Tema 5: Servicios y Calidad de Servicio
Cálculo de Redes

November 6, 2025

Contenido I

1. Introducción
2. Álgebra min-plus
   - Convolución
   - Deconvolución
   - Desviación horizontal y vertical
3. Modelado con álgebra min-plus
   - Curvas de llegadas
   - Leaky bucket
   - Curvas de servicio
   - Concatenación
4. Resultados fundamentales
   - Retardo y backlog
   - Cotas retardo/backlog
   - Prioridades
   - Weighted Fair Queuing
   - Ejemplo

Introducción

Introducción

El cálculo de redes [LBT01] modela flujos

- electricidad
- fluidos
- **tráfico en internet**

Nos sirve para modelar:

- conformado de tráfico
- políticas de tráfico
- averiguar métricas de lantencia y tamaño en cola

Álgebra min-plus

Álgebra min-plus: Convolución

Definición (Convolución min-plus)

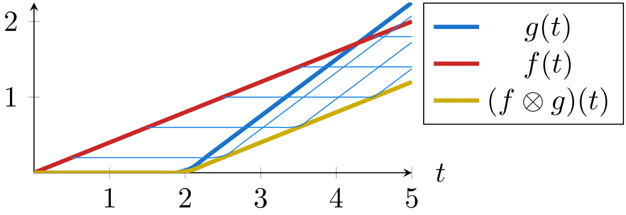
La convolución min-plus $\otimes$ *de dos funciones crecientes* $f, g \in \mathcal{F}$ *se define como*

$$(f \otimes g)(t) = \inf_{0 \leq s \leq t} \{ f(t-s) + g(s) \} \tag{1}$$

Nota: equivalente a la convolución "clásica":

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

Si $f(0) = g(0) = 0$ se puede calcular comenzando a dibujar cada función sobre todo punto de la otra y tomando el mínimo.



En este ejemplo: $f(t) = rt, g(t) = \beta_{R,T}(t)$ con $R > r > 0$.

Álgebra min-plus: Deconvolución

Definición (Deconvolución min-plus)

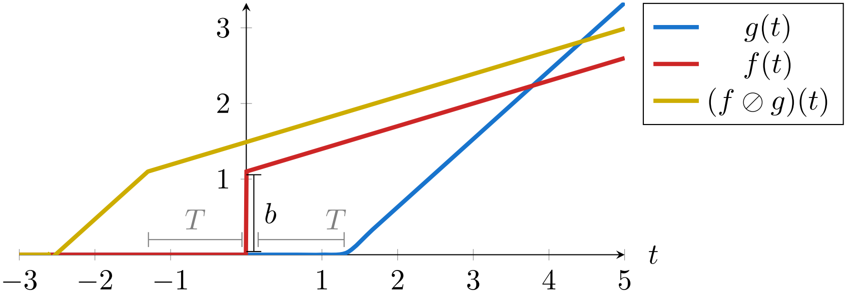
La deconvolución[a] *min-plus* $\oslash$ *de dos funciones crecientes* $f, g \in \mathcal{F}$ *se define como*

$$(f \oslash g)(t) = \sup_{u \geq 0}\{f(t+u) - g(u)\} \tag{2}$$

[a]No esta definida si $\exists t : f(t) = +\infty$ o $g(t) = +\infty$.

Truco: es como la convolución min-plus pero sustituyendo los $+$ por $-$.

La deconvolución de $f(t) = \gamma_{r,b}(t) = rt + b$ y $g(t) = \beta_{R,T}(t)$ es[1]



Nota: nótese que $(f \oslash g)(t) \notin \mathcal{F}$ porque no es cero para $t < 0$.

[1] con $R > r$

Álgebra min-plus: Desviación horizontal y vertical

Definición (Desviación horizontal/Vertical)

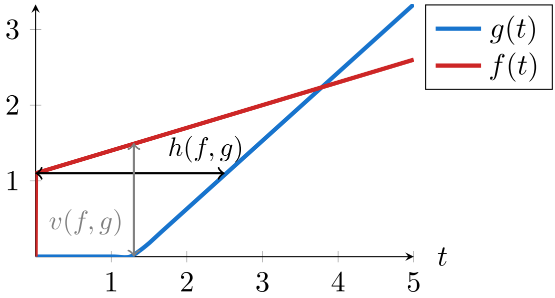
Sean $f, g \in \mathcal{F}$, su desviación horizontal es

$$h(f,g) = \sup_{t \geq 0} \left\{ \inf_{d \geq 0} \{ d : f(t) \leq g(t+d) \} \right\}$$

y su desviación vertical es

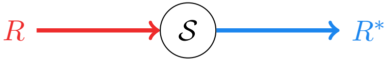
$$v(f,g) = \sup_{t \geq 0} \{ f(t) - g(t) \}$$

Gráficamente las desviaciones verticales y horizontales se ven así:

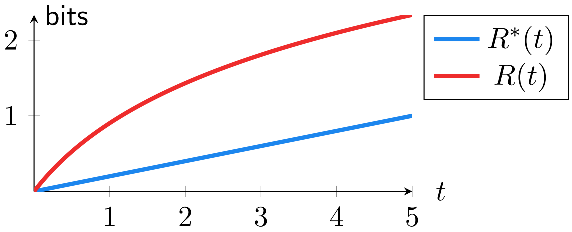


Modelado con álgebra min-plus

Modelado con álgebra min-plus



- $\mathcal{S}$: el sistema atravesado (shaper,router,colas,etc.)
- $R \in \mathcal{F}$: flujo entrante
- $R^* \in \mathcal{F}$: flujo saliente



Modelado con álgebra min-plus: Curvas de llegadas

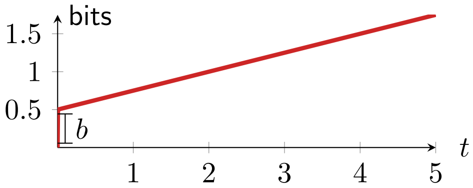
Definición (Curva de llegadas)

La función $\alpha \in \mathcal{F}$ es una curva de llegadas para el flujo $R \in \mathcal{F}$ si y solo si

$$\forall s \leq t, \quad R(t) - R(s) \leq \alpha(t - s)$$

También diremos que le flujo R es α-suave.

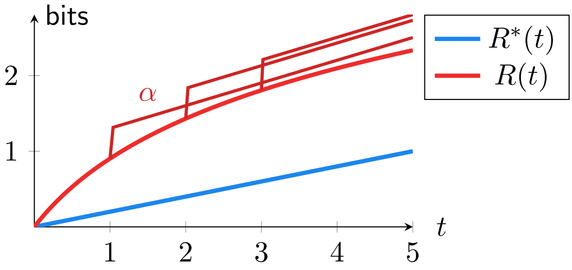
Ejemplo: una curva de llegadas famosa es la función afín $\alpha(t) = rt + b$, también conocida como la curva "rate-burst" $\gamma_{r,b}(t)$.



Modelado con álgebra min-plus: Curvas de llegadas

También, el flujo R es α-suave si $R \leq R \otimes \alpha$.

Interpretación: R queda por debajo de $\alpha = \gamma_{r,b}$ puesta encima.

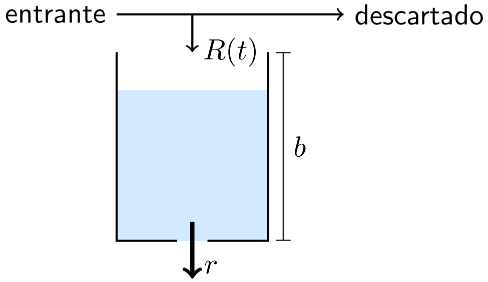


Modelado con álgebra min-plus

El tráfico conformado y no descartado por un leaky bucket $R(t)$ es α-suave con

$$\alpha(t) = rt + b$$

con r la tasa de pérdidas de la cubeta y b el tamaño de cubeta.

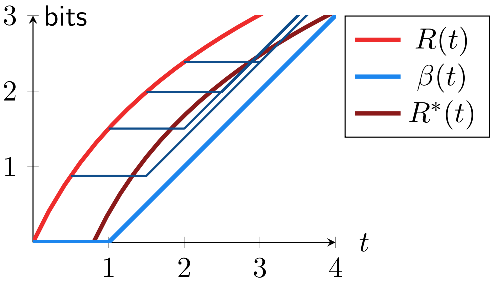


Modelado con álgebra min-plus: Curvas de servicio

Definition (Curva de servicio)

Sea un flujo R atravesando un sistema $\mathcal{S}$, decimos que $\beta \in \mathcal{F}$ es una curva de servicio de $R \in \mathcal{F}$ si y solo si

$$R^* \geq R \otimes \beta$$

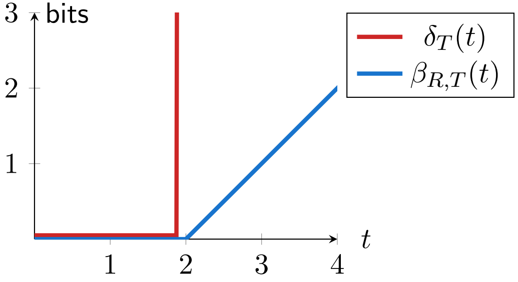


Modelado con álgebra min-plus: Curvas de servicio

La curva de servicio "rate-latency" es $\beta_{R,T}(t) = R[t - T]^{+}$.

La curva de servicio "burst-delay" es

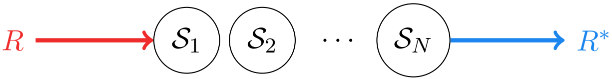
$$\delta_T(t) = \begin{cases} +\infty, & t > T \\ 0, & t \leq T \end{cases}$$



Modelado con álgebra min-plus: Concatenación

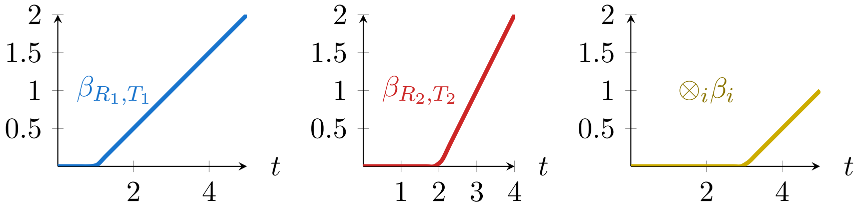
Si atravesamos N sistemas, cada uno con una curva de servicio $\beta_i, i = 1, \ldots, N$; la curva de servicio del sistema total es

$$\beta = \beta_1 \otimes \beta_2 \ldots \otimes \beta_N = \otimes_{i=1}^{N} \beta_i$$



<u>Ejemplo</u>: concatenación de curvas rate-latency.

$$(\beta_{R_1,T_1} \otimes \beta_{R_2,T_2}) = \beta_{\min\{R_1,R_2\},T_1+T_2}$$



Resultados fundamentales

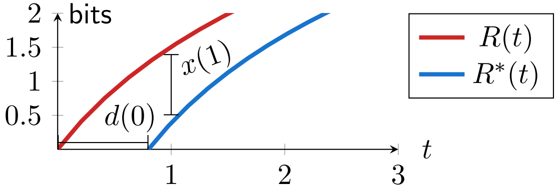
Definición (Retardo y backlog)

Sea un flujo $R \in \mathcal{F}$ que al atravesar el sistema $\mathcal{S}$ sale como $R^* \in \mathcal{F}$, el retardo de un bit que llega en t se define como

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

y el backlog en el instante t como

$$x(t) = R(t) - R^*(t)$$



Resultados fundamentales: Cotas retardo/backlog

Lema (Cota retardo/backlog)

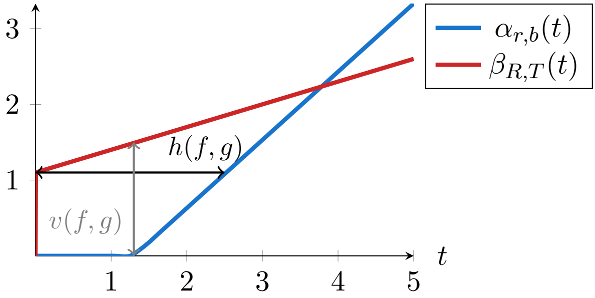
Sea un flujo $R \in \mathcal{F}$ *con curva de llegada* α *que atraviesa un sistema* $\mathcal{S}$ *con curva de servicio* β*, el retardo y backlog cumplen*

$$d(t) \leq h(\alpha, \beta)$$
$$x(t) \leq v(\alpha, \beta)$$

Resultados fundamentales: Cotas retardo/backlog

Ejemplo: si tenemos las curvas $\alpha = \gamma_{r,b}$ y $\beta_{R,T}$, las cotas de retardo y backlog son

$$d(t) \leq \frac{b}{R} + T$$
$$x(t) \leq rT + b$$



Resultados fundamentales: Prioridades

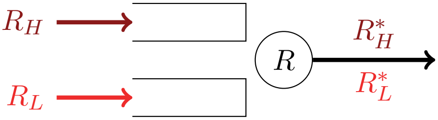
Lema (Curvas de servicio sistema con prioridades)

Sean dos flujos $R_L, R_H \in \mathcal{F}$ *con curvas de llegadas* α_L, α_H *que atraviesan un sistema* $\mathcal{S}$ *con curva de servicio* β *que da prioridad a* R_H*, las curvas de servicio que experimentan los flujos de alta/baja prioridad son*

$$\beta_H = [\beta - l_{\max}^L]^+$$
$$\beta_L = [\beta - \alpha_H]^+$$

con $l_{\max}^L$ *el tamaño máximo de paquete del flujo de baja prioridad.*

Ejemplo: la cola de prioridad luego atraviesa un sistema con curva de servicio rate latency $\beta_{R,T}$. Además el flujo de prioridad R_H es $\gamma_{r,b}$-suave.



En este caso se tiene

$$\beta_H = \beta_{R, \frac{l_{\max}^L}{R}}$$

$$\beta_L = \beta_{R-r, \frac{b}{R-r}}$$

Resultados fundamentales: Weighted Fair Queuing

En el caso del WQF cada flujo i recibe un peso $(w_i/\sum_{j\neq i} w_j)C$. Por tanto el flujo R_i experimenta una curva de servicio

$$\beta_{\frac{w_i}{\sum_{j\neq i} w_j}C,0}$$

Resultados fundamentales: Ejemplo I

Considere un tráfico de baja prioridad que

1. se conforma por un leaky bucket $b = 2\,\mathrm{Mb}, r = \frac{1}{4}\,\mathrm{Mbps}$
2. atraviesa un switch con velocidad de rtansmisión de $R = 2\,\mathrm{Mbps}$ con tráfico de alta prioridad a tasa $r = 1\,\mathrm{Mbps}$ con ráfagas de $b = 2\,\mathrm{Mb}$
3. y sale por un switch on WFQ y velocidad de transmisión de $C = 10\,\mathrm{Mbps}$ que otorga un peso de $\frac{1}{20}$ al tráfico de baja prioridad

¿Cuál es el retardo máximo que sufrirá el tráfico conformado de baja prioridad?

Resultados fundamentales: Ejemplo II

Solución:

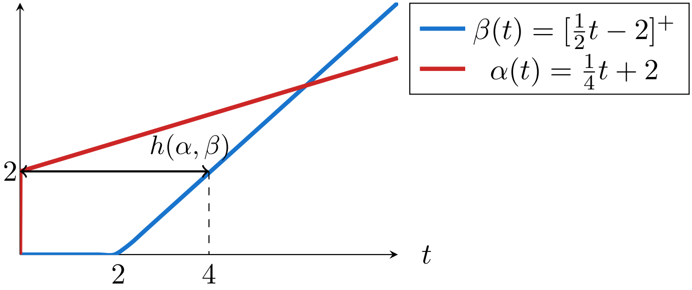
- tráfico conformado $\alpha(t) = \frac{1}{4}t + 2$
- servicio baja prioridad $\beta_1(t) = [(R-r)t - \frac{b}{R-r}]^+ = [t-2]^+ = \beta_{1,2}$
- servicio WFQ $\beta_2(t) = \frac{C}{20}t = [\frac{1}{2}t - 0]^+ = \beta_{\frac{1}{2},0}$

La curva de servicio resultante es

$$\beta = \beta_1 \otimes \beta_2 = \beta_{1,2} \otimes \beta_{\frac{1}{2},0} = \beta_{\frac{1}{2},2}$$

el máximo retardo está acotado por la desviación horizontal de la curva de llegadas del tráfico conformado $\alpha(t)$ con la curva de servicio $\beta(t)$

Resultados fundamentales: Ejemplo III



Por tanto el retardo siempre está acotado por la desviacón horizontal

$$d(t) \leq h(\alpha, \beta) = 4\,\text{sec}$$

Referencias I

Jean-Yves Le Boudec and Patrick Thiran, Network calculus: a theory of deterministic queuing systems for the internet, Springer, 2001.