# D.5 Prioritätssysteme

- Aufträge werden Prioritäten $r$ ($r = 1, 2, \ldots , R$) zugeordnet
- $1$ ist die niedrigste Priorität und $R$ die höchste !
- Der nächste zu bedienende Auftrag ist derjenige mit der höchsten Prioritätszahl $r$
- Haben mehrere Aufträge diesselbe Priorität (gehören derselben Prioritätsklasse an) so erfolgt innerhalb der Prioritätsklasse die Auswahl nach FCFS

# 1 Prioritätssysteme ohne Verdrängung

- Die **mittlere Wartezeit** eines Auftrags der Prioritätsklasse $r$ $\overline{W}_r$ besteht aus drei Komponenten:

  - Die mittlere Restbedienzeit $\overline{W}_0$ des Auftrags in der Bedieneinheit

  - Die mittlere Bedienzeit der Aufträge in der WS, die vor dem betrachteten Auftrag bedient werden. Das sind Aufträge mit derselben und höherer Priorität als der betrachtete Auftrag.

  - Die mittlere Bedienzeit der Aufträge, die während der Wartezeit des betrachteten Auftrags ankommen und vor ihm bedient werden. Das sind Aufträge mit höherer Priotität als der betrachtete Auftrag.

- Definition:
  - $\overline{N}_{ir}$: Mittlere Anzahl von Aufträgen der Priorität $i$, die in der WS sind wenn der betrachtete Auftrag der Priorität $r$ ankommt und die vor ihm bedient werden
  - $\overline{M}_{ir}$: Mittlere Anzahl von Aufträgen der Priorität $i$, die während der Wartezeit des betrachteten Auftrags der Priorität $r$ ankommen und vor ihm bedient werden

- Dann erhält man für die mittlere Wartezeit eines Auftrags der Priorität $r$:

$$\overline{W}_r = \overline{W}_0 + \sum_{i=1}^{R} \overline{N}_{ir} \cdot \frac{1}{\mu_i} + \sum_{i=1}^{R} \overline{M}_{ir} \cdot \frac{1}{\mu_i}$$

Und für Multiple-Server-Systeme (m > 1):

$$\overline{W}_r = \overline{W}_0 + \sum_{i=1}^{R} \frac{\overline{N}_{ir}}{m} \frac{1}{\mu_i} + \sum_{i=1}^{R} \frac{\overline{M}_{ir}}{m} \cdot \frac{1}{\mu_i}$$

◆ Für $\overline{N}_{ir}$ und $\overline{M}_{ir}$ erhält man:

$$\overline{N}_{ir} = 0 \quad i < r$$

$$\overline{M}_{ir} = 0 \quad i \le r$$

Und mit Little's Gesetz:

$$\overline{N}_{ir} = \lambda_i \overline{W}_i \quad i \ge r$$

$$\overline{M}_{ir} = \lambda_i \overline{W}_r \quad i > r.$$

- Damit ergibt sich für die mittlere Wartezeit eines Auftrages der Priorität $r$:

$$\overline{W}_r = \frac{\overline{W}_0}{(1-\sigma_r)(1-\sigma_{r+1})}$$

mit:

$$\sigma_r = \sum_{i=r}^{R} \rho_i$$

- Mittlere Wartezeit über alle Prioritätsklassen:

$$\overline{W} = \sum_{i=1}^{R} \frac{\lambda_i}{\lambda} \cdot \overline{W}_i$$

◆ Mittlere Restbedienzeit

➤ M/M/1:

$$\overline{W}_{0,\mathrm{M/M/1}} = \sum_{i=1}^{R} \rho_i \frac{1}{\mu_i}$$

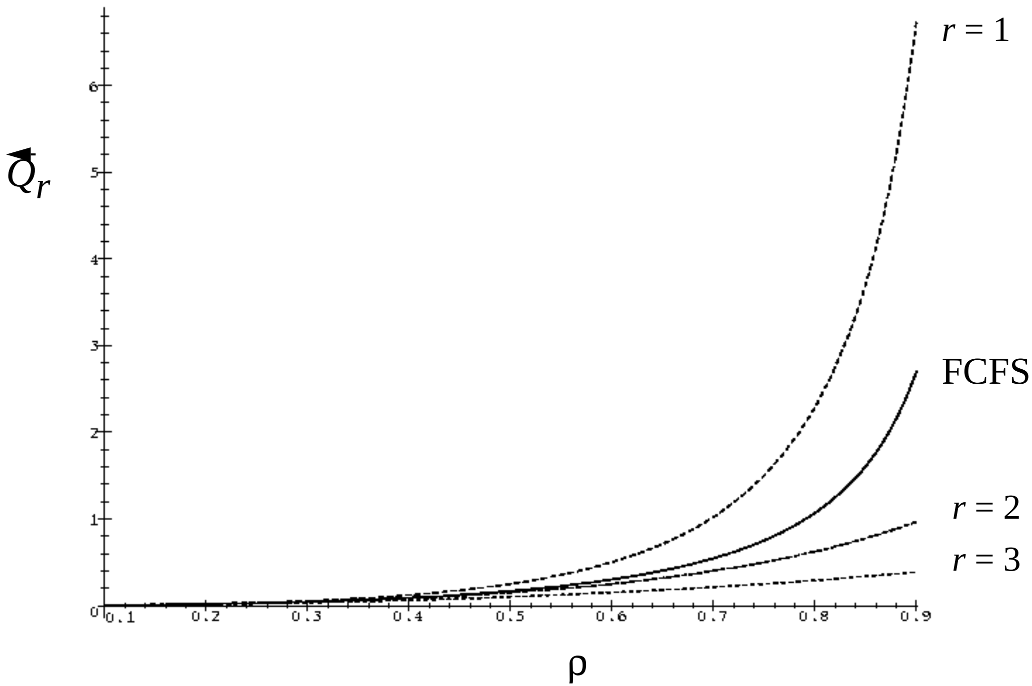
➤ M/G/1:

$$\overline{W}_{0,\mathrm{M/G/1}} = \sum_{i=1}^{R} \rho_i \cdot \frac{1 + c_{B_i}^2}{2\mu_i}$$

➤ M/M/m:

$$\overline{W}_{0,\mathrm{M/M/m}} = \frac{P_m}{m\rho} \sum_{i=1}^{R} \rho_i \cdot \frac{1}{\mu_i}$$

Mittlere Warteschlangenlänge $\overleftarrow{Q}_r$ eines M/M/1-Prioritätssytems ohne Vergrängung:



# 2 Wartezeiterhaltungsgesetz

Ähnlich dem Energie- Impuls- oder Krafterhaltungsgesetz in der Physik gibt es in der Warteschlangentheorie ein Wartezeiterhaltungsgesetz. Das bedeutet, daß ein Auftrag, der z.B. wegen einer höheren Priorität nur eine kurze Wartezeit hat bewirkt, daß andere Aufträge entsprechend länger warten müssen:

➤ Wartezeiterhaltungsgesetz:

$$\frac{1}{\rho}\sum_{i=1}^{R}\rho_i\overline{W}_i = \frac{\overline{W}_0}{1-\rho} = \overline{W}_{\mathrm{FCFS}}$$

◆ Das Wartezeiterhaltungsgesetz gilt unter folgenden Voraussetzungen:

- Keine Bedieneinheit ist leer solange ein Auftrag in der WS ist
- Kein Auftrag verlässt das System ohne fertig bedient bedient zu sein
- Die Bedienzeiten und die Zwischenankunftszeiten sind beliebig verteilt (mit kleinen Einschränkungen)
- Die Bedienzeiten sind unabhängig von WS-Disziplin
- Verdrängung ist nur erlaubt, wenn alle Aufträge dieselbe Exponentialverteilung besitzen und die Verdrängung vom Typ "premptive resume" ist
- Bei G/G/m-Systemen müssen alle Prioritätsklassen dieselbe Bedienzeitverteilung haben (Nicht notwendig bei G/G/1-Systemen)

# 3 Prioritätssysteme mit Verdrängung

- Verdrängung nach der Strategie: "preemptive resume"
- Verdrängung erfolgt zeitlos.
- Auftrag der Priorität $r$ wird nicht beeinflusst durch Aufträge der Prioritäten $1, 2, 3, \ldots, r-1$
- Um die Wartezeit $\overline{W}_r$ zu berechnen, müssen nur die Klassen $r, r+1, \ldots, R$ berücksichtigt werden
- Es kann ersetzt werden:

$$\rho = \sum_{i=1}^{R} \rho_i \quad \text{durch} \quad \sigma_r = \sum_{i=r}^{R} \rho_i$$

und

$$\overline{W}_{\mathrm{FCFS}} \quad \text{durch} \quad \overline{W}^r = \frac{\overline{W}_0^r}{1-\sigma_r}$$

➤ Für die mittlere Restbedienzeit gilt:

$$\overline{W}_0^r = \frac{P_m^r}{2m\sigma_r} \cdot \sum_{i=r}^{R} \rho_i \frac{1+c_B^2}{\mu_i}$$

➤ Zur Bestimmung der mittleren Wartezeit $\overline{W}_r$ eines Auftrages der Priorität $r$ wird das Wartezeiterhaltungsgesetz zweimal angewendet:

$$\sigma_r \cdot \overline{W}^r = \sum_{i=r}^{R} \rho_i \overline{W}_i \, ,$$

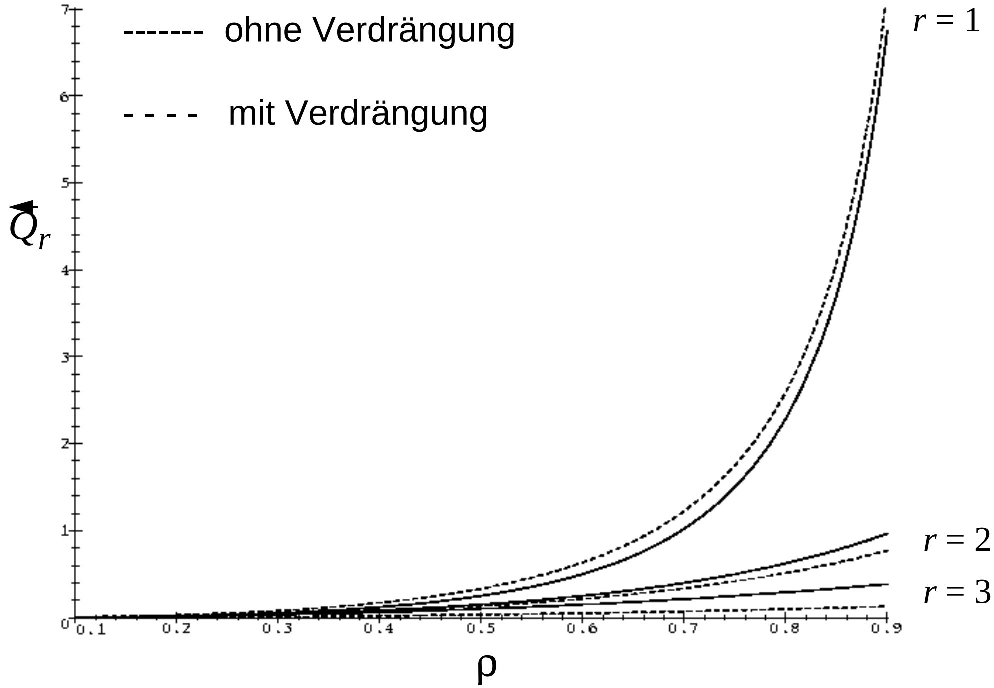
$$\sigma_{r+1} \cdot \overline{W}^{r+1} = \sum_{i=r+1}^{R} \rho_i \cdot \overline{W}_i$$

- Durch Subtraktion ergibt sich für die mittlere Wartezeit eines Auftrags der Priorität $r$:

$$\overline{W}_r = \frac{1}{\rho_r}\left(\sigma_r \overline{W}^r - \sigma_{r+1}\overline{W}^{r+1}\right)$$

- Das Ergebnis ist exakt für:
  - M/M/1
  - M/G/1
  - M/M/m
- Für andere Systeme ist es eine gute Approximation

# Mittlere Warteschlangenlänge für ein M/M/1-Prioritätssystem mit und ohne Verdrängung



# 4 Zeitabhängige Prioritäten

- In vielen Systemen ändern sich die Prioritäten in Abhängigkeit von der Zeit um Aufträge, die schon lange warten zu bevorzugen, z.B.:
  - ➤ Realzeitsysteme
  - ➤ Mobilfunksysteme
  - ➤ Internet

- Es wird eine Prioritätsfunktion eingeführt, die angibt welchen Wert die Priorität eines Auftrags einer bestimmten Prioritätsklasse $r$ zu einer bestimmten Zeit hat:

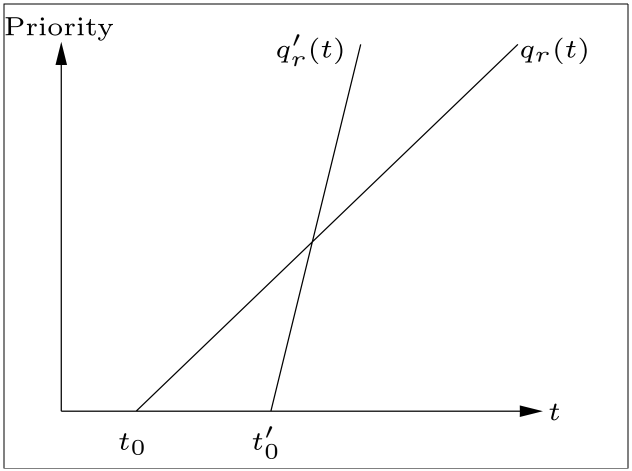
$$\boldsymbol{q_r(t)} = \text{Priority of class } \boldsymbol{r} \text{ at time } \boldsymbol{t}$$

- Prioritätsfunktion mit zeitabhängiger Anstiegsrate:

$$q_r(t) = (t - t_0) \cdot b_r$$

mit:

$$0 \leq b_1 \leq b_2 \leq \ldots \leq b_r$$

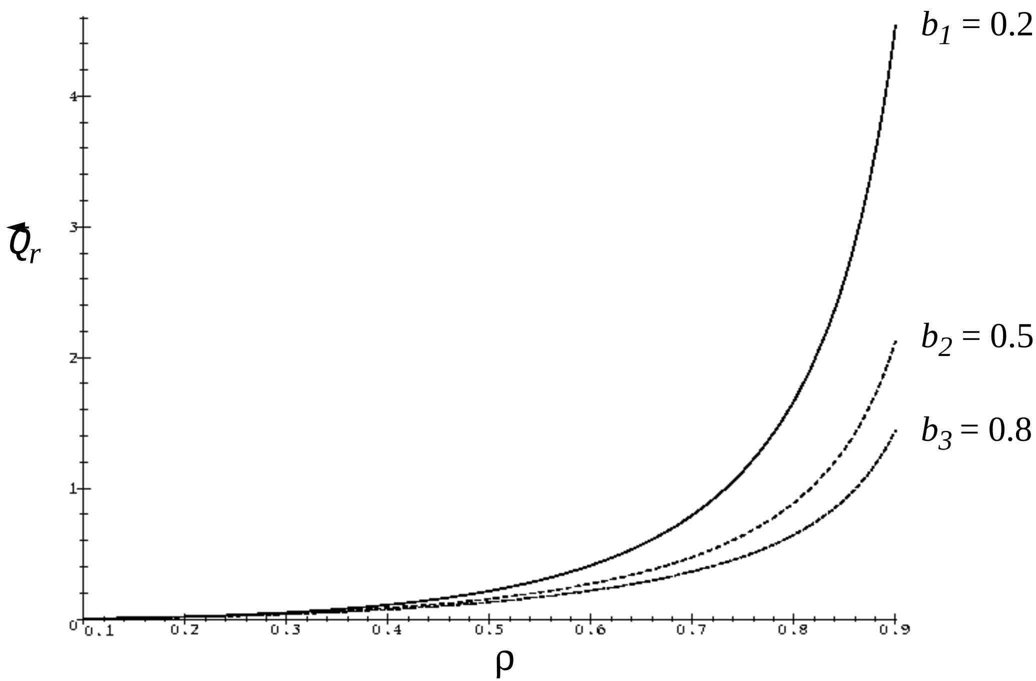


➤ Mittlere Wartezeit eines Auftrags der Prioritätsklasse $r$ :

$$\overline{W}_r = \frac{\dfrac{\overline{W}_0}{1-\rho} - \sum_{i=1}^{r-1} \rho_i \overline{W}_i \left(1 - \dfrac{b_i}{b_r}\right)}{1 - \sum_{i=r+1}^{R} \rho_i \left(1 - \dfrac{b_r}{b_i}\right)}$$

Gilt für beliebiges G/G/m-System

Mittlere Warteschlangenlänge $\overleftarrow{Q}_r$ eines M/M/1-Prioritätssytems und einer Prioritätsfunktion mit zeitabhängiger Anstiegsrate ohne Verdrängung:

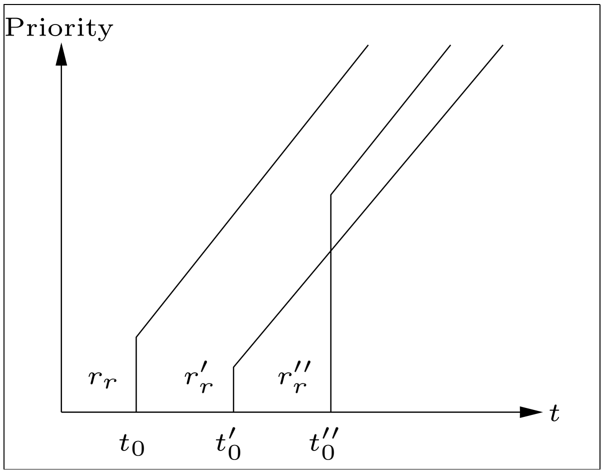


- Prioritätsfunktion mit zeitabhängiger Anfangspriorität $r_r$:

$$q_r(t) = r_r + t - t_0$$

mit:

$$0 \leq r_1 \leq r_2 \leq \ldots \leq r_r$$



◆ Mittlere Wartezeit eines Auftrags der Prioritätsklasse $r$ :

➤ "Heavy-Traffic"-Approximation (ρ --> 1):

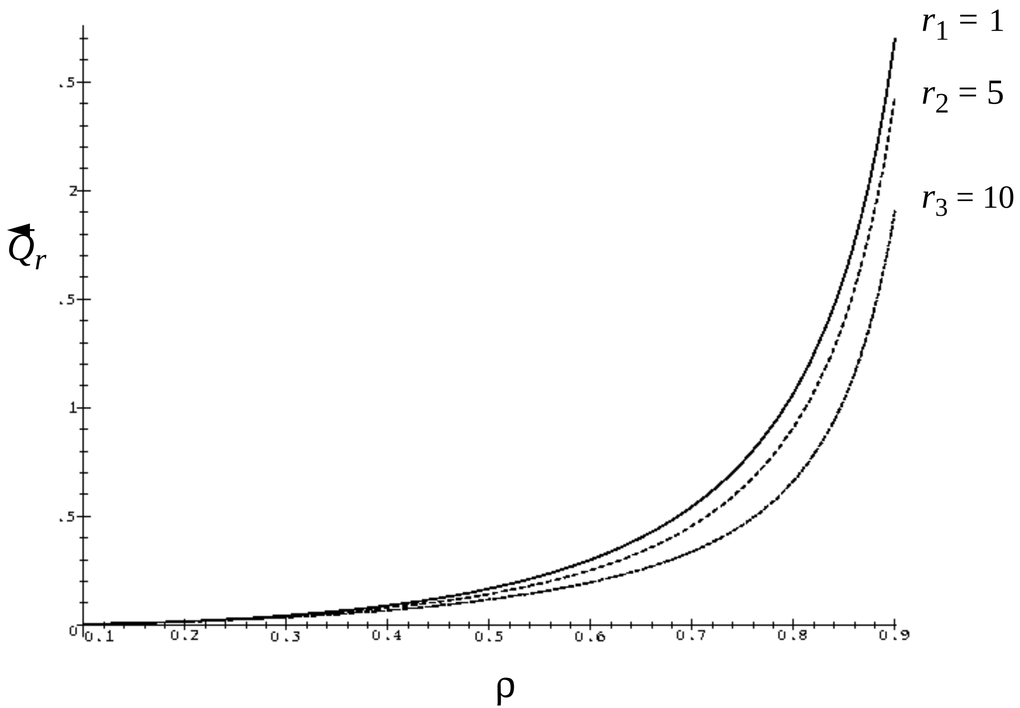
$$\overline{W}_r \approx \frac{\overline{W}_0}{1-\rho} - P_m \cdot \sum_{i=1}^{R} \rho_i (r_r - r_i)$$

➤ Genauere Approximation (0 < ρ < 1):

$$\overline{W}_r \approx \frac{\overline{W}_0}{1-\rho} - \sum_{i=1}^{r-1} \rho_i \overline{W}_i \left(1 - \exp\left(\frac{P_m (r_i - r_r)}{\overline{W}_i}\right)\right)$$

➤ Gilt für beliebige G/G/m-Systeme ($m = 1$ --> $P_m = \rho$)

Mittlere Warteschlangenlänge $\overline{Q}_r$ eines M/M/1-Prioritätssytems und einer Prioritätsfunktion mit zeitabhängiger Anfangspriorität ohne Verdrängung:



## "Deadline-orientierte" Prioritätsfunktion:

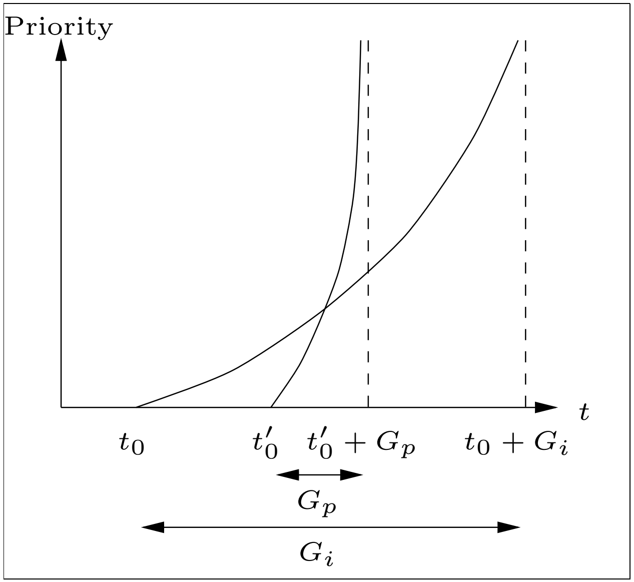
- In vielen Realzeitsystemen muss ein Auftrag zu einem bestimmtem Zeitpunkt fertigbearbeitet sein ("deadline", Fertigstellungszeitpunkt)

- Dann ist eine Prioritätsfunktion günstig, die von 0 bis oo zwischen dem Ankunftszeitpunkt $t_0$ und dem Fertigstellungszeitpunkt $u_r$ ansteigt:

- Eine geeignete Prioritätsfunktion ist:

$$q_r(t) = \begin{cases} (t - t_0)/(u_r - t + t_0) & t_0 < t \leq u_r + t_0\,, \\ \infty & u_r + t_0 \leq t\,. \end{cases}$$

➤ "Deadline-orientierte" Prioritätsfunktion:




$(G_r = u_r)$

◆ Mittlere Wartezeit eines Auftrags der Prioritätsklasse *r* :

➤ "Heavy-Traffic"-Approximation (ρ --> 1):

$$\overline{W}_r \approx \left( \frac{\overline{W}_0}{1-\rho} - P_m \sum_{i=1}^{r-1} \rho_i (u_i - u_r) \right) \left( 1 - (1 - P_m) \sum_{i=r+1}^{R} \rho_i \left( 1 - \frac{u_i}{u_r} \right) \right)^{-1}$$

➤ Genauere Approximation (0 < ρ < 1):

$$\overline{W}_r \approx \left( \frac{\overline{W}_0}{1-\rho} - \sum_{i=1}^{r-1} \rho_i \cdot \overline{W}_i \left( 1 - \frac{u_r}{u_i} \right) \left( 1 - P_m \exp\left( -\frac{\rho u_i}{\overline{W}_i} \right) \right) \right) \cdot \left( 1 - \sum_{i=r+1}^{R} \rho_i \left( 1 - \frac{u_i}{u_r} \right) \left( 1 - P_m \cdot \exp\left( -\frac{\rho u_r}{\overline{W}_r} \right) \right) \right)^{-1} .$$

Mittlere Wartezeit $\overleftarrow{W}_r$ eines M/M/1-Prioritätssytems und einer deadline-orientierten Prioritätsfunktion ohne Verdrängung: