Q8. Zeitschrankenberechnung für Warteschlangennetze

Gliederung

1. Einfache Schranken für Stationen
2. Schranken für Ankunftsraten
3. Schranken für Bedienraten
4. Mathematische Grundlagen
5. Einzelne Stationen
6. Tandemnetze

Literatur

Neues Gebiet, das bisher noch keinen Eingang in die einschlägigen Lehrbücher gefunden hat!

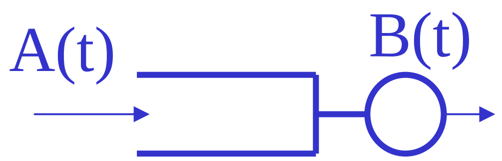
Einige Originalquellen

- J. Y. Le Boudec, P. Thiran
Network Calculus, Springer 2001 (Auszüge aus Teil 1)
- C. S. Chang. Performance Guarantees in Communication Networks. Springer 2000.
- F. Baccelli, G. Cohen, G. J. Olsder, J. P. Quadrat
Synchronization and Linearity, Wiley 1992 (Auszüge aus Teil I)
- L. Thiele, S. Chakraborty
Real-Time Calculus for Scheduling Hard Real-Time Systems
Proc. ISCAS 2000

Ziele:

- Idee der Schrankenberechnung in Warteschlangen kennen lernen
- Ankunfts- und Bedienzeitkurven als Basiskonzepte der Schrankenberechnung einordnen und definieren können
- Basiskonzepte des Network- und Real-Time-Calculus kennen lernen
- Mathematische Grundlagen der Schrankenberechnung kennen lernen
- Analyseansätze für einfache Tandemnetze einsetzen können

8.1 Einfache Schranken für Stationen



Bisherige Sichtweise:

- Zwischenankunftszeiten und Bedienzeiten sind Zufallsvariablen
- Berechnet wird der Erwartungswert der Leistungsgrößen
(in einigen Fällen auch die Verteilung)
- Bei unendlichen Puffern können keine Aussagen über den schlechtesten Fall gemacht werden

$\Rightarrow$Nur sehr eingeschränkt für die Dimensionierung von (harten) Realzeitsystemen geeignet!

Sichtweise in diesem Kapitel

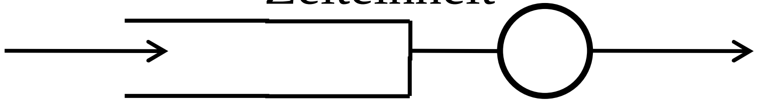
- Schranken für Ankunfts- und Bedienzeiten
  - untere Schranke für die Zwischenankunftszeit
  - obere Schranke für die Bedienzeit

  liefern Schranken für die Verweilzeit, Pufferbelegung …
- Systematische Analysemethode durch Annahme kontinuierlicher Ströme

Sichtweise

Ankunftsprozess A(t):
N,N,N,N,….
generiert Last
(jedes N ist eine Lasteinheit)

Bedienrate
(erst einmal) konstant c
Lasteinheiten pro
Zeiteinheit



Abgangsprozess A‘(t):
N‘,N‘,N‘,….
beschreibt die Last für
nachfolgende Stationen

Beschreibung des Ankunftsprozesses im Intervall [0, T]:

- A(t) Summe der Lasteinheiten, die im Intervall [0,t] angekommen sind

t kann diskret oder kontinuierlich sein

A(t) sei rechts- oder links-kontinuierlich, so dass $dA/dt = a(t)$ für alle t ∈ [0,T]

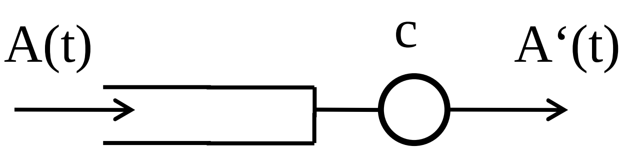
existiert und $A(t) = \int_0^t a(s)ds$

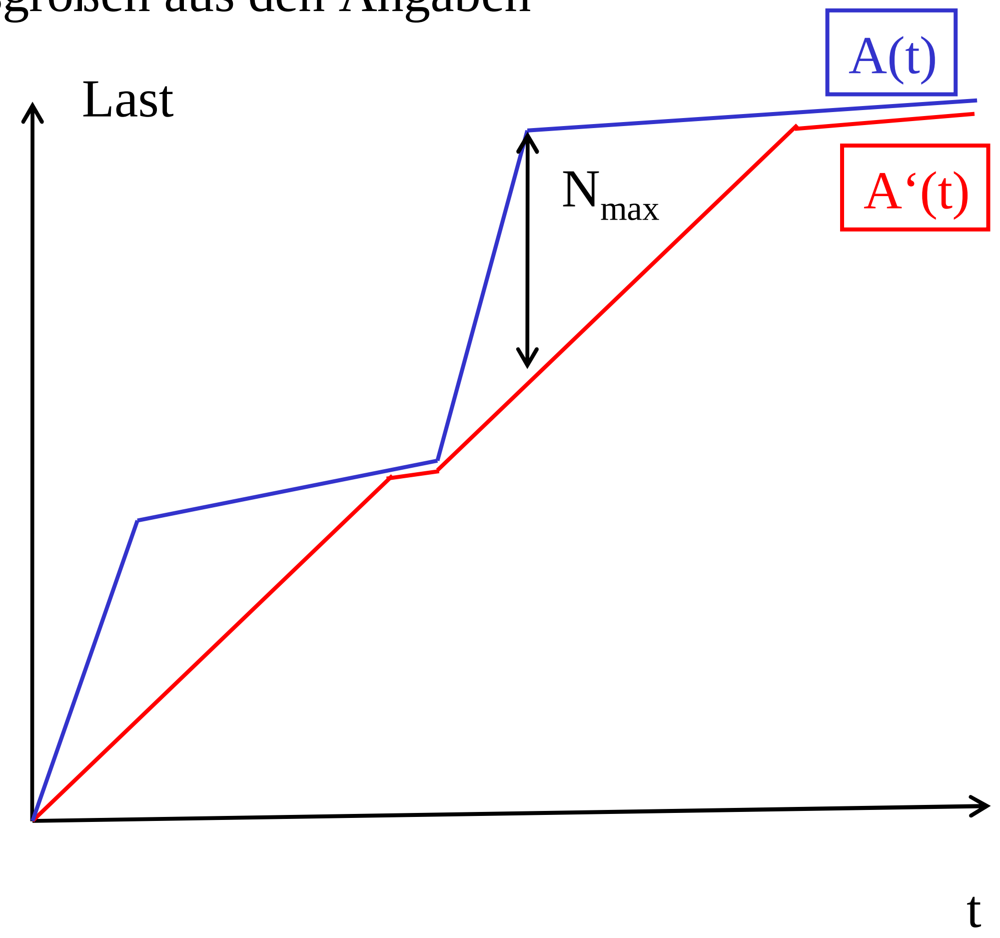
- Jeder kontinuierliche Strom kann diskretisiert werden

sei Δ>0 der Zeittakt, dann ist D(kΔ) = A(kΔ) mit k = 0,1,2, … und

A(t) = D(⌈t/Δ⌉) ist das kontinuierliche Modell, dass aus dem diskreten resultiert

Ziel Berechnung von Leistungsgrößen aus den Angaben





Annahme: Keine Verluste!

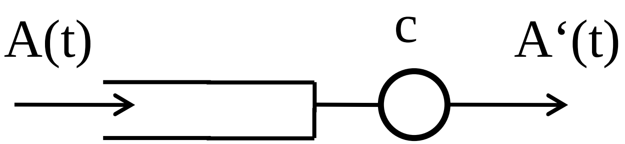
Population zum Zeitpunkt t:

$N(t) = A(t) – A‘(t)$

$= \max_{s \leq t}(A(t) – A(s) – c(t-s))$

Maximale Population in [0,T]:

$N_{max} = \max_{T \geq t \geq 0}(A(t)-A‘(t)) = \max_{0 \leq t \leq T}(\max_{s \leq t}(A(t) – A(s) – c(t-s)))$



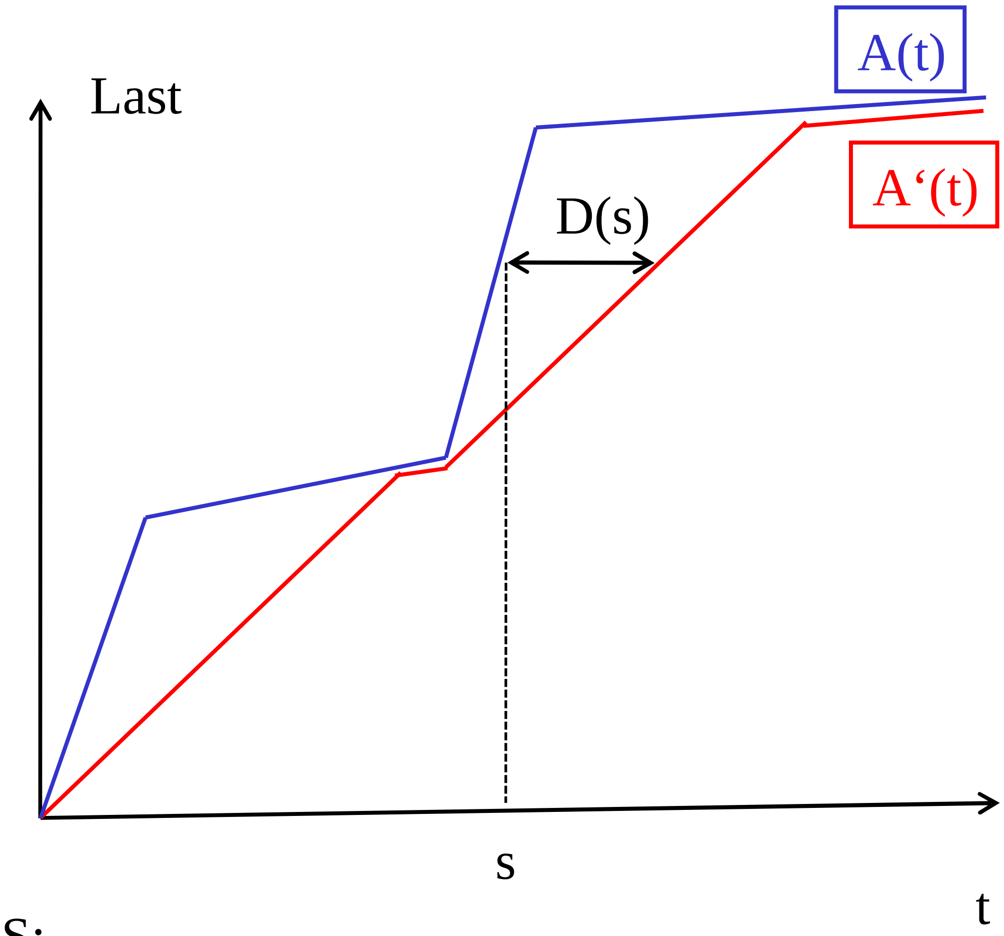
Bestimmung von A‘(t):

$A'(t) = \min_{s \leq t}(A(s) + c(t-s))$

Verzögerung zum Zeitpunkt t,

falls Bediendisziplin FCFS

$D(t) = \min\{s \geq 0 : A(t) \leq A'(t+s)\}$



Maximale Verzögerung für FCFS:

$D_{max} = \max_{T \geq t \geq 0}(\min\{s \geq 0 : A(t) \leq A'(t+s)\})$

8.2 Schranken für Ankunftsraten

Ankunftsströme sind oft nicht deterministisch, sondern variieren

Schranken können nur berechnet werden, wenn Schranken für Ankunftsraten und Bedienraten bekannt!

In diesem Kapitel werden Schranken für Ankunftsprozesse bestimmt:

Sei $\alpha: \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}_{\geq 0}$ eine streng monoton nicht fallende Funktion

- α^u ist eine obere Schranke für einen Ankunftsprozess, falls gilt:

$\forall t \geq s \geq 0: A(t) - A(s) \leq \alpha^u(t-s)$ (upper arrival curve)

- α^l ist eine untere Schranke für einen Ankunftsprozess, falls gilt:

$\forall t \geq s \geq 0: A(t) - A(s) \geq \alpha^l(t-s)$ (lower arrival curve)

- α ist subadditiv, falls $\alpha(t+s) \leq \alpha(t) + \alpha(s)$ (α^u soll subadditiv sein)

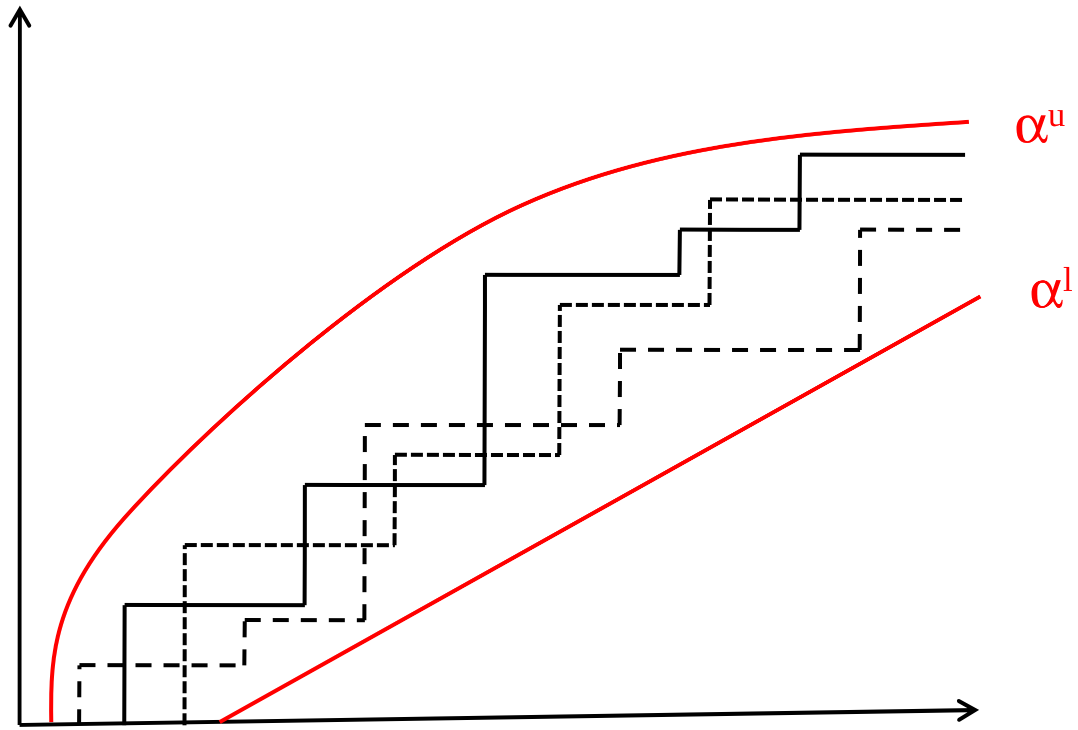
- α ist superadditiv, falls $\alpha(t+s) \geq \alpha(t) + \alpha(s)$ (α^l soll superadditiv sein)

Bestimmung von Ankunftskurven:

- Aus Messungen
  - An realen Systemen
  - Simulations-Traces
- Aus stochastischen Beschreibungen
- Aus dem System
  - Flusskontrollmechanismen (Leaky bucket)
  - Systembeschränkungen (Hardwaretakt)
- Aus Modellen
  - Automatenmodellen zu Spezifikation
  - UML Spezifikationen

Beobachtungen aus Traces (Messung oder Simulation)

Üblicherweise Ankunft von Lastmengen zu diskreten Zeitpunkten

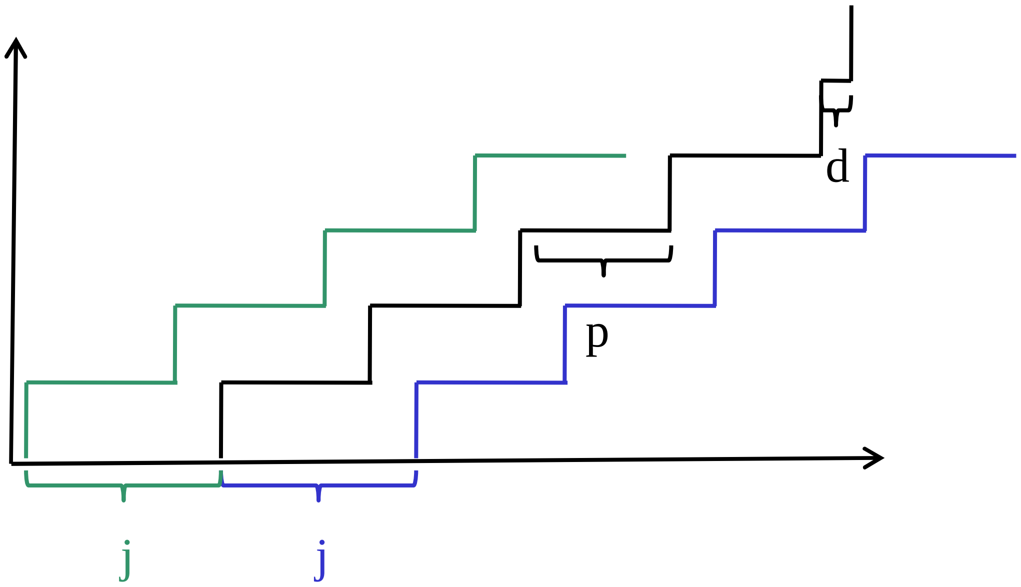


- Genügend große Zahl von Abläufen ist zu beobachten
- Funktionen zur Beschreibung der Ankünfte sollten
  - eine kompakte Darstellung haben,
  - subadditiv (bzw. superadditiv) sein
- Analyse durch stochastische Auswertung der Messungen

Ableitung aus stochastischen Beschreibungen

Definition eines Ankunftsstroms

- Ankunftsabstände (p, j, d) mit
  - p Periode (mittlerer Abstand zwischen zwei Paketen)
  - j Jitter (maximale Abweichung von der Periode)
  - d minimaler Abstand zwischen zwei Paketen
- Paketgröße konstant c=1 (eine Lasteinheit)



$$\alpha^l(s) = \left\lfloor \frac{s-j}{p} \right\rfloor$$

$$\alpha^u(s) = \min\left\{ \left\lceil \frac{s+j}{p} \right\rceil, \left\lceil \frac{s}{d} \right\rceil \right\}$$

Ableitung aus stochastischen Beschreibungen

Falls Abstand zwischen zwei Ankünften durch eine Verteilung mit Verteilungsfunktion F(t) (mit F(0) = 0) und Verteilungsdichte f(t) gegeben und jede Ankunft eine Lasteinheit erzeugt, dann ergeben sich folgende Schranken:

$$\alpha^u(0) = \alpha^l(0) = 0$$

Sei $\alpha^l(t)$ und $\alpha^u(t)$ bekannt, dann:

$$\Delta^l = \sup\{s : f(s) > 0\} \;:\; \alpha^l(t + \Delta^l) \leq \alpha^l(t) + 1$$

$$\Delta^u = \inf\{s : f(s) > 0\} \;:\; \alpha^u(t + \Delta^u) \geq \alpha^u(t) + 1$$

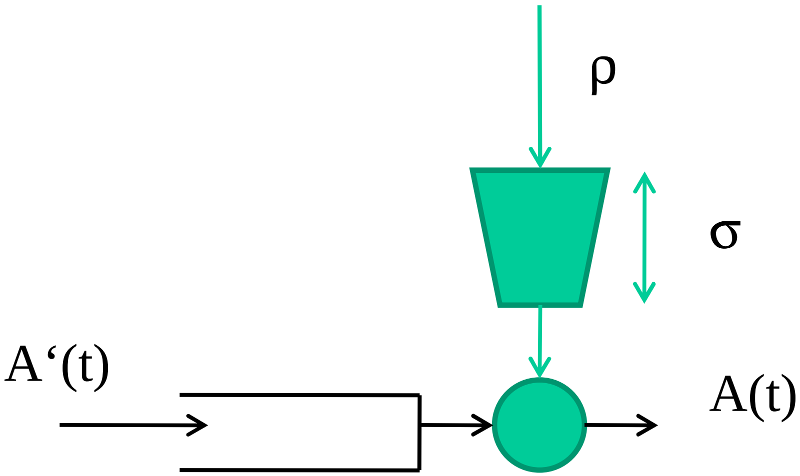
Falls f(0)>0 gilt $\alpha^u(t) = \infty$ für alle t, falls $f(\infty)>0$ gilt $\alpha^l(t)=0$ für alle t!

Probabilistische Schranken mit Irrtumswahrscheinlichkeit ε erhält man zum Beispiel durch die folgende Berechnung von Δ^l und Δ^u:

$\Delta^l = \sup\{s: F(s) > 1 - \varepsilon\}$ und $\Delta^u = \inf\{s: F(s) > \varepsilon \wedge f(s) > 0\}$

Ableitung aus dem System

Beispiel leaky bucket (lecker Eimer)



Obere Schranke für A(t):

$A(t) \leq \rho t + \sigma$

Untere Schranke für A(t) =
min (untere Schranke A‘(t),
$\rho t + \sigma$)

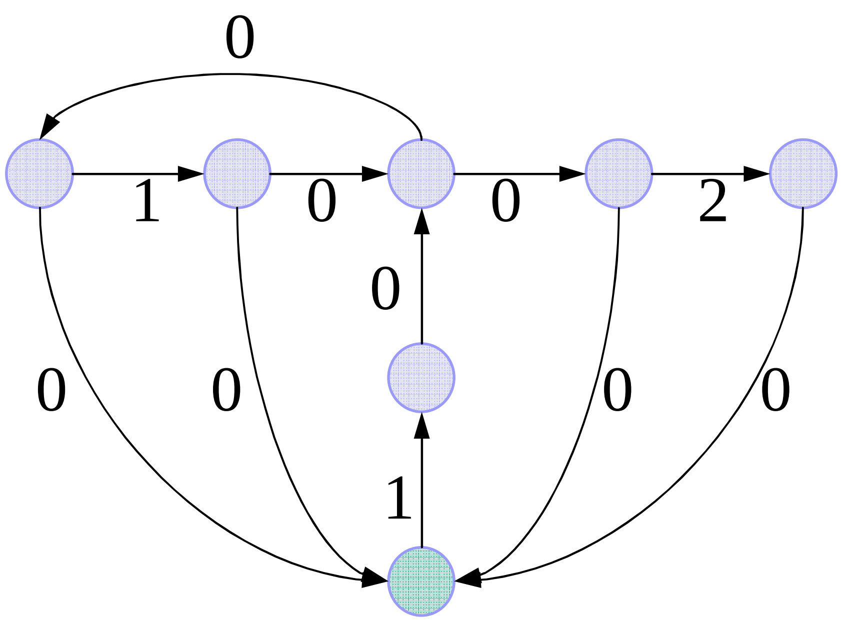
Weit verbreitetes Modell für Ankunftsströme!

Ableitung aus Modellen

Hier endliche Automaten zur Spezifikation von Programmen, Komponenten etc.

Automaten mit Kantengewichten zur Beschreibung der generierten Lasteinheiten

Periodisches Verhalten, d.h. eine Transition pro Zeiteinheit



A(t) ($t \in \mathbb{N}$):

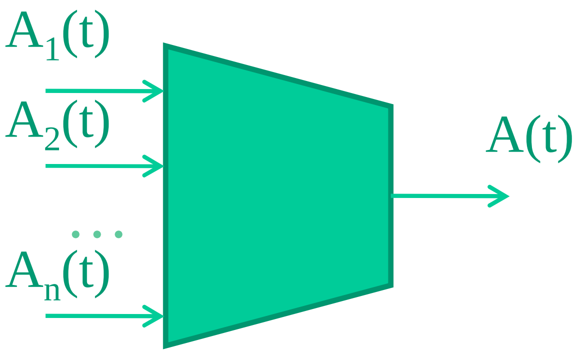
Pfadgewicht eines Pfades der Länge t, der entweder

- im grünen Zustand oder
- in einem beliebigen Zustand

beginnt

Schranken durch Berechnung minimaler/maximaler Pfadgewichte

Überlagerung (Multiplexing) von Ankunftsströmen



Sei $\alpha_i^l(t) \leq A_i(t) \leq \alpha_i^u(t)$

$$\Rightarrow \alpha^l(t) = \sum_{i=1}^{n} \alpha_i^l(t) \leq A(t) \leq \alpha^u(t) = \sum_{i=1}^{n} \alpha_i^u(t)$$

Beispiel: $Ai(t) \leq \alpha_i^u(t) = \rho_i t + \sigma_i$

$$\Rightarrow A(t) \leq \left(\sum_{i=1}^{n} \rho_i\right) \cdot t + \sum_{i=1}^{n} \sigma_i$$

Es gilt dann natürlich auch:

$$\Rightarrow \alpha^u(t) - \alpha^l(t) = \sum_{i=1}^{n} \alpha_i^u(t) - \sum_{i=1}^{n} \alpha_i^l(t)$$

Überlagerte Schranken, die aus stochastischen Überlegungen resultieren, erlauben in der Regel die Bestimmung engerer Schranken für den überlagerten Prozess!

8.3 Schranken für Bedienraten

Vorgehen ähnlich zur Schrankenberechnung für Ankunftsprozesse:

- Sei B(t) die Lastmenge, die im Intervall [0,t] bearbeitet/bedient würde, wenn der Bediener permanent arbeitet

Sei β: $\mathbb{R}_{\geq 0} \rightarrow \mathbb{R}_{\geq 0}$ eine streng monoton nicht fallende Funktion

- β^u ist eine obere Schranke für einen Bedienprozessprozess, falls gilt:

$\forall t \geq s \geq 0$: $B(t) - B(s) \leq \beta^u(t-s)$ (upper service curve)

- β^l ist eine untere Schranke für einen Bedienprozess, falls gilt:

$\forall t \geq s \geq 0$: $B(t) - B(s) \geq \beta^l(t-s)$ (lower service curve)

Beispiele:

- Konstante Bedienrate c: $\beta^l(t) = \beta^u(t) = ct$
- Konstante Bedienrate c und maximale Verzögerung d:

  $\beta^l(t) = \max(0, c(t-d))$ und $\beta^u(t) = ct$
- Konstante Bedienrate c, konstante Verzögerung d und Jitter j ($j < d$):

  $\beta^l(t) = \max(0, c(t-d-j))$ und $\beta^u(t) = \max(0, c(t-d+j))$
- TDMA, es werden alle p Zeiteinheiten bis zu c_Δ Lasteinheiten bedient:

  $\beta^l(t) = \lfloor t/p \rfloor \cdot c_\Delta$ und $\beta^u(t) = \lceil t/p \rceil \cdot c_\Delta$
- Stochastische Bedienzeiten pro Lasteinheit mit Verteilungsfunktion F(s),

  $F(s) = 0$ für $0 < s < \Delta^l$ und $F(s) = 1$ für $s < \Delta^u < \infty$:

  $\beta^l(t) = \lfloor t/\Delta^u \rfloor$ und $\beta^u(t) = \lfloor t/\Delta^l \rfloor$
- Stochastische Bedienzeiten pro Lasteinheit mit Verteilungsfunktion F(s) unter Einbeziehung einer Irrtumswahrscheinlichkeit ε :

  $\Delta^l = \sup\{s: F(s) > 1 - \varepsilon\}$ und $\Delta^u = \inf\{s: F(s) > \varepsilon \wedge f(s) > 0\}$

  $\beta^l(t) = \lfloor t/\Delta^u \rfloor$ und $\beta^u(t) = \lfloor t/\Delta^l \rfloor$

8.4 Mathematische Grundlagen

Bevor wir uns mit der Analyse von Systemen beschäftigen, ein kurzer Überblick über die mathematischen Grundlagen: max/+- und min/+ Algebra (weitere Details in der Literatur z.B. Buch von Baccelli et al.)

Mathematische Struktur der Seminring: ($\mathbb{K}$, +, ·,0,1) mit

- $\mathbb{K}$ ist eine Menge mit den binären Operationen + und ·, die folgenden Axiome erfüllen:
- + und · are assoziativ
- + ist kommutativ
- + und · sind rechts- und links-distributiv
- 0 und 1 sind die additiven und multiplikativen Identitäten mit $0 \neq 1$
- für alle $k \in \mathbb{K}$ gilt $k \cdot 0 = 0 \cdot k = 0$.

Übliches Beispiel

- $(\mathbb{R}_{\geq 0}, +, \cdot, 0, 1)$:

$a \cdot (b + c) = a \cdot b + a \cdot c$

Wir nutzen im Folgenden:

- Min/+: $(\mathbb{R}_{\geq 0} \cup \{+\infty\}, \wedge, +, +\infty, 0)$ mit $\wedge$ = min bzw. inf:

$a + (b \wedge c) = a + b \wedge a + c = \min(a+b, a+c)$

- Max/+: $(\mathbb{R}_{\geq 0} \cup \{-\infty\}, \vee, +, -\infty, 0)$ mit $\vee$ = max bzw. sup:

$a + (b \vee c) = a + b \vee a + c = \max(a+b, a+c)$

Übertragung auf Funktionen $f,g: \mathbb{R} \rightarrow \mathbb{R}$:

$$\left.\begin{array}{l} (f+g)(t) = f(t) + g(t) \\ (f \vee g)(t) = f(t) \vee g(t) \\ (f \wedge g)(t) = f(t) \wedge g(t) \end{array}\right\} \text{für alle } t$$

Notationen

$f \leq g \Rightarrow f(t) \leq g(t)$ für alle t,
entsprechend für =, <, >, $\geq$

Integral einer Funktion $\int_0^t f(s)ds$ wird zu $\inf_{0 \le s \le t}(f(s))$ bzw. $\sup_{0 \le s \le t}(f(s))$

Sei $\mathcal{F}$ die Menge der monoton nicht fallenden Funktionen mit f(t)=0 für t<0

Faltung von Funktionen (seien f, g $\in \mathcal{F}$):

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

Faltung von Funktionen in min/+ (seien f, g $\in \mathcal{F}$):

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

Faltung von Funktionen in max/+ (seien f, g $\in \mathcal{F}$):

$$(f \overline{\otimes} g)(t) = \sup_{0 \le s \le t} (f(t-s) + g(s))$$

Subadditivität: $f(s+t) \leq f(s) + f(t)$

Subadditiver Abschluss:

Sei $f \in \mathcal{F}$, dann ist $\bar{f}(t)$ rekursiv definiert als

1. $\bar{f}(0) = 0$
2. $\bar{f}(t) = \min_{0 \leq s \leq t}(f(t), \bar{f}(s) + \bar{f}(t-s))$ für $t > 0$

Eigenschaften von $\bar{f}, \bar{g}$ für $f,g \in \mathcal{F}$:

- $\bar{f}$ ist subadditiv
- $\bar{f} \leq f$
- Falls f subadditiv, dann ist $\bar{f} = f$, falls $f(0)=0$ gilt auch die Umkehrung
- Monotonie $f \leq g \Rightarrow \bar{f} \leq \bar{g}$
- Subadditiver Abschluss des Minimums $\overline{f \wedge g} = \bar{f} \underline{\otimes} \bar{g}$

Superadditivität: $f(s+t) \geq f(s) + f(t)$

Superadditiver Abschluss:

Sei $f \in \mathcal{F}$, dann ist $\underline{f}(t)$ rekursiv definiert als

1. $\underline{f}(0) = 0$
2. $\underline{f}(t) = \max_{0 \leq s \leq t}(f(t), \underline{f}(s) + \underline{f}(t\text{-}s))$ für $t > 0$

Eigenschaften von $\underline{f}, \underline{g}$ für $f, g \in \mathcal{F}$:

- $\underline{f}$ ist superadditiv
- $\underline{f} \geq f$
- Falls f superadditiv, dann ist $\underline{f} = f$, falls $f(0)=0$ gilt auch die Umkehrung
- Monotonie $f \leq g \Rightarrow \underline{f} \leq \underline{g}$
- Subadditiver Abschluss des Minimums $\underline{f \wedge g} = \underline{f} \otimes \underline{g}$

- α^u ist eine obere Schranke für A $\Leftrightarrow A \leq A \underline{\otimes} \alpha^u$
- α^l ist eine unter Schranke für A $\Leftrightarrow A \geq A \overline{\otimes} \alpha^l$
- β^u ist eine obere Schranke für B $\Leftrightarrow B \leq B \underline{\otimes} \beta^u$
- β^l ist eine unter Schranke für B $\Leftrightarrow B \geq B \overline{\otimes} \beta^l$

Subadditive und superadditive Schranken können durch Bestimmung des subadditiven und superadditiven Abschlusses erreicht werden

Entfaltung von Funktionen in min/+ (seien f, g $\in \mathcal{F}$):

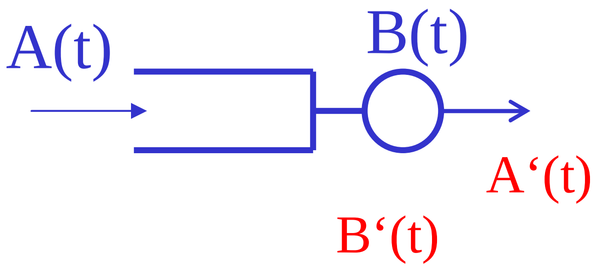
$$(f \underline{\oslash} g)(t) = \sup_{0 \leq s} \left(f(t+s) - g(s) \right)$$

Entfaltung von Funktionen in max/+ (seien f, g $\in \mathcal{F}$):

$$(f \overline{\oslash} g)(t) = \inf_{0 \leq s} \left(f(t+s) - g(s) \right)$$

Entfaltung zweier Funktionen aus $\mathcal{F}$ muss nicht in $\mathcal{F}$ liegen

8.5 Einzelne Stationen



Schranken α^l, α^u, β^l und β^u bekannt

Unendlicher Puffer

FCFS Bedienung

Bediener arbeitet, wenn Last vorhanden

Ziele: Bestimmung von

- Schranken für Abgangsprozess $\alpha'^l \leq A' \leq \alpha'^u$

$$\alpha'^l = \inf\left\{(\alpha^l \underline{\oslash}\, \beta^u) \underline{\otimes}\, \beta^l, \beta^l\right\},\ \alpha'^u = \inf\left\{(\alpha^u \underline{\otimes}\, \beta^u) \underline{\oslash}\, \beta^l, \beta^u\right\}$$

- Schranken für freie Bedienkapazität $\beta'^l \leq B' \leq \beta'^u$

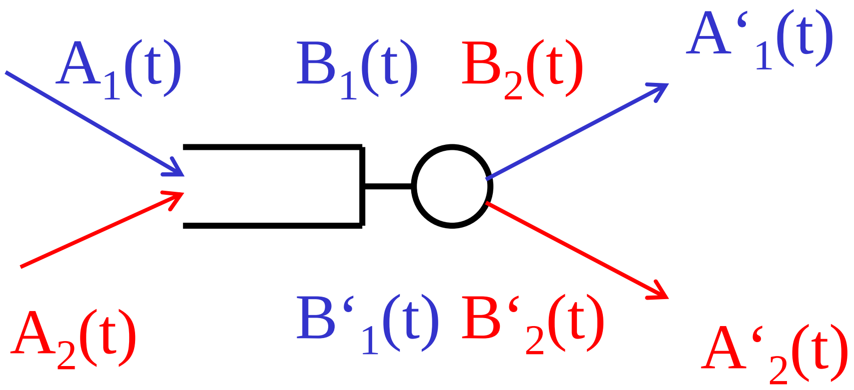
$$\beta'^l = (\beta^l - \alpha^u)\ \overline{\otimes}\ 0,\ \beta'^u = (\beta^u - \alpha^l)\ \overline{\oslash}\ 0$$

- Schranken für Population

$$\sup_{s\geq 0}\left(\alpha^l(s) - \beta^u(s)\right) \leq N_{\max} \leq \sup_{s\geq 0}\left(\alpha^u(s) - \beta^l(s)\right)$$

- Schranken für die Verweilzeit

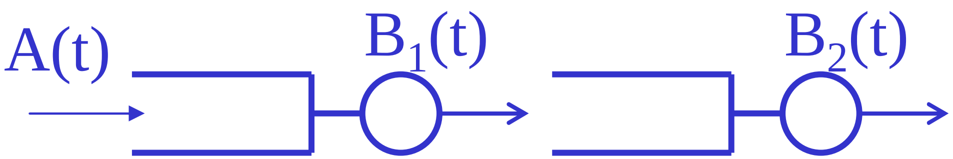
$$\sup_{t\geq 0}\left(\inf\left\{s \geq 0 : \alpha^l(t) \leq \beta^u(t+s)\right\}\right) \leq D_{\max} \leq \sup_{t\geq 0}\left(\inf\left\{s \geq 0 : \alpha^u(t) \leq \beta^l(t+s)\right\}\right)$$



α_i^l, α_i^u , β_i^l, β_i^u sind die Schranken für die individuellen Ankunfts- und Bedienprozesse

- Es gilt $\alpha'^l \leq A'_1(t) + A'_2(t) \leq \alpha'^u$ und $\beta'^l \leq \max(B'_1(t) - B'_2(t), 0) \leq \beta'^u$ für System mit Ankunftskurve $[\alpha^l = \alpha_1^l+\alpha_2^l , \alpha^u = \alpha_1^u+\alpha_2^u]$ und Bedienkurve $[\beta^l = \beta_1^l = \beta_2^l , \beta^u = \beta_1^u = \beta_2^u]$
  Berechnung wie vorher
- Unterbrechende Prioritäten (1 vor 2):
  - Analysiere System für Klasse 1 $\Rightarrow$ $\alpha_1'^l , \alpha_1'^u$ sowie
  - Analysiere Systeme für Klasse 2 mit neuen Schranken für $B_2(t)$ $[\max(\beta_2^l - \beta'^u_1, 0), \max(\beta_2^u - \beta'^l_1 ,0)]$
- Generalized Processor Sharing (GPS) mit Aufteilung $(p_1, p_2 = 1-p_1)$ durch Analyse von zwei Systemen mit $[\beta_i^l , \beta_i^u]$ und $[p_i \alpha_i^l , p_i \alpha_i^u]$

8.4 Tandemnetze



- Berechnung der ersten Station liefert $A_2(t)$ =A'(t)

  $\Rightarrow$ Berechnung der zweiten Station mit Eingangsprozess $A_2(t)$

Beispiel:

- Ankunftsprozess nach oben beschränkt durch $\alpha_1^u(t) = \sigma + \rho t$
- Bedienprozesse nach unten beschränkt durch $\beta_1^l(t) = \min(0, c_i(t-d_i))$

sei $c_i > \rho$

Bestimmung Abgangsprozess Station 1: $\alpha'^{u}_{1}(t) = \sigma + \rho(t+d_1)$
= Ankunftsprozess Station 2

Obere Schranke für die Verweilzeit in Station 1 (FCFS): $D_1^u = d_1 + \sigma / c_1$

Maximale Population in Station 1: $N_1^u = \sigma + \rho d_1$

Analyse von Station 2 mit Ankunftsprozessschranke: $\alpha_1^u(t) = \sigma + \rho(t+d_1)$
analog

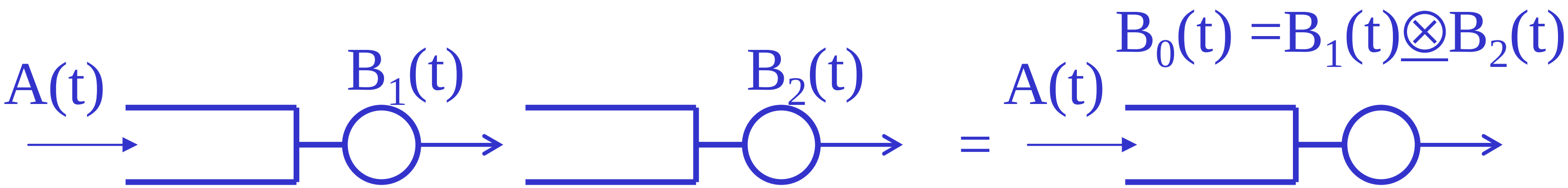
Bestimmung Abgangsprozess Station 1: $\alpha'^{u}_{2}(t) = \sigma + \rho(t+d_1+d_2)$

Obere Schranke für die Verweilzeit in Station 2 (FCFS): $D_2^u = d_2 + (\sigma + \rho d_1) / c_2$

Ober Schranke für die Population in Station 2: $N_1^u = \sigma + \rho(d_1 + d_2)$

Schranke für die Gesamtverzögerung:
$D_1^u + D_2^u = \quad d_1 + d_2 + \sigma / c_1 + (\sigma + \rho d_1) / c_2$

Unter FCFCS gilt:



$\beta_0^l(t) = \min(0, \min(c_1, c_2)\ (t-d_1-d_2))$

Analyse liefert:

Bestimmung Abgangsprozess Station 0: $\alpha'^u_0(t) = \sigma + \rho(t+d_1+d_2)$

Obere Schranke für die Verweilzeit in Station 0 (FCFS):

$$D_0^u = d_1 + d_2 + \sigma / \min(c_1, c_2)$$

Ober Schranke für die Population in Station 1: $N_0^u = \sigma + \rho(d_1 + d_2)$

Offensichtlich gilt $D_0 < D_1+D_2$

Gründe Bursts treten nur einmal auf, werden bei der Zerlegung aber mehrfach berücksichtigt bzw. verstärkt (pay bursts only once)