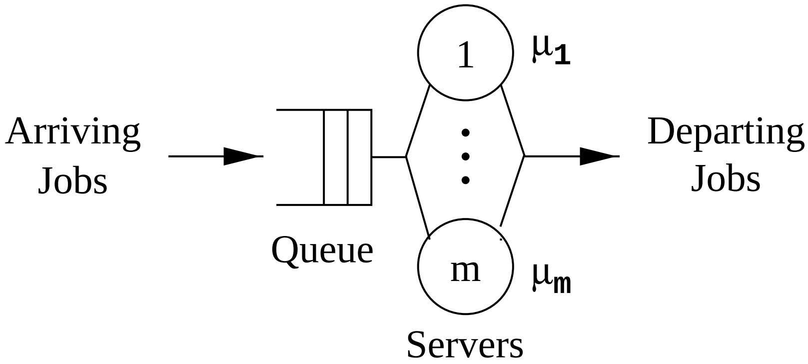
D.6 Heterogene Wartesysteme



- G/G/m - Systeme mit unterschiedlichen Bedienraten μ_i ($1 \leq i \leq m$)
- Auswahlstrategien für freie Bedieneinheiten (Bei identischen Bedienein-heiten hat Auswahlstrategie keinen Einfluss !):
  - ➤ Auswahl der **schnellsten** Bedieneinheit: **FFS** (Fastest Free Server)
  - ➤ **Zufällige** Auswahl: **Random**
- Grosse praktische Bedeutung, z.B. bei Fertigungssytemen

Reproduktion jeder Art oder Verwendung dieser Unterlage, außer zu Lehrzwecken an der Universität Erlangen-Nürnberg, bedarf der Zustimmung des Autors.

1 Approximative Analyse

- "Heavy traffic"-Aprroximation, d.h. Auslastung ρ ist nahe bei 1
  - Es ist meistens keine oder nur eine Bedieneinheit frei
  - Auswahlstrategie hat keinen Einfluss

- Wahrscheinlichkeit p_k $(1 \leq k \leq m)$, dass ein Auftrag von der Bedieneinheit k bedient wird ist proportional zur Bedienrate μ_k :

$$p_k \approx \frac{\mu_k}{\sum_{i=1}^{m} \mu_i}$$

- Damit ergibt sich für die Auslastung der Bedieneinheit k:

$$\rho_k \approx \frac{\lambda}{\mu_k} \cdot p_k = \frac{\lambda}{\mu_k} \cdot \frac{\mu_k}{\sum_{i=1}^{m} \mu_i} = \frac{\lambda}{\sum_{i=1}^{m} \mu_i}$$

Reproduktion jeder Art oder Verwendung dieser Unterlage, außer zu Lehrzwecken an der Universität Erlangen-Nürnberg, bedarf der Zustimmung des Autors.

◆ Die Auslastung aller Bedieneinheiten ist identisch und man erhält für die mittlere Auslastung des Gesamtsystems:

$$\rho = \rho_k$$

◆ Mit den Formeln für homogene M/M/m-, M/G/m- und G/G/m-Systeme können damit die Leistungsgrößen auch für heterogene Systeme berechnet werden, z.B. gilt für homogene M/M/m - Systeme:

$$\overline{K} = m\rho + \frac{\rho}{1-\rho} \cdot P_m$$

mit:

$$P_m = P(K \geq m) = \sum_{k=m}^{\infty} \pi_k$$

$$= \frac{(m\rho)^m}{m!(1-\rho)} \cdot \pi_0$$

Reproduktion jeder Art oder Verwendung dieser Unterlage, außer zu Lehrzwecken an der Universität Erlangen-Nürnberg, bedarf der Zustimmung des Autors.

- In der Tabelle sind approximative und zum Vergleich exakte Werte für die Leistunggsgrößen heterogener a) M/M/5- und b) M/E_2/5- Systeme gegeben:

	λ	μ_k	$\rho(A)$	$\rho(S_1)$	$\rho(S_2)$	$\overline{Q}(A)$	$\overline{Q}(S_1)$	$\overline{Q}(S_2)$
(a)	73	10,15,20,20,25	0.811	0.799	0.819	2.472	2.367	2.495
	73	16,17,18,19,20	0.811	0.807	0.812	2.472	2.476	2.500
	73	5,10,18,22,35	0.811	0.808	0.844	2.472	2.443	2.626
	81.11	4,8,16,32,40	0.811	0.826	0.858	2.472	2.549	2.713
	81.11	8,9,20,31,32	0.811	0.807	0.839	2.472	2.456	2.606
	81.11	8,14,20,26,32	0.811	0.801	0.830	2.472	2.442	2.569
	81.11	10,15,20,25,30	0.811	0.799	0.824	2.472	2.425	2.523
(b)	81.11	10,15,20,25,30	0.811	0.800	0.825	1.854	1.854	1.962

A: Approximation, S_1: FFS, S_2: Random

Reproduktion jeder Art oder Verwendung dieser Unterlage, außer zu Lehrzwecken an der Universität Erlangen-Nürnberg, bedarf der Zustimmung des Autors.

2 Heterogenes M/M/2 - System

- Strategie: **FFS**
- Systemzustand: (k_1, k_2)
  - $k_1 \geq 0$: Anzahl der Aufträge in der WS + Auftrag in der schnellen BE
  - $k_2 \in \{0, 1\}$: Anzahl der Aufträge in der langsamen BE
- Zustandswahrscheinlichkeiten (mithilfe der Markovanalyse):

$$\pi(k,1) = c\pi(k-1,1) = c^{k-1}\pi(1,1), \qquad \text{for} \qquad k > 1$$

mit:

$$c = \frac{\lambda}{\mu_1 + \mu_2}$$

Reproduktion jeder Art oder Verwendung dieser Unterlage, außer zu Lehrzwecken an der Universität Erlangen-Nürnberg, bedarf der Zustimmung des Autors.

und

$$\pi(0,1) = \frac{c}{1+2c}\frac{\lambda}{\mu_2}\pi(0,0)\,,$$

$$\pi(1,0) = \frac{1+c}{1+2c}\frac{\lambda}{\mu_1}\pi(0,0)\,,$$

$$\pi(1,1) = \frac{c}{1+2c}\frac{\lambda(\lambda+\mu_2)}{\mu_1\mu_2}\pi(0,0)\,.$$

und mithilfe der Normalisierungsbedingung ($\Sigma\, \pi(k_1, k_2) = 1$):

$$\pi(0,0) = [1 + \frac{\lambda(\lambda+\mu_2)}{\mu_1\mu_2(1+2c)(1-c)}]^{-1}$$

MLR
Reproduktion jeder Art oder Verwendung dieser Unterlage, außer zu Lehrzwecken an der Universität Erlangen-Nürnberg, bedarf der Zustimmung des Autors.

◆ Auslastung der BE'n:

$$\rho_1 = 1 - \pi(0,0) - \pi(0,1)$$
$$\rho_2 = 1 - \pi(0,0) - \pi(1,0)$$

◆ Mittlere Anzahl der Aufträge im System:

$$\overline{K} = \frac{1}{A(1-c)^2}$$

mit:

$$A = \left[\frac{\mu_1\mu_2(1+2c)}{\lambda(\lambda+\mu_2)} + \frac{1}{1-c}\right]$$

Reproduktion jeder Art oder Verwendung dieser Unterlage, außer zu Lehrzwecken an der Universität Erlangen-Nürnberg, bedarf der Zustimmung des Autors.

3 Heterogenes M/M/m - System

- Vorgehen in **zwei Schritten**:

  - Erst wird ein heterogenes **Verlustsystem** analysiert

    - Verlustsystem hat keine WS

    - Aufträge, die ankommen, wenn alle BE'n aktiv sind, gehen verloren

  - Dann werden mithilfe der Leistungsgrößen für das heterogene Verlustsystem die Leistungsgrößen für das entsprechende heterogene Wartesystem ermittelt

Reproduktion jeder Art oder Verwendung dieser Unterlage, außer zu Lehrzwecken an der Universität Erlangen-Nürnberg, bedarf der Zustimmung des Autors.

Heterogenes M/M/m - Verlustsystem:

◆ Strategie**: Random**

➤ Zustandswahrscheinlichkeiten:

$$\pi_g = \frac{(m-|g|)!}{m!} \cdot \prod_{k \in g} \frac{\lambda}{\mu_k} \cdot \pi_\emptyset$$

$$\text{for all } g \subseteq G \text{ with } g \neq \emptyset \text{ and } G = \{1, 2, \ldots, m\}$$

π_0 wird mithilfe der Normalisierungsbedingung ermittelt:

$$\sum_{g \subseteq G} \pi_g = 1 \qquad \text{with} \qquad G = \{1, 2, \ldots, m\}$$

Reproduktion jeder Art oder Verwendung dieser Unterlage, außer zu Lehrzwecken an der Universität Erlangen-Nürnberg, bedarf der Zustimmung des Autors.

➤ Verlustwahrscheinlichkeit (P(alle BE’n sind aktiv)):

$$\pi_m^{(L)} = \pi_{\{1,2,\ldots,m\}} = \pi_G$$

➤ Auslastung der k-ten BE: Summe der Wahrscheinlichkeiten für die Zustände bei denen die k-te BE aktiv ist:

$$\rho_k^L = \sum_{g\,:\,k \in g} \pi_g$$

Reproduktion jeder Art oder Verwendung dieser Unterlage, außer zu Lehrzwecken an der Universität Erlangen-Nürnberg, bedarf der Zustimmung des Autors.

◆ Strategie: **FFS**

➤ Abkürzungen:

$$\mu^m = (\mu_1, \mu_2, \cdots, \mu_m)$$

und

$$(\mu^{k-1}, \mu_m) = (\mu_1, \mu_2, \cdots, \mu_{k-1}, \mu_m)$$

➤ Verlustwahrscheinlichkeit:

$$\begin{aligned}\pi_m^{(L)} &= \pi_{\{1,\ldots,m\}}(\mu^m) = B_m(\mu^m) \\ &= B_{m-1}(\mu^{m-1}) \cdot \left[1 + \frac{\mu_m}{\lambda} \cdot \prod_{k=1}^{m-1} \frac{B_k(\mu^{k-1}, \mu_m)}{B_k(\mu^{k-1}, \mu_k + \mu_m)}\right]^{-1}\end{aligned}$$

Reproduktion jeder Art oder Verwendung dieser Unterlage, außer zu Lehrzwecken an der Universität Erlangen-Nürnberg, bedarf der Zustimmung des Autors.

➤ mit:

$$B_1(\mu^1) = \pi_{\{1\}}(\mu_1) = \frac{\lambda}{\lambda + \mu_1}$$

➤ Auslastung der einzelnen BE’n:

$$\rho_k = \frac{\lambda}{\mu_k}[B_{k-1}(\mu^{k-1}) - B_k(\mu^k)] \qquad \text{with} \qquad B_0(\mu^0) = 1$$

Reproduktion jeder Art oder Verwendung dieser Unterlage, außer zu Lehrzwecken an der Universität Erlangen-Nürnberg, bedarf der Zustimmung des Autors.

Heterogenes M/M/m - Wartesystem:

- Zustandswahrscheinlichkeiten:

$$\pi_i^{(W)} = \left(\sum_{k=1}^{m} \frac{\mu_k}{\lambda} \right)^{m-i} \cdot \pi_m^{(W)} \qquad \text{for} \qquad i > m$$

mit

$$\pi_m^{(W)} = \frac{\pi_m^{(L)}}{N}$$

und

$$N = 1 + \frac{\pi_m^{(L)} \cdot c}{1 - c} \qquad \text{with} \qquad c = \frac{\lambda}{\sum_{k=1}^{m} \mu_k}$$

Reproduktion jeder Art oder Verwendung dieser Unterlage, außer zu Lehrzwecken an der Universität Erlangen-Nürnberg, bedarf der Zustimmung des Autors.

◆ Wartewahrscheinlichkeit:

$$P_m = \sum_{i=m}^{\infty} \pi_i^{(W)} = \frac{1}{1-c} \cdot \pi_m^{(W)} = \frac{1}{1-c} \cdot \frac{\pi_m^{(L)}}{N}$$

◆ Mittlere Warteschlangenlänge:

$$\overline{Q} = \sum_{i=m+1}^{\infty} (i-m) \cdot \pi_i^{(W)} = \frac{P_m \cdot c}{1-c}$$

Reproduktion jeder Art oder Verwendung dieser Unterlage, außer zu Lehrzwecken an der Universität Erlangen-Nürnberg, bedarf der Zustimmung des Autors.

◆ Auslastung der einzelnen Bedieneinheiten:

$$\rho_k^{(W)} = \frac{\rho_k^{(L)}}{N} + \sum_{i=m+1}^{\infty} \pi_i^{(W)} = \frac{\rho_k^{(L)} + \pi_m^{(L)} \cdot c/(1-c)}{N}$$

◆ Gesamtauslastung (Mittlerer Anteil aktiver BE'n):

$$\rho = \frac{m}{\sum_{k=1}^{m} \rho_k^{-1}}$$

Reproduktion jeder Art oder Verwendung dieser Unterlage, außer zu Lehrzwecken an der Universität Erlangen-Nürnberg, bedarf der Zustimmung des Autors.

◆ Leistungsgrößen für ein heterogenes M/M/2 - System mit $\lambda = 0.2$, $\mu_1 = 0.5$ und $\mu_2 = 0.25$:

Strategy	Appr.	FFS(M/M/2)	FFS(M/M/m)	Random
$\boldsymbol{\rho}$	0.267	(0.267)	0.234	0.275
$\boldsymbol{\rho_1}$	0.267	0.305	0.306	0.230
$\boldsymbol{\rho_2}$	0.267	0.189	0.189	0.341
$\overline{\boldsymbol{K}}$	0.575	0.533	0.533	0.615
$\overline{\boldsymbol{T}}$	2.875	2.663	2.665	3.074

FFS: Fastest Free Server

Reproduktion jeder Art oder Verwendung dieser Unterlage, außer zu Lehrzwecken an der Universität Erlangen-Nürnberg, bedarf der Zustimmung des Autors.

- Vergleich der mittleren Antwortzeit $\overline{T}$ eines heterogenen M/M/2 -Systems (Strategie: **FFS**) mit den mittleren Antwortzeiten zweier M/M/1 - Systeme $\overline{T}_1$ und $\overline{T}_2$ ($\lambda = 0.2$, $\mu_2 = 0.25$ und $\mu_1 = \alpha\mu_2$):

$\boldsymbol{\alpha}$	1	2	3	4	5
$\overline{\boldsymbol{T}}_{\mathbf{1}}$	20	3.33	1.818	1.25	0.55
$\overline{\boldsymbol{T}}_{\mathbf{2}}$	20	20	20	20	20
$\overline{\boldsymbol{T}}$	4.762	2.662	1.875	1.459	1.20

➤ Manchmal ist es besser die langsame BE eines heterogenen M/M/m - Systems nicht zu verwenden, wenn die mittlere Antwortzeit minimal sein soll. Dies gilt wenn die Bedienraten sehr unterschiedlich sind und die Auslastung sehr gering ist.

Reproduktion jeder Art oder Verwendung dieser Unterlage, außer zu Lehrzwecken an der Universität Erlangen-Nürnberg, bedarf der Zustimmung des Autors.

D.7 Batch-Systeme

- Die Bearbeitung mehrerer Aufträge (eines **Batch**) wird gleichzeitig gestartet und durchgeführt.

- **Full-Batch**-Strategie (FB):
  Die Bearbeitung eines Batch wird gestartet, wenn alle b Aufträge des Batch angekommen sind.

- **Minimum-Batch**-Strategie (MB):
  Die Bearbeitung eines Batch wird gestartet, wenn mindestens
  a Aufträge des Batch angekommen sind.
  Warten mehr als b Aufträge, dann werden b Aufträge zu einem Batch zusammengefasst und gemeinsam bearbeitet.

- Ein Spezialfall der MB-Strategie ist die **GREEDY**-Strategie:
  Die Beabeitung beginnt, wenn $a = 1$ Auftrag in der WS ist.

Reproduktion jeder Art oder Verwendung dieser Unterlage, außer zu Lehrzwecken an der Universität Erlangen-Nürnberg, bedarf der Zustimmung des Autors.

◆ Kendall’sche Notation für Batch-Systeme:

$$\mathrm{G/G}^{[\boldsymbol{b},\boldsymbol{b}]}/\mathrm{m}, \text{ Multiserver system with FB policy}$$

$$\mathrm{G/G}^{[\boldsymbol{a},\boldsymbol{b}]}/\mathrm{m}, \text{ Multiserver system with MB policy}$$

Reproduktion jeder Art oder Verwendung dieser Unterlage, außer zu Lehrzwecken an der Universität Erlangen-Nürnberg, bedarf der Zustimmung des Autors.

◆ Full-Batch (FB):

➤ Berechnung der mittleren WS-Länge $\hat{Q}$ für die Batches mithilfe der bekannten Formeln für G/G/m-Systeme, mit:

➤ Ankunftsrate der Batches (λ : Ankunftsrate der einzelnen Aufträge):

$$\hat{\lambda} = \frac{\lambda}{b}$$

➤ Variationskoeffizient der Zwischenankunftszeit der Batches ($c_A{}^2$: Variationskoeffizient der Zwischenankunftszeit der einzelnen Aufträge):

$$\hat{c}_A^2 = \frac{c_A^2}{b}$$

➤ Mittlere WS-Länge eines einzelnen Auftrags in einen Full-Batch-System:

$$\overline{Q} \approx b \cdot \hat{Q} + \frac{b-1}{2}$$

Reproduktion jeder Art oder Verwendung dieser Unterlage, außer zu Lehrzwecken an der Universität Erlangen-Nürnberg, bedarf der Zustimmung des Autors.

◆ Minimum-Batch (MB):

➤ Mittlere WS-Länge eines einzelnen Auftrages:

$$\overline{Q}^{[a,b]} \approx b \cdot \hat{Q} + P_m \cdot \frac{b-1}{2} + (1 - P_m) \cdot \frac{a-1}{2}$$

mit P_m: Wartewahrscheinlichkeit des entsprechenden M/M/m-Systems

Reproduktion jeder Art oder Verwendung dieser Unterlage, außer zu Lehrzwecken an der Universität Erlangen-Nürnberg, bedarf der Zustimmung des Autors.

◆ Beispiel: $M/M^{[b,b]}/m$ - System

➤ Fullbatchsystem mit Batchlänge b

➤ Exponentiell verteilte Zwischenankunftszeit und Bedienzeit der einzelnen Aufträge eines Batches

➤ Bedienzeit des Batches ist auch exponentiell verteilt, da die Aufträge parallel bearbeitet werden

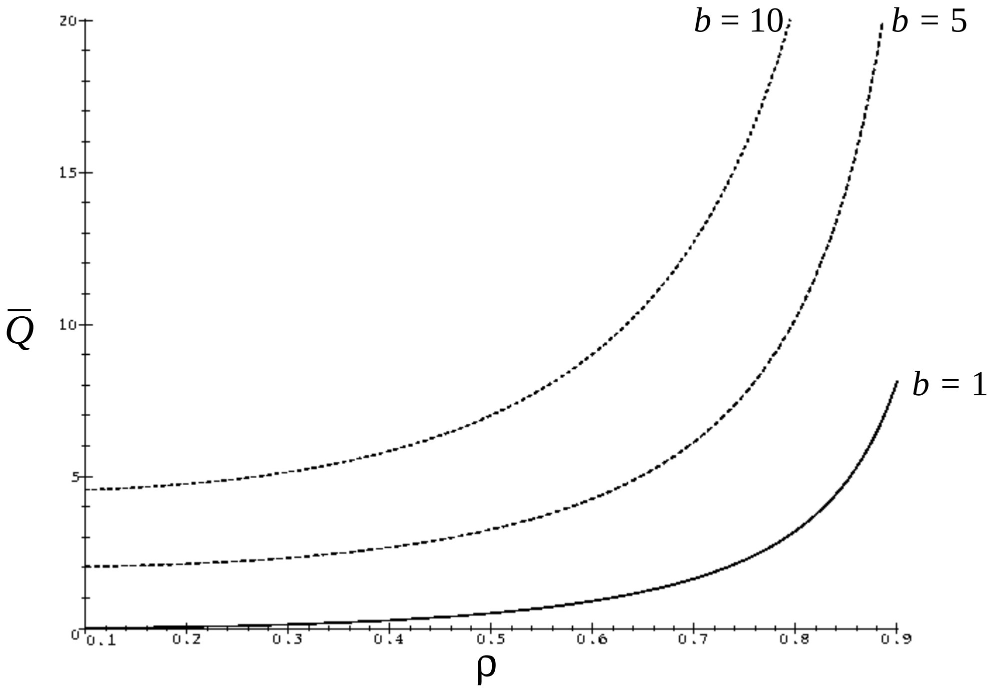
➤ Zwischenankunftszeit eines Batches ist E_b-verteilt, da erst b Aufträge ankommen müssen bis mit der Bearbeitung begonnen wird.
(Summe von exponentiell verteilten Zeiten ergibt eine Erlangverteilung der Gesamtzeit)

➤ Zur Berechnung der mittleren WS-Länge $\hat{Q}$ für die Batches muss man daher ein E_b/M/m-Wartesystem verwenden

➤ Daraus berechnet man dann mit den angegeben Formeln die mittlere WS-Länge $\overline{Q}$ für einzelne Aufträge

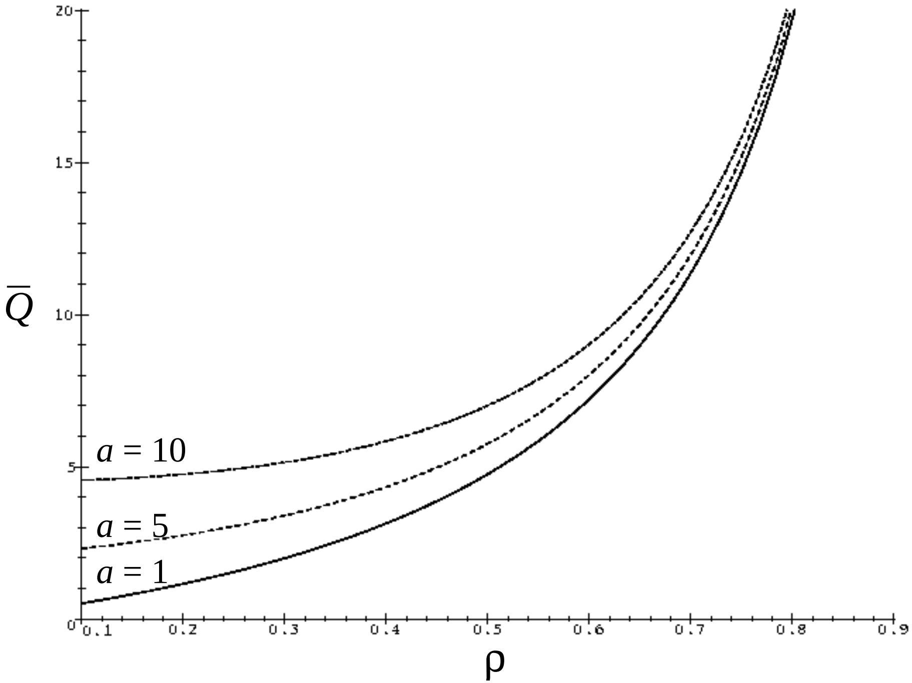
Reproduktion jeder Art oder Verwendung dieser Unterlage, außer zu Lehrzwecken an der Universität Erlangen-Nürnberg, bedarf der Zustimmung des Autors.

Mittlere WS-Länge $\overline{Q}$ für einzelne Aufträge in einem FB-System



Reproduktion jeder Art oder Verwendung dieser Unterlage, außer zu Lehrzwecken an der Universität Erlangen-Nürnberg, bedarf der Zustimmung des Autors.

Mittlere WS-Länge $\overline{Q}$ für einzelne Aufträge in einem MB-System ($b = 10$)



Reproduktion jeder Art oder Verwendung dieser Unterlage, außer zu Lehrzwecken an der Universität Erlangen-Nürnberg, bedarf der Zustimmung des Autors.