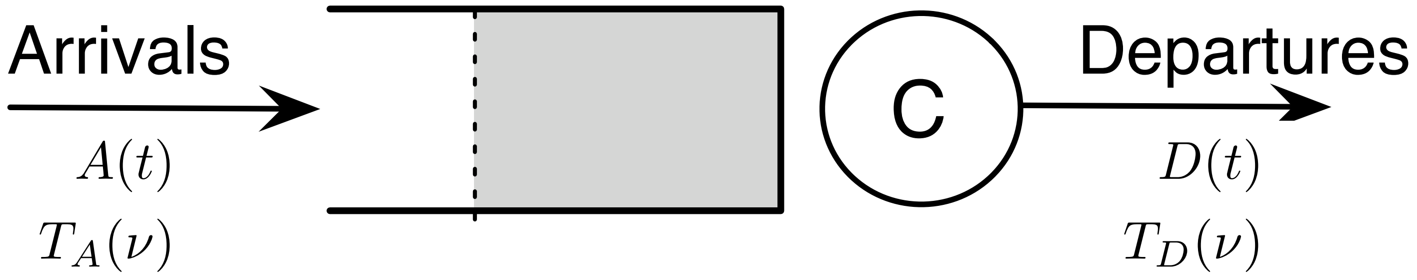
Notation

Min-plus	
$A(t)$	Arrivals until time t
$D(t)$	Departures until time t
$W(t)$	virtual delay at time t
$B(t)$	Backlog at time t
$S(t)$	Service curve
$E(t)$	Envelope ('arrival curve')
$\oslash, \otimes$	(De-) Convolution
$F \in \mathcal{F}_o$	f is left-continuous, non-decreasing, $F(t) = 0$ if $t \leq 0$

Max-plus	
$T_A(\nu)$	Arrival time of bit ν
$T_D(\nu)$	Departure time of bit ν
$W(\nu)$	Delay of bit ν:
$B^a(\nu)$	Backlog at arrival of ν
$B^d(\nu)$	Backlog at departure of ν
$\gamma_S(\nu)$	Service curve
$\lambda_E(\nu)$	Envelope ('arrival curve')
$\overline{\oslash}, \overline{\otimes}$	(De-) Convolution
$F \in \mathcal{T}_o$	f is right-continuous, non-decreasing, $F(t) = -\infty$ if $t < 0$

Buffered Link

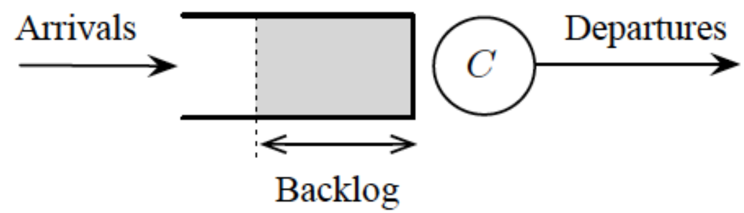
- Work-conserving link with fixed rate C



- Offers an exact (min-plus) service curve: $S(t) = Ct$
  such that $D(t) = A \otimes S(t)$

- The corresponding max-plus service curve should be: $\gamma_S(\nu) = \frac{\nu}{C}$
  with $T_D(\nu) = T_A \overline{\otimes} \gamma_S(\nu)$

Buffered Link

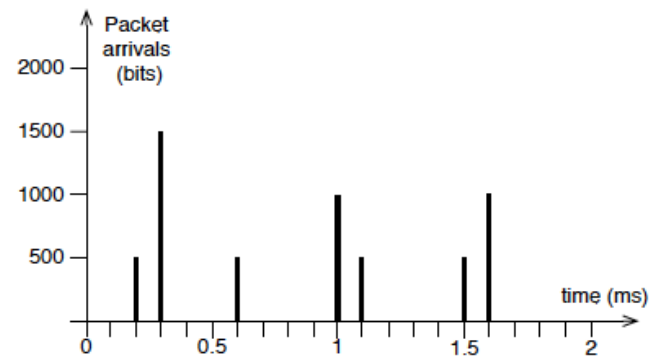


$T_A^p(n)$ Arrival time of n-th packet

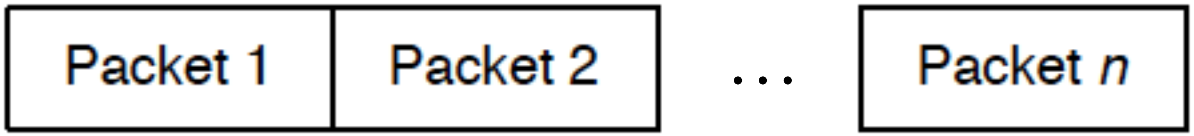
$T_D^p(n)$ Departure time of n-th packet

ℓ_n Size of n-th packet (in bits)

$L_n = \sum_{j=1}^{n} \ell_n$



Recursion for Departure time of nth Packet



$$
\begin{aligned}
T_D^p(n) &= \max\left\{T_A^p(n), T_D^p(n-1)\right\} + \frac{\ell_n}{C} \\
&= \max\left\{T_A^p(n) + \frac{\ell_n}{C}\ ,\ T_A^p(n-1) + \frac{\ell_{n-1}+\ell_n}{C}\ ,\ \ldots\right. \\
&\qquad\qquad \left.\ldots\ ,\ T_A^p(1) + \frac{\ell_1 + \ldots + \ell_n}{C}\right\} \\
&= \max_{0 \le k \le n-1}\left\{T_A^p(n-k) + \frac{\ell_{n-k} + \ldots + \ell_n}{C}\right\}
\end{aligned}
$$

with $T_D^p(0) = 0$.

Bit-level View of Packets

ℓ_n	Size of n-th packet (in bits)
L_n	$= \sum_{j=1}^{n} \ell_n$

We number the bits of the packets

$$\underbrace{0, 1, \ldots, \ell_1 - 1}_{\text{Packet 1}}, \underbrace{\ell_1, \ldots, L_2 - 1}_{\text{Packet 2}}, \quad \cdots \quad, \underbrace{L_{n-1}, \ldots, L_n - 1}_{\text{Packet } n}$$

Departure of bit ν:

$$\begin{aligned} T_D(\nu) &= \max\Big\{T_A(\nu), T_D(\nu - 1)\Big\} + \frac{1}{C} \\ &= \max\Big\{T_A(\nu) + \frac{1}{C}, T_A(\nu - 1) + \frac{2}{C}, \ldots, T_A(0) + \frac{\nu}{C}\Big\} \\ &= \max_{\kappa=0,1,\ldots,n} \Big\{T_A(\nu - \kappa) + \frac{\kappa + 1}{C}\Big\} \end{aligned}$$

Bit-level View of Packets

With

$$F \overline{\otimes} G(\nu) = \max_{\kappa=0,1,\ldots,\nu} \{F(\nu-\kappa) + G(\kappa)\},$$

we get either

$$T_D(\nu) = T_A \overline{\otimes} \gamma_S(\nu) \text{ with } \quad \gamma_S(\nu) = \frac{\nu+1}{C}$$

or

$$T_D(\nu) = T_A \overline{\otimes} \gamma'_S(\nu) + \frac{1}{C} \text{ with } \quad \gamma'_S(\nu) = \frac{\nu}{C}$$

Towards a Continuous-Space View

If we measure traffic in $\frac{1}{k}$-th of a bit:

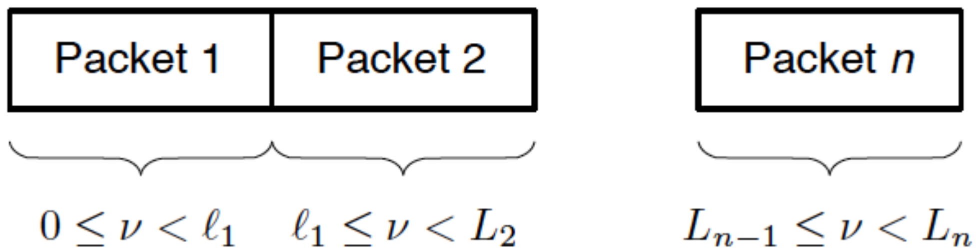
$$
\begin{aligned}
T_D(\nu) &= \max\{T_A(\nu), T_D(\nu - \tfrac{1}{k})\} + \frac{1}{kC} \\
&= \max_{\kappa=0,\frac{1}{k},\frac{2}{k}\dots,\nu} \left\{ T_A(\nu-\kappa) + \frac{\kappa + \frac{1}{k}}{C} \right\} \\
&= T_A \overline{\otimes} \gamma_S(\nu) + \frac{1}{k} \qquad \text{with } \gamma_S(\nu) = \tfrac{\nu}{C}
\end{aligned}
$$

For $k \to \infty$:

$$
T_D(\nu) = T_A \overline{\otimes} \gamma_S(\nu) \qquad \text{with } \gamma_S(\nu) = \tfrac{\nu}{C}
$$

Continuous-space View of Packets

Viewing bits as real numbers:



$$
\begin{aligned}
T_D(L_n^-) &= \sup_{0 \le \kappa \le L_n^-} \left\{ T_A(L_n^- - \kappa) + \frac{\kappa}{C} \right\} \\
&= T_A \overline{\otimes} \gamma(L_n^-)
\end{aligned}
$$

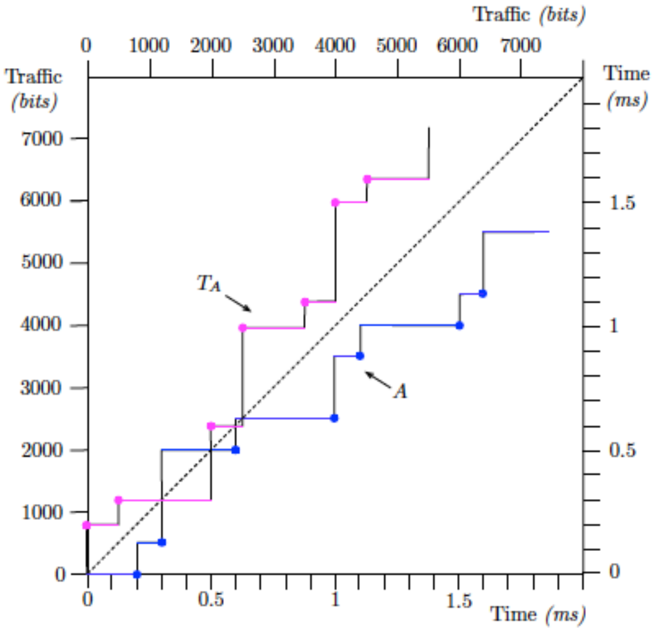
with $\gamma_S(\nu) = \frac{\nu}{C}$

Zwischenfazit (Interim conclusions)

- Continuous-space view results in: $S(t) = Ct \leftrightarrow \gamma_S(\nu) = \frac{\nu}{C}$
- In a packet-level or bit-level view:
  - 'Extra term' $\frac{\ell_n}{C}$ for packets (or $\frac{1}{C}$ for bits) reflects a packetization (or 'bit'-ization)
  - 'Extra term' is the root cause for reported discrepancies between min-plus and max-plus network calculus
- **Next:** Continuous-space max-plus NC and continuous-time min-plus NC are isomporphic $\Rightarrow$ Pseudo-inverse functions

Motivation for Pseudo-inverses

- A and the T_A are diagonal reflections of each other
- If functions are continuous and strictly increasing, diagonal reflection are the inverses
- Since A and T_A are neither, inverse functions do not exist
  $\Rightarrow$ Pseudo-inverse functions



Pseudo-inverses

For a non-decreasing function F:

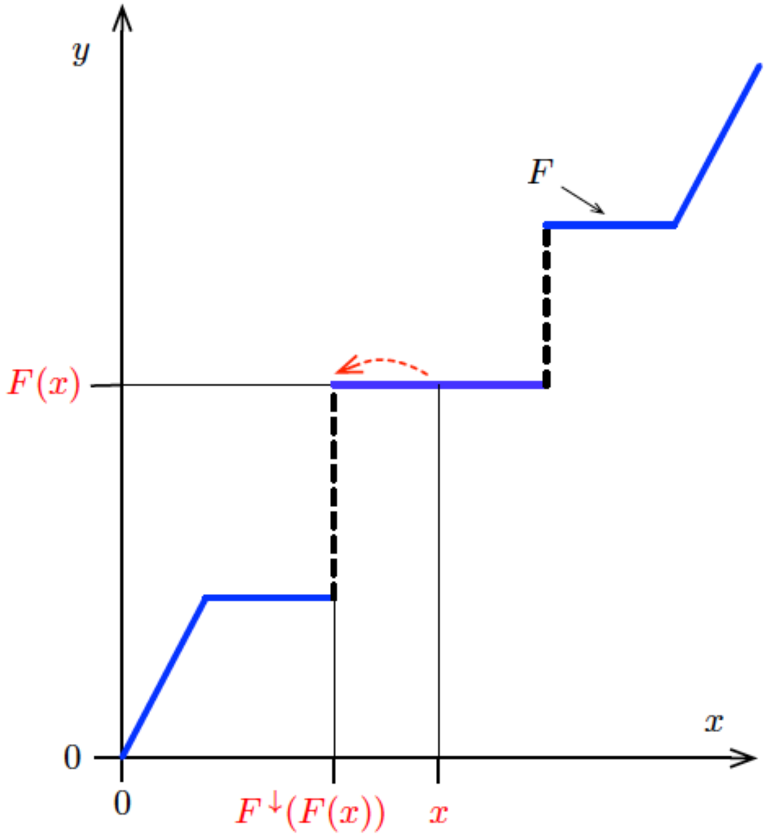
- Lower pseudo-inverse:

$$F^{\downarrow}(y) = \inf\{x \mid F(x) \geq y\} = \sup\{x \mid F(x) < y\}$$

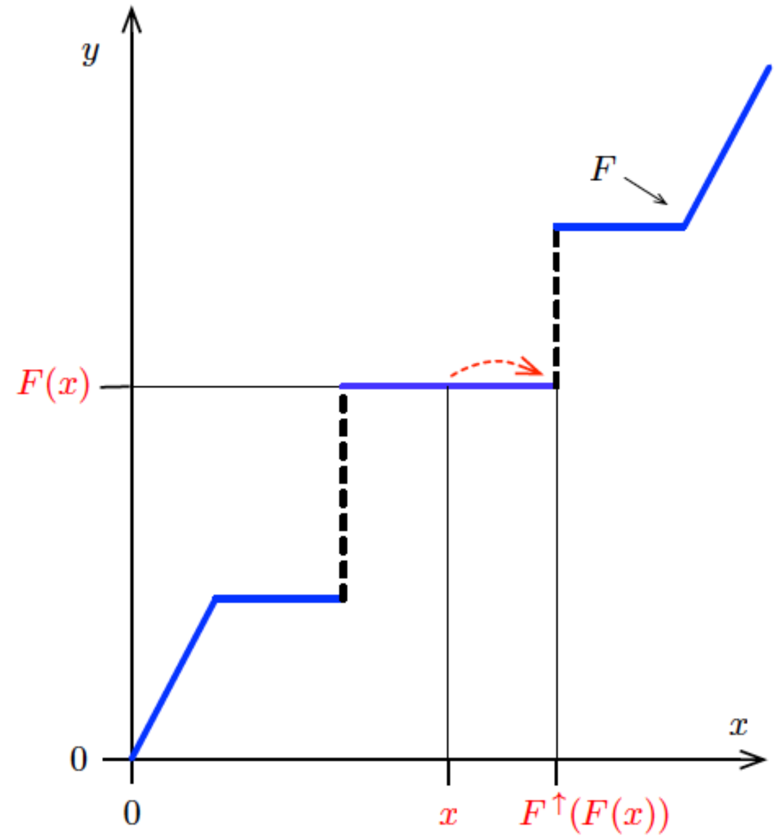
- Upper pseudo-inverse:

$$F^{\uparrow}(y) = \sup\{x \mid F(x) \leq y\} = \inf\{x \mid F(x) > y\}$$

Pseudo-inverses



Lower pseudo-inverse



Upper pseudo-inverse

Properties of pseudo-inverses

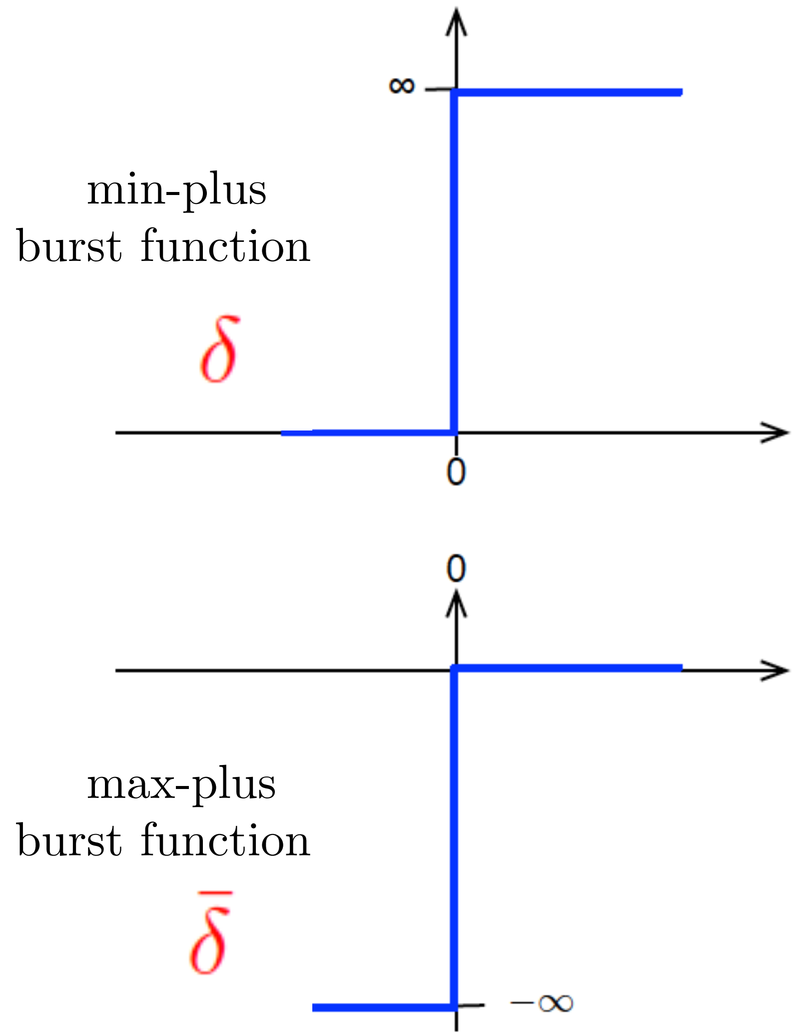
For non-decreasing functions F and G:

1. $F^{\downarrow}$ and $F^{\uparrow}$ are non-decreasing
2. $F^{\downarrow}$ is left-continuous and $F^{\uparrow}$ is right-continuous
3. F is left-continuous $\Rightarrow F = \left(F^{\uparrow}\right)^{\downarrow}$
4. F is right-continuous $\Rightarrow F = \left(F^{\downarrow}\right)^{\uparrow}$
5. Order-reversing: $F \geq G \Rightarrow F^{\uparrow} \leq G^{\uparrow}$, $F^{\downarrow} \leq G^{\downarrow}$

$$A = \left(A^{\uparrow}\right)^{\downarrow} = T_A^{\downarrow}$$
$$T_A = \left(T_A^{\downarrow}\right)^{\uparrow} = A^{\uparrow}$$

Mapping functions between time and space domain

1. $F \in \mathcal{F}_o \Rightarrow F^{\uparrow} \in \mathcal{T}_o$
2. $F \in \mathcal{T}_o \Rightarrow F^{\downarrow} \in \mathcal{F}_o$
3. $\delta^{\uparrow} = \bar{\delta}$ and $\bar{\delta}^{\downarrow} = \delta$.



Mapping between min-plus and max-plus algebras

Min-plus $\rightarrow$ Max-plus:

1. $(F \wedge G)^{\uparrow} = F^{\uparrow} \vee G^{\uparrow}$
2. $(F \otimes G)^{\uparrow} = F^{\uparrow} \overline{\otimes} G^{\uparrow}$
3. $(F \oslash G)^{\uparrow} = F^{\uparrow} \overline{\oslash} G^{\uparrow}$
4. $(F + G)^{\uparrow}(\nu) = \inf_{0 \leq \kappa \leq \nu} \max\{F^{\uparrow}(\kappa), G^{\uparrow}(\nu - \kappa)\}$
5. $F \in \mathcal{F}_o$ subadditive $\Rightarrow F^{\uparrow}$ superadditive

Max-plus $\rightarrow$ Min-plus:

1. $(F \vee G)^{\downarrow} = F^{\downarrow} \wedge G^{\downarrow}$.
2. ...

Mapping traffic envelopes

Notation:

$$\left\{\begin{array}{ll} A \sim E : & E \\ T_A \sim \lambda_E : & \lambda_E \end{array}\right\} \text{ is a } \left\{\begin{array}{l}\text{min-plus} \\ \text{max-plus}\end{array}\right\} \text{ traffic envelope for } \left\{\begin{array}{c} A \\ T_A \end{array}\right\}$$

1. $A \sim E \Longrightarrow A^{\uparrow} \sim E^{\uparrow}$
2. $T_A \sim \lambda_E \Longrightarrow {T_A}^{\downarrow} \sim {\lambda_E}^{\downarrow}$

Example: Token Bucket

$$E(t) = b + rt \quad \Rightarrow \quad E^{\uparrow}(\nu) = \left[\frac{\nu - b}{r}\right]^{+}$$

Mapping service curves

1. $D = A \otimes S \Rightarrow D^{\uparrow} = A^{\uparrow} \overline{\otimes} S^{\uparrow}$
2. $D \geq A \otimes S \Rightarrow D^{\uparrow} \leq A^{\uparrow} \overline{\otimes} S^{\uparrow}$
3. $D \leq A \otimes S \Rightarrow D^{\uparrow} \geq A^{\uparrow} \overline{\otimes} S^{\uparrow}$

1. $T_D = T_A \overline{\otimes} \gamma_S \Rightarrow T_D^{\downarrow} = T_A^{\downarrow} \otimes \gamma_S^{\downarrow}$
2. $T_D \leq T_A \overline{\otimes} \gamma_S \Rightarrow T_D^{\downarrow} \geq T_A^{\downarrow} \otimes \gamma_S^{\downarrow}$
3. $T_D \geq T_A \overline{\otimes} \gamma_S \Rightarrow T_D^{\downarrow} \leq T_A^{\downarrow} \otimes \gamma_S^{\downarrow}$

Example: Latency-rate service curve

$$S(t) = R(t-T)I_{t>T} \quad \Rightarrow \quad S^{\uparrow}(\nu) = \begin{cases} -\infty\,, & \text{if } \nu < 0\,, \\ \dfrac{\nu}{R} + T\,, & \text{if } \nu \geq 0\,. \end{cases}$$

Example: Residual service curve

$$S(t) = [Ct - E_c(t)]^{+} \quad \Rightarrow \quad S^{\uparrow}(\nu) = \frac{1}{C}\left(\inf\left\{x \mid \lambda_c(x) \geq \frac{x+\nu}{C}\right\} + \nu\right)$$

Why do we care?

- Backlog easier with min-plus: $B(t) = A(t) - D(t)$
  Delay is easier with max-plus: $W(\nu) = T_D(\nu) - T_A(\nu)$
- **Aggregate of flows:**

$$\boxed{A(t) = \sum_{j=1}^{N} A_j(t)} \qquad \textit{(min-plus)}$$

$$\boxed{T_A(\nu) = \inf_{\substack{\nu_1,\ldots,\nu_N \\ \nu=\nu_1+\ldots+\nu_N}} \max_{j=1,\ldots,N} T_{A_j}(\nu_j)} \qquad \textit{(max-plus)}$$

Min-plus and max-plus network calculus are complementary:

- Capacity provisioning is easier with min-plus network calculus
- Traffic algorithms are easier in a max-plus view

Example: SCED

- **Computing timestamps:** Deadline computation at a SCED scheduler :

$$\boxed{D\ell(\nu) = T_A \overline{\otimes} \gamma_S(\nu)}$$ *(max-plus)*

vs.

$$\boxed{D\ell(t) = \sup\{x | A \otimes S(x) \leq A(t)\}}$$ *(min-plus)*

$$= (A \otimes S)^{\uparrow}(A(t))$$

- **Schedulability:** Condition for SCED schedulability:

$$\boxed{\inf_{\substack{\nu_1,\ldots,\nu_N \\ \nu=\nu_1+\ldots+\nu_N}} \max_{j=1,\ldots,N} \lambda_{E_j} \overline{\otimes} \gamma_{S_j}(\nu_j) \geq \frac{\nu}{C}, \quad \forall \nu \geq 0}$$ *(max-plus)*

vs.

$$\boxed{\sum_{j=1}^{N} E_j \otimes S_j(t) \leq Ct, \quad \forall t \geq 0}$$ *(min-plus)*

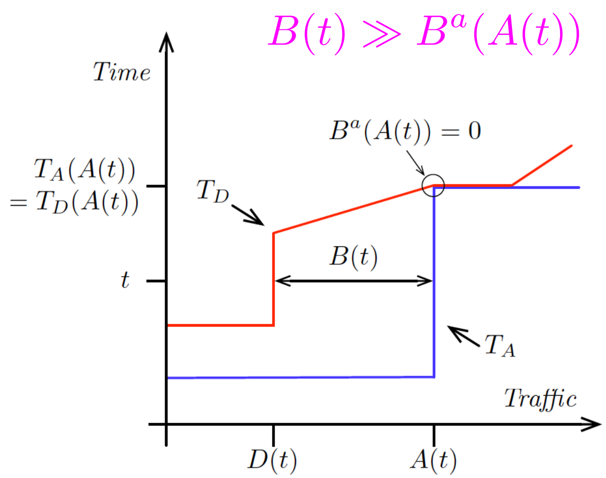
Everything maps nicely, right?
Not quite!

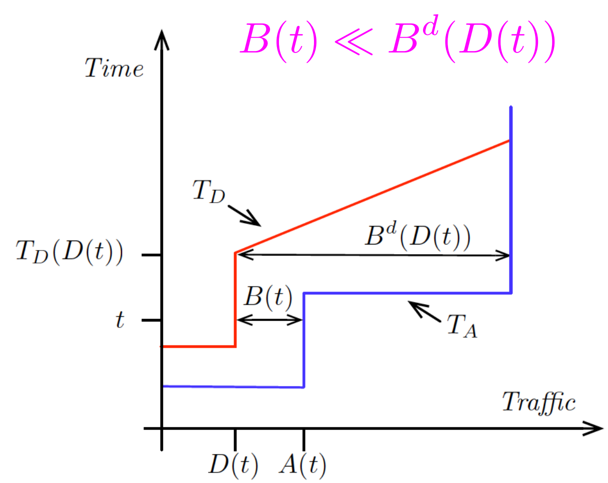
Backlog and Delay

- Backlog and delay cannot be mapped exactly with pseudo-inverses
- We can only provide bounds, e.g.,

$$B^a(A(t)) \leq B(t) \leq B^d(D(t))$$

which can be quite loose:





- Good news: If A and D are continuous at $T_A(\nu)$ then

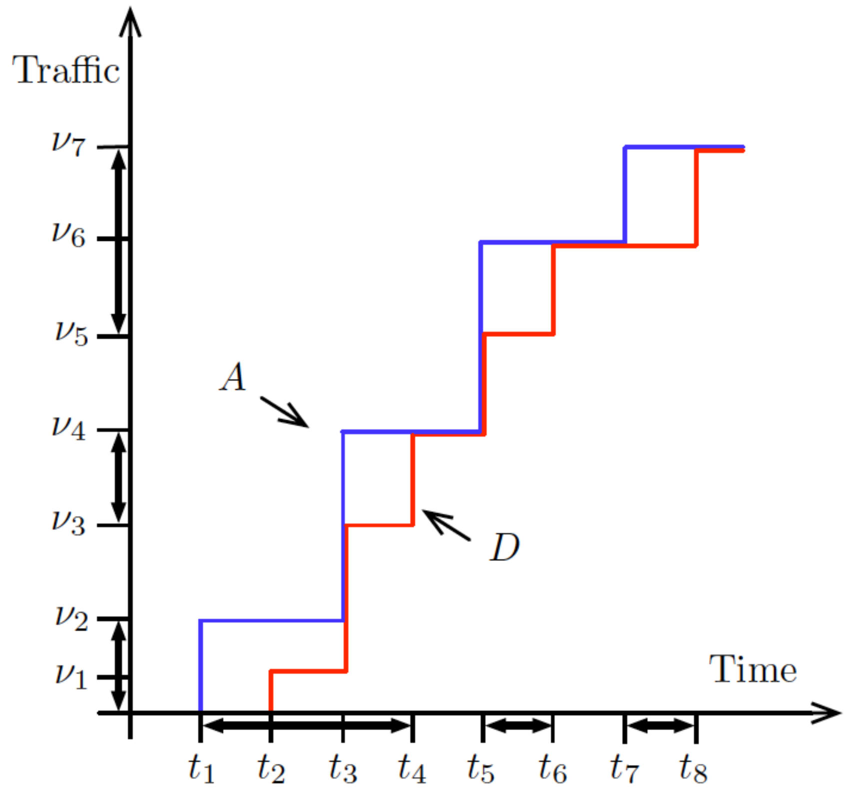
$$B(T_A(\nu)) = B^a(\nu)$$

Busy periods and busy sequences

- In general, busy periods cannot be described using expressions of the max-plus network calculus
- We define the concept of busy sequence as a maximal sequence of bits with non-zero delays
- Busy sequence also helps with defining a strict max-plus service curve

Problems with busy periods and sequences

- A single busy periods may cover multiple busy sequences
- A single busy sequence may cover multiple busy periods



Summary

- Clarification of the relationship between min-plus and max-plus network calculus
  - Dispenses with the frequently made assumption of constant packet sizes for a max-plus analysis
- Now: Can switch between a min-plus or max-plus viewpoint in the same analysis
- Filled a few holes in the max-plus literature, e.g.,
  - busy sequence
  - strict max-plus service curve
  - adaptive max-plus service curve

Supplemental Slides

Strict max-plus service curve

A strict max-plus service curve $\gamma_S \in \mathcal{T}_o$ satisfies for all ν and μ in the same busy sequence, it holds that

start of bsy sequence

$$T_D(\nu) - T_D(\mu) \leq \gamma_S(\nu - \mu)\,, \text{ if } \underline{\nu} < \mu\,,$$

$$\text{and } T_D(\nu) - T_A(\mu) \leq \gamma_S(\nu - \mu)\,, \text{ if } \underline{\nu} = \mu\,.$$

- Cannot define (general) strict max-plus service curve with busy period !

Adaptive max-plus service curve

An adaptive max-plus service curve γ_S for a network element satisfies for all $\nu \geq 0$,

$$T_D(\nu) \leq \inf_{\mu \leq \nu} \left\{ \max \left[T_D(\mu) + \gamma_S(\nu - \mu) \, , T_A \overline{\otimes}_{\mu} \gamma_S(\nu) \right] \right\} .$$

where

$$F \overline{\otimes}_{\mu} G(\nu) = \sup_{\mu \leq \kappa \leq \nu} \left\{ F(\kappa) + G(\nu - \kappa) \right\}$$

Max-plus convolution by Chang/Lin

- Chapter 6 in Chang's Book
- Let F and G be non-decreasing real-valued functions, and $H(n)$ a non-decreasing integer-valued function:

$$F \odot_H G(n) = \max_{0 \leq k \leq n} \{F(k) + G(H(n) - H(k))\} ,$$

- Setting $\ell'_n = \ell_{n+1}$ and $L'(n) = \sum_{k=0}^{n-1} \ell'_k$ the output of the buffered link is

$$T_D^p(n+1) = \tau \odot_{L'} \gamma_S(n) + \frac{\ell'_n}{C}$$

with $\gamma_S(\nu) = \frac{\nu}{C}$.

- If $\ell_n = 1$, operations $\vee$ and $\odot_{L'}$ yield a dioid.

Mappings of Dioids

- See Chapter 4, in "Synchronization and Linearity: ...".
- Applies residuation theory for lattices to establish isomorphisms between dioids
- Terminology:

residual	$\longrightarrow$	upper pseudo-inverse
dual residual	$\longrightarrow$	lower pseudo-inverse
isotone mapping	$\longrightarrow$	non-decreasing function
isotone and upper semi-continuous	$\longrightarrow$	right-continuous
isotone and lower semi-continuous	$\longrightarrow$	left-continuous
residuated mapping	$\longrightarrow$	left-continuous and non-decreasing function
dual residuated mapping	$\longrightarrow$	right-continuous and non-decreasing function

Lindley equation

$$\overline{W}_n = \max\{0, \overline{W}_{n-1} + S_{n-1} - A_{n-1}\},$$

with

$\overline{W}_n$	queueing time of n-th packet,
S_{n-1}	service time of $(n-1)$-th packet, and
A_{n-1}	is time between arrivals of packets $(n-1)$ and n

- With $W_n = \overline{W}_n + S_n$ we can rewrite Lindley equation as

$$W_n = \max\{0, W_{n-1} - A_{n-1}\} + S_n .$$

- Since $W_n = T_D^p(n) - T_A^p(n)$ we can write

$$\begin{aligned} T_D^p(n) &= W_n + T_A^p(n) \\ &= \max\{T_A^p(n), T_D^p(n-1))\} + \frac{\ell_n}{C} \end{aligned}$$