Kapitel 11

Systeme von Funktionseinheiten

11.1 Einleitung

Handlungen, Aktionen oder Transitionen wurden in den vorangehenden Kapiteln wiederholt betrachtet. Hier wird zunächst auf den philpsophischen Handlungsbegriff eingegangen. Später werden Handlungen als Aufträge zu diese ausführende Funktionseinheiten in Beziehung gesetzt. Auftragsbearbeitung wird bei großen Auftragsmengen stochastisch beschrieben. Mit dem Gesetz von Little und dem Begriff der Sättigungsfüllung wird in diese Betrachtungsweise eingeführt. Diese Modellierungstechnik führt in weiterführenden Vorlesungen zur Verkehrsanalyse durch Wartesysteme.

11.2 Handlung, Auftrag, Funktionseinheit

Da es unser Ziel ist, Vorgänge in Systemen zu beschreiben, befassen wir uns zunächst mit dem elementaren Begriff der Handlung. Als Handlung können Phänomene sehr unterschiedlicher Komplexität verstanden werden: ein Rechensystem aus mehreren Prozessoren und peripheren Geräten vollbringt eine Handlung, ebenso der Benutzer, der auf einen Knopf drückt.

In beiden Beispielen wurde nicht gesagt, wie diese Handlung durchgeführt wurde, sondern die vollzogene Änderung beschrieben. Daher wollen wir, möglichst allgemein, eine Handlung durch ihre Wirkung definieren.

Definition 11.1 *Unter einer* Handlung *verstehen wir die (möglichst zweckvolle) Herstellung oder Änderung eines Objektes.*

Dieser Begriff stellt eine beträchtliche Einengung des philosophischen Handlungsbegriffes dar. Schon in der Antike unterscheidet Aristoteles in seiner Nikomachischen Ethik (Aristoteles 69) das *Herstellen* vom *Handeln*. Herstellen hat sein Ziel außerhalb des Tuns in einer von ihm ablösbaren (abstrahierbaren) Sache. Handeln hat dagegen sein Ziel im Tun, z.B. wenn wir mit anderen verständnisvoll kommunizieren. Danach können Maschinen nicht handeln, da sie keinen freien Willen haben und nur Menschen zum ,,Verstehen" fähig sind.

Der philosophische Handlungsbegriff wurde in neuer Zeit z.B. von J. Habermas weiterentwickelt, indem er rationales (d.h. vernünftiges) Handeln in Bezug auf *kognitiv-instrumentelle Rationalität* einerseits und *kommunikative Rationalität* andererseits differenziert (Habermas 81). Im hier relevanten Kontext - ist diese Unterscheidung wichtig, weil zunehmend, insbesondere im Sprachgebrauch über die ,,Mensch-Maschine-Kommunikation", die Tendenz besteht, den Begriff auch für den Menschen auf seine instrumentelle Komponente zu verkürzen (siehe auch (Schnädelbach 84)).

In Def. 11.1 liegt eine *extensionale* Begriffsbildung vor. Die Extension eines Begriff es ist sein Umfang, d.h. die Menge der Gegenstände oder Situationen, auf die er zutrifft. In unserem Falle bedeutet dies, dass Handlungen, die die gleiche Änderungen bewirken, als gleich oder zumindest ununterscheidbar betrachtet werden. Auch für andere Begriffe dieser Vorlesung gilt das Zitat aus (Klaus 69): "Es ist immer anzustreben, wissenschaftliche Formulierungen, Theorien usw. in einer extensionalen Sprache darzustellen, da diese Sprachen bedeutend leichter zu handhaben sind, sich präziser formulieren lassen. Besonders wichtig ist, dass sich die Probleme der maschinellen Rechentechnik wesentlich leichter bewältigen lassen, wenn der Bereich, in den das betreffende Problem fällt, in einer extensionalen Sprache dargestellt ist."

Handlungen können auf verschiedene Arten festgelegt und beschrieben werden. Bekannt sind Darstellungen durch Transitionen in Netzen, durch Algorithmen, Programme, Nachrichten, Texte usw. Jede solche Beschreibung, festgehalten auf einem Datenträger, nennen wir *Handlungsanweisung* oder kürzer *Anweisung*. Eine Anweisung kann i.a. eine Menge von Handlungen beschreiben.

Ist eine Handlung in Teilhandlungen zerlegbar, so entsprechen diesen meist auch Teilanweisungen. In einer gegebenen Betrachtungsebene (z.B. höhere Programmsprache, Maschinensprache) sind solche Verfeinerungen nur bis zu einem gewissen Grade ökonomisch und sinnvoll. Meist sind nicht mehr weiter teilbare Handlungen, die wir *elementar* nennen, durch einen Handelnden oder Prozessor festgelegt. Dieser wird interpretationsneutral *Funktionseinheit* genannt.

Einerseits ist also die Spezifikation einer Handlungsanweisung nur sinnvoll in Hinblick auf eine Funktionseinheit – andererseits nennen wir alles, was Anweisungen ausführen kann, Funktionseinheiten. Ein ,,Auftrag" bindet die Handlungsanweisung an eine geeignete Funktionseinheit.

Definition 11.2 *Unter einer* Funktionseinheit FE *(functional unit) verstehen wir ein durch Aufgabe und Wirkung abgrenzbares Gebilde, wobei Aufgabe oder Wirkung in Handlungen bestehen. Hierzu kann etwa die Verarbeitung von Daten ebenso wie die Übertragung oder Speicherung von Daten gehören.*

Der erste Halbsatz dieser Definition entspricht *DIN 44300* bzw. *DIN 66200*. Einschränkend fassen wir z.B. eine Nachricht nicht als Funktionseinheit auf. Eine Anweisung, die an eine Funktionseinheit gegeben wird, um sie zu einer Handlung zu verpflichten, begründet einen Auftrag an eine Funktionseinheit.

Definition 11.3 *Ein* Auftrag *(task) ist eine Verpflichtung zur Ausführung einer Handlung. Ein Auftrag wird durch Übergabe einer* (Handlungs-)Anweisung *(statement), der* Auftragsbeschreibung*(task specification), an die verpflichtete, den Auftrag ausführende Funktionseinheit begründet (vgl.* DIN 66200*). Wir stellen Aufträge daher oft als Paar* a $= (a, R)$ *dar, wobei* a *die Anweisung und* R *die verpflichtete Funktionseinheit darstellt.*

Definition 11.4 *Die* Auftragsbeschreibung *ist die Beschreibung der auszuführenden Handlung eines Auftrages. Sie ist in einer* Auftragssprache *auf einem* Datenträger *formuliert. Anweisungen in Form eines Textes nennen wir auch* Programm*,* Protokoll *oder* Prozessanweisung*. Zur graphischen Darstellung von Anweisungen benutzen wir statt Flußdiagrammen* Netze*.*

Eine Handlung in einem Rechensystem kommt durch die Verpflichtung einer Funktionseinheit durch einen entsprechenden Auftrag zustande (Def. 11.3). Verschiedene Aufträge bzw. Funktionseinheiten können durch Verknüpfung komplexe Strukturen bilden, die wir Auftragssysteme bzw. Systeme von Funktionseinheiten nennen. In diesem Abschnitt wird beschrieben, wie Auftragssysteme in Systemen von Funktionseinheiten bearbeitet werden.

Eine Funktionseinheit ist (nach Def. 11.2) ein durch Aufgabe oder Wirkung abgrenzbares Gebilde, wobei Aufgabe oder Wirkung in Handlungen bestehen. Die Funktionseinheit braucht nicht durch ein Gerät (wie z.B. einen Drucker) gegeben sein: Programme, Kommunikationssysteme sind solche Funktionseinheiten. Ein Nachrichtenkanal kann z.B. sowohl als Gerät, wie auch als Funktionseinheit betrachtet werden. In *DIN 40150* ist deshalb der Oberbegriff *Betrachtungseinheit* als ,,ein nach Aufgabe und Umfang abgegrenzter Gegenstand einer Betrachtung" definiert worden. Darunter fallen ,,Funktionseinheit", deren Abgrenzung nach Aufgabe und Wirkung erfolgt, wie auch ,,Baueinheit", deren Abgrenzung nach Aufbau und Zusammensetzung erfolgt. Es werden aber auch andere Beispiele von Betrachtungskriterien genannt:

- Betriebs- / Nutzungseinheit
- Instandhaltungseinheit
- Prüfeinheit
- Austauscheinheit
- Fertigungseinheit

Betrachtungsebenen können hierarchisch geordnet sein. Dies entspricht einer Vorstellung von Verfeinerung und Vergröberung von Handlungen. *DIN 40150* schlägt eine Unterteilung in vier Betrachtungsebenen vor :

1. *Element*, als in Abhängigkeit von der Betrachtung als *unteilbar* aufgefaßten Einheit der *untersten Betrachtungsebene*

2. *Gruppe*, als Zusammenfassung von Elementen in einer höheren Betrachtungsebene zu einer noch *nicht selbstständig verwendeten* Betrachtungseinheit
3. *Einrichtung*, als Zusammenfassung von Elementen und/oder Gruppen in einer nächsthöheren Betrachtungsebene zu einer *selbstständig verwendbaren* Betrachtungseinheit
4. *System*, als Gesamtheit der zur selbstständigen Erfüllung eines Aufgabenkomplexes erforderlichen technischen und/oder organisatorischen und/oder anderen Mittel der *obersten Betrachtungsebene* (zu organisatorischen Mitteln zählen auch Personal und Ausbildung).

Betrachtungskriterien können mit Betrachtungsebenen kombiniert werden.

Betrachtungs-	-Kriterien					
		Funktionseinheit	Baueinheit	Betriebseinheit	Instandhaltungseinheit	...
-Ebenen	Element	Funktionselement	Bauelement	...		
	Gruppe	Funktionsgruppe	Baugruppe	...		
	Einrichtung	funktionale Einrichtung	bauliche Einrichtung	...		
	System	funktional abgegrenztes System	baulich abgegrenztes System	...		

Diese Begriffe sind zum Gebrauch für alle Bereiche der Technik bestimmt. Wir empfehlen ihre Benutzung in der Informatik.

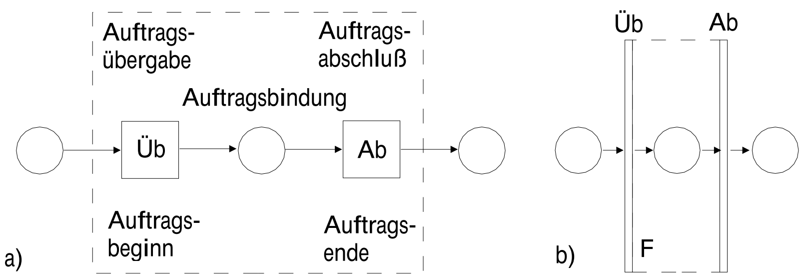


Abbildung 11.1: Auftrag und Funktionseinheit: a) Begriff e, b) Kurzdarstellung

Um Begriffe wie Füllung, Durchsatz und Verweilzeit definieren zu können, ist es notwendig, den Vorgang der Auftragsausführung durch eine Funktionseinheit F genauer zu betrachten, wobei man sich an der Abb. 11.1 orientieren kann.

Definition 11.5 *Ein* Auftrag *ist eine Verpflichtung einer Funktionseinheit F zur Ausführung einer Handlung mittels einer Auftragsbeschreibung (vgl. Def. 11.3). Der Zustand der Verpflichtung heißt* Auftragsbindung. *Verschiedene Auftragsbindungen werden durch* Auftragskennzeichen *unterschieden. F heißt* Auftragsnehmer. *Die* Übergabe *der Auftragsbeschreibung (der Anweisung) heißt auch* Auftragserteilung. Auftragsbeginn *bzw.* Auftragsende *ist das Ereignis, das Beginn bzw. Ende einer Auftragsbindung definiert, i.a. Auftragserteilung und Abschluß der zugehörigen Handlung. Mit Auftragsende ist der Auftrag* erledigt (ausgeführt).

Als Beispiel betrachten wir die Anweisung

$$\boldsymbol{do}\ i \leq 2 \rightarrow i := i + 1\ \boldsymbol{od}$$

und eine Funktionseinheit F (wie in Abb. 11.1), die den Speicherplatz i mit Anfangswert $i = 1$ und die Operation „erhöhe um 1" verfügbar haben möge.

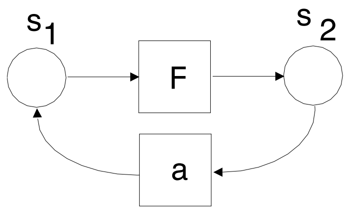


Abbildung 11.2: Wiederholt nutzbare Funktionseinheit

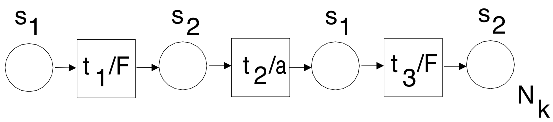


Abbildung 11.3: Prozess zum P/T-Netz N in Abb. 11.2

Abbildung 11.2 zeigt die Funktionseinheit F (als Vergröberung von Abb. 11.1 b)). Wir interpretieren das Schalten von F als Auftragsbearbeitung und das Schalten von a als Vorbereitung zur erneuten Übergabe. Abbildung 11.3 zeigt einen Prozess N_K des P/T-Netzes N (im Sinne von Def. 5.26). Die Transitionen t_1 und t_3 unterscheiden als Auftragskennzeichen die beiden Auftragsbindungen der Anweisung „$i := i{+}1$" an F. Anders als in diesem Beispiel können Funktionseinheiten zu einem Zeitpunkt mehrere Auftragsbindungen eingegangen sein. Ihre Anzahl heißt Füllung der Funktionseinheit.

Definition 11.6 *Die* Füllung *f (number of tasks, population) einer Funktionseinheit ist die Anzahl der Aufträge, die der Funktionseinheit (zu einem Zeitpunkt) übergeben, aber noch nicht* *erledigt worden sind. Bei Füllung $f = 0$ heißt die Funktionseinheit* frei *(idle), bei $f > 0$* beschäftigt *(busy). Bei $f > 1$ sprechen wir auch von* Simultanbetrieb *(parallel mode). Existiert eine maximale Füllung, dann heißt diese* Kapazität *k (capacity) der Funktionseinheit. Ist $k = 1$, dann heißt die Funktionseinheit* einfach *(simple), bei $f = k$ heißt sie* belegt *(occupied). $\phi := \frac{f}{k}$ heißt relative* Füllung *(relative population) der Funktionseinheit.*

Die Kapazität einer Funktionseinheit F kann in der Netzdarstellung durch die Kapazität einer Stelle (Abb. 11.4 a)) oder explizit, aber äquivalent, durch eine komplementäre Stelle $\overline{f}$ (Abb. 11.4 b)) ausgedrückt werden.

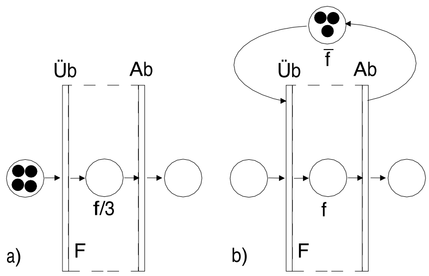


Abbildung 11.4: Funktionseinheit mit Kapazität 3

Simultanbetrieb einer Funktionseinheit kann z. B. durch Lagerung von Teilaufträgen (Abb. 11.5 a)) oder durch kollaterale Handlungen (Def. 5.27) in Untereinheiten F_i (Abb. 11.5 b)) entstehen. Abbildung 11.6 zeigt eine mögliche Verfeinerung von Abb. 11.5 a) (die Stellen ,,Auftragspuffer" und ,,Ergebnispuffer" sollen die Kapazität 1 haben). Die Funktionseinheit enthält eine einfache Funktionseinheit ,,Bearbeitung" und eine nicht einfache Funktionseinheit ,,Lagerung" mit Kapazität 5. Wegen der beiden Puffer hat die gesamte Funktionseinheit F die Kapazität 8.

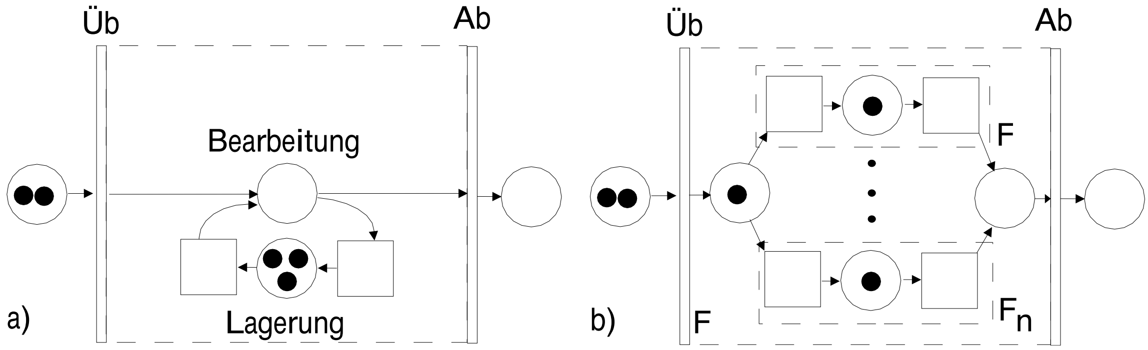


Abbildung 11.5: Simultanbetrieb einer Funktionseinheit : Lagerung und Parallelarbeit

Um komplexe Systeme einfacher darzustellen bzw. große P/T Systeme zu vereinfachen, werden gefärbte Netze eingesetzt. Das Netz in Abb. 11.7 enthält z. B. die individuellen Marken a,b,c,d

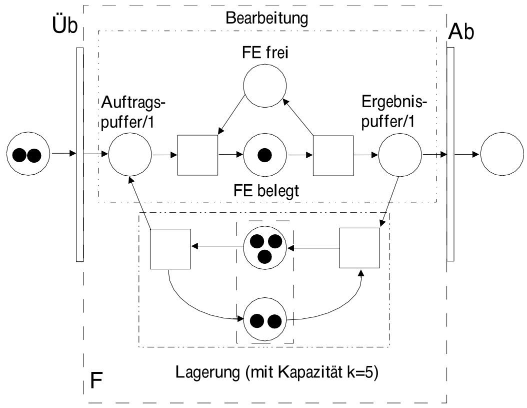


Abbildung 11.6: Verfeinerung von Abb. 11.5 a)

für die entsprechenden Aufträge.

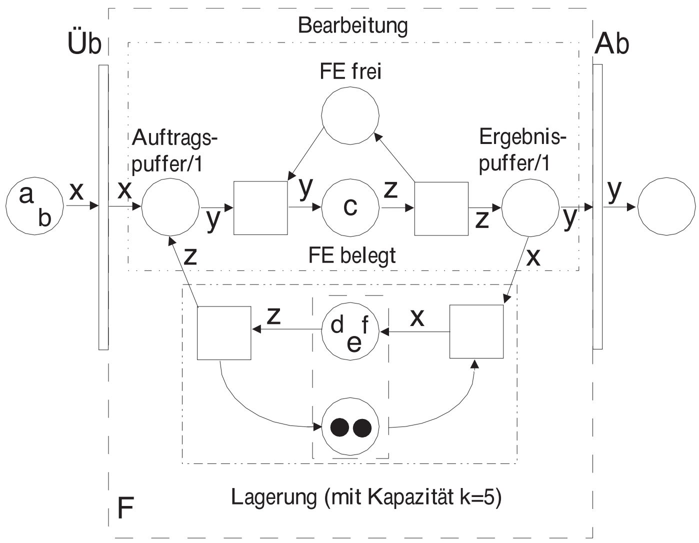


Abbildung 11.7: Funktionseinheit F von Abb. 11.6 als gefärbtes Netz

Definition 11.7 *Es sei a ein Auftrag, der zum Zeitpunkt t_1 an eine Funktionseinheit F übergeben und zum Zeitpunkt t_2 abgeschlossen wird. $y = t_2 - t_1$ heißt dann* Verweilzeit *(sojourn time, delay) von a in F. Der* Durchsatz *(throughput) $d(t_1, t_2)$ einer Funktionseinheit F im Zeitintervall $(t_1 \ ; t_2)$ mit $(t_1 < t_2)$ ist die Zahl der Auftragsende-Ereignisse aller von F im Intervall $(t_1 \ ; t_2)$ ausgeführten Aufträge geteilt durch $t_2 - t_1$. Besteht – bestimmt durch die Art der Aufträge und der Funktionseinheit – ein maximaler Durchsatz, so heißt dieser* Grenzdurchsatz *(bandwidth) c von F.* Relativer Durchsatz *oder* Auslastung *(utilization) von F heißt der Quotient $\rho = \frac{d}{c}$.*

Besonders übersichtliche Verhältnisse liegen bei einfachen Funktionseinheiten (Def. 11.6) vor, die ein wichtiges Modellierungsmittel für Rechensysteme sind. Im Intervall $(t_1 \; ; t_2)$ möge eine solche Funktionseinheit Aufträge verarbeiten, die Verweilzeiten $y_j (j \in \{1 \dots n\})$ (Def. 11.7) erfordern. Da eine einfache Funktionseinheit nach Auftragsbeginn eines einzigen Auftrages belegt ist (Def. 11.6), verursachen die Aufträge die Belegtzeit

$$x = \sum_{j=1}^{n} y_j \tag{11.1}$$

der Funktionseinheit (Def. 11.8). Die mittlere erforderliche Verweilzeit (Def. 11.7) der betrachteten Aufträge ist

$$\overline{y} = \frac{x}{n} = \frac{1}{n} \sum_{j=1}^{n} y_j \tag{11.2}$$

Von Aufträgen dieser Charakteristik kann die Funktionseinheit einen Grenzdurchsatz (Def. 11.7)

$$c = \frac{1}{\overline{y}} = \frac{n}{\sum\limits_{j=1}^{n} y_j} \tag{11.3}$$

leisten; sie hat eine Auslastung (Def. 11.7)

$$\rho = \frac{d}{c} = \frac{n \cdot \sum\limits_{j=1}^{n} y_j}{(t_2 - t_1) \cdot n} = \frac{1}{t_2 - t_1} \cdot \sum_{j=1}^{n} y_j \tag{11.4}$$

Mithin ist in $(t_2 \; ; t_1)$ ihre Auslastung gleich der relativen Belegtzeit $x \cdot (t_2 - t_1)^{-1}$! Man beachte, dass diese naheliegende Auffassung von ,,Auslastung" aus der wesentlich allgemeineren Definition 11.7 nur für den Speziallfall der einfachen Funktionseinheit folgt.

Definition 11.8 Belegtzeit *einer Funktionseinheit ist innerhalb eines Betrachtungsintervalls $(t_1 \; ; t_2)$ die Zeit, in der die Funktionseinheit belegt (Def. 11.6) ist. Die* Bedienzeit *b (service time) eines Auftrags in einer Funktionseinheit ist seine Verweilzeit (Def. 11.7) bei Füllung 1, d.h. wenn kein anderer Auftrag in der Funktionseinheit verweilt.*

Der Begriff Bedienzeit wird üblicherweise dazu benutzt, diejenige Verweilzeit zu bezeichnen, die sich bei günstigster Bedienung aufgrund der Charakteristik des Auftrags und der ausführenden Instanz ergibt, also ohne irgendwelche Verzögerung durch Bedienung von Mitbewerbern oder gar durch mutwillige Verschleppung. Mit der von uns gebrauchten Definition bemühen wir uns wieder, den Begriff auf einfach nachweisbare Größen zurückzuführen, da sich i.a. bei Füllung 1 der günstigste Fall für die Verweilzeit eines Auftrages ergibt.

Da also auch die Verweilzeit eines Auftrages einer einfachen Funktioneinheit eine solche Bedienzeit b mit dem Mittelwert $\overline{b}$ ist, folgt für den Grenzdurchsatz der einfachen Funktionseinheit nach 11.3:

$$c = \frac{1}{\overline{b}} \tag{11.5}$$

Diese Definitionen setzen wegen ihres Bezugs auf Zeitpunkte einen sychronen Prozess voraus (Def. 5.27), d.h. eine Bestimmung der Zeitpunkte a(Üb) = e(Üb) = t_1 und a(Ab) = e(Ab) = t_2, wenn man diese Vorgänge als zeitlos idealisiert.

Bei Funktionseinheiten mit relativ zur Informationsübertragungsgeschwindigkeit weit auseinanderliegenden Untereinheiten für Übergabe und Abgabe von Aufträgen kann dies zu Problemen führen. Man denke z. B. an eine Funktionseinheit zur Nachrichtenübertragung im Weltall (FE = Sender + Empfänger) oder an nebenläufige Schaltelemente in Hochleistungsrechnern mit sehr kurzen Schaltzeiten

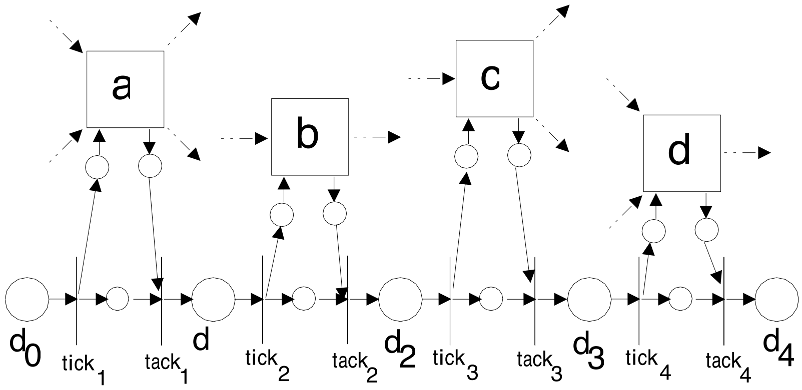


Abbildung 11.8: Getaktete Ereignisse

Die Bestimmung von Ereigniszeitpunkten kann durch *Messen* oder durch *Takten* erfolgen. Beides setzt eine für die fraglichen Ereignisse zentrale Funktionseinheit ,,Uhr" voraus, mit der kommuniziert wird. Abbildung 11.8 zeigt einen Prozess, in dem die (zeitlosen) Ereignisse $tick_i$ und $tack_j$ einer Funktionseinheit ,,Uhr" auf der hervorgehobenen Linie durch $d_0, d_1, \ldots, d_4$ liegen. Die ,,Ereigniss-Intervalle" (e($tick_i$) ; a($tack_j$)) $(1 \leq i \leq 4)$ bestimmen den ,,Tack", d.h. wann Ereignisse a, ..., d stattzufinden haben. Die ,,Taktintervalle" (e($tick_i$) ; a($tack_j$)) $(1 \leq i \leq 4)$ werden auch durch die Uhr bestimmt, und sind i.a. groß gegenüber den Ereignisintervallen (siehe auch Richter 85).

Abbildung 11.9 zeigt eine Funktionseinheit F, die durch eine Uhr getaktet wird. Zu jedem Ereignisintervall werden drei Aufträge übernommen. Jeder Auftrag bleibt 4 Taktintervalle in der Funktionseinheit. Folglich beträgt die Verweilzeit y = 4 Einheiten. Der Durchsatz beträgt d = 3, da 3 Abgänge in jedem Ereignisintervall stattfinden. Die Füllung ist f = 12.

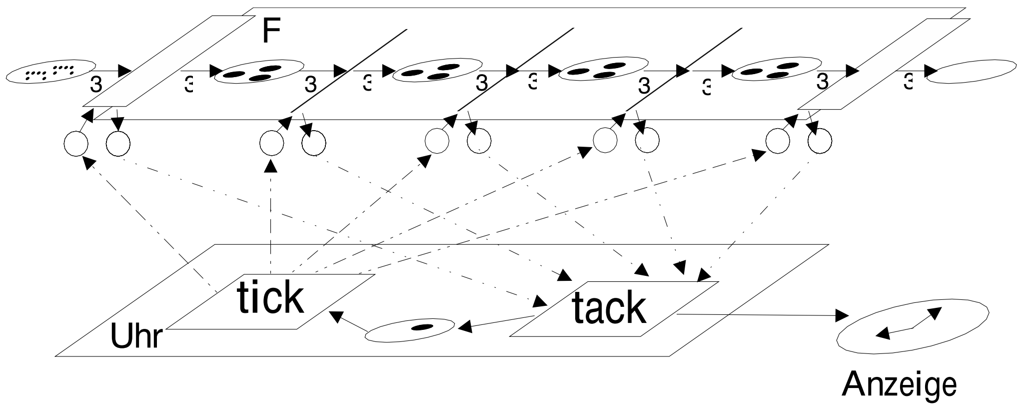


Abbildung 11.9: Funktionseinheit mit Uhr zum Gesetz von Little

Im Abschnitt 11.3 wird Little's Formel behandelt werden. Sie sagt aus, dass unter geeigneten Umständen der zeitliche Mittelwert der Füllung f gleich dem Produkt aus der mittleren Verweilzeit y und dem Durchsatz d ist. Das vorstehende Beispiel stellt einen sehr einfachen (deterministischen) Fall dieser Formel

$$f = y \cdot d$$

dar und veranschaulicht das allgemeine Gesetz.

In Definition 11.5 wurde ein Auftrag als Verpflichtung einer Funktionseinheit F zur Ausführung einer Handlung definiert. Kann man eine Instanz angeben, von der die Verpflichtung ausgesprochen wird, so nennt man diese *Auftraggeber*. *DIN 66200* behandelt den Fall, wo F ein Rechensystem und der Auftraggeber eine Person, ein *Benutzer*, ist.

Abbildung 11.10 zeigt den Rechenbetrieb als Netz einschließlich der dazu in *DIN 66200* definierten Begriffe. Die Kanten zwischen „Ein-" bzw. „Ausgabekanal" und „Bearbeiten" stellen Interaktion des Benutzes mit bereits in Bearbeitung befindlichen Aufträgen dar, wie z.B. Dialogbetrieb. Die Verantwortung für einen Auftrag kann natürlich nur bei der Person des Benutzers liegen. Wir zitieren aus der DIN-Norm:

„Durch die Verwendung der Wörter *Auftrag* und *benutzen* bei der Festlegung des Begriffsystems wird deutlich, dass man von der Funktion des Rechensystems unterschiedliche Auffassungen haben kann: auf der einen Seite die Funktion einen (Geschäfts-) *Partners* dem man im Rahmen eines allgemienen Vertragsverhältnisses *Aufträge erteilt*, auf der anderen Seite die eines *Werkzeugs*, das man für die Lösung bestimmter Aufgaben *benutzt*. Diese beiden Auffassungen lassen sich durch die Vorstellung miteinander vereinbaren, dass die Rolle des Auftragsnehmers teilweise von einer (gegebenenfalls juristischen) Person auf ein Werkzeug übergehen kann. Dieser Vorgang geschieht durch *Delegation* von Aufgaben."

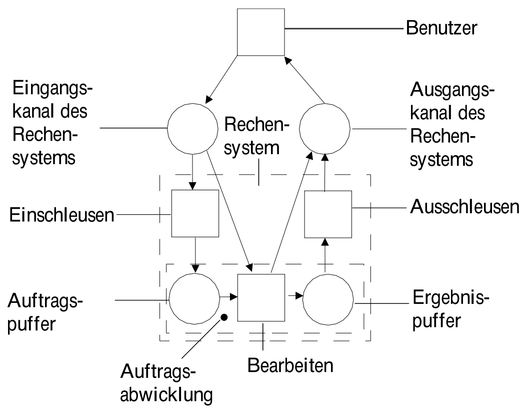


Abbildung 11.10: Rechenbetrieb als Netz

Bei der Modellierung gefärbter Petrinetze können statt Marken auch ganze Auftragssysteme (siehe Abbildung 5.9) als Individuen genutzt werden. Wir betrachten nun ein System FS von Funktionseinheiten mit n_F Funktionseinheiten $1, \ldots, n_F$, welches als gefärbte Petrinetze dargestellt werden kann, aber nicht muß. Im Intervall (t_1 , t_2) mögen n_A Aufräge $1, \ldots, n_A$ von dem System FE erledigt werden.

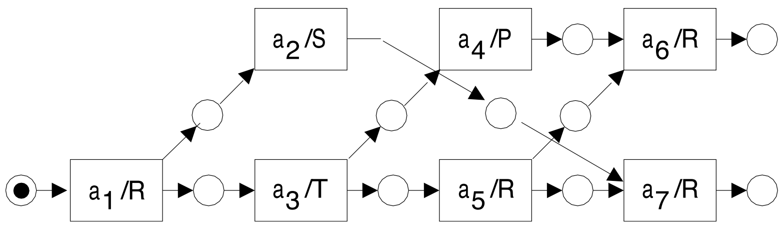


Abbildung 11.11: Auftrag mit Anschrift zugehöriger Funktionseinheit

Bei Abarbeitung eines an FS gegebenen Aufrags j komme es in der Funktionseinheit i zu v_j Teilaufträgen der Verweilzeit $y_{ijl}(l \in \{1, \ldots v_j)$. v_{ij} heißt Besuchszahl. Beispielsweise ist die Besuchszahl v_{ij} für den Auftrag (oder das Auftragssystem) von Abb. 11.11 in der Funktionseinheit i = R gerade gleich 4. Dabei muß nicht vorausgesetzt werden, dass die Teilaufträge durch eine Folge von Handlungen realisiert werden (a_2 und a_3 können z.B. kollateral in S und T bearbeitet werden). Letzteres ist vielmehr als Spezialfall in unserem Ansatz enthalten.

Definition 11.9 *Die Zahl der Teilaufträge, die bei Erledigung eines Auftrages an eine Funktionseinheit i gegeben wird, heißt* Besuchszahl *(visit ratio) v_{ij} des Auftrages j in der Funktionseinheit i. Später werden wir die Besuchszahl statt auf eine Funktionseinheit auch auf einen Funktionseinheittyp beziehen.*

Die Ablaufdarstellung des Workflow-Beispiels (Abb. 4.35 und 11.12) kann auch hier als Beispiel

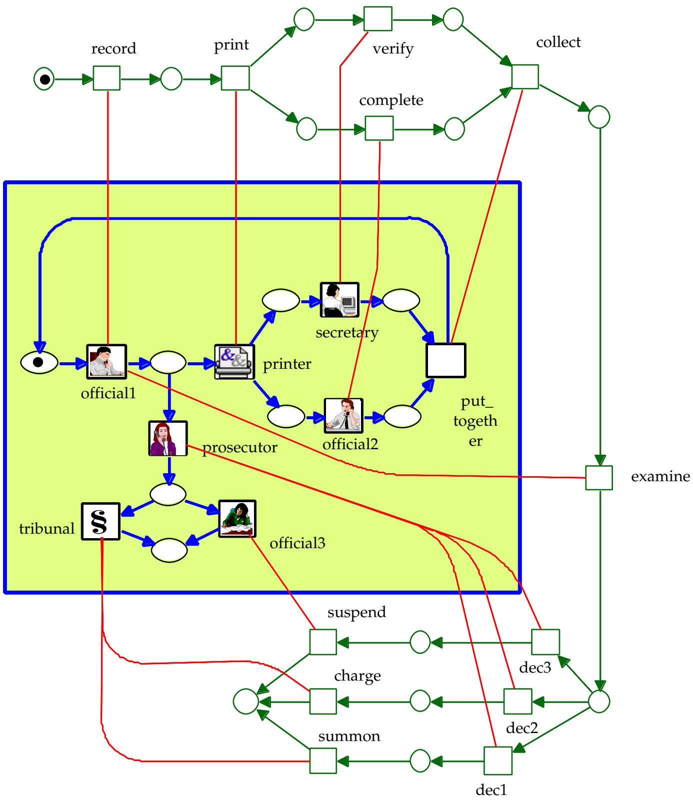


Abbildung 11.12: Workflow: Beispiel für Besuchszahl

herangezogen werden: die Funktionseinheit *official1* wird von den beiden Teilaufträgen *record* und *examine* besucht.

11.3 Das Gesetz von Little

Die Untersuchung der Bearbeitung zusammengesetzter Aufträge in einem System, das aus Funktionseinheiten zusammengesetzt ist, führte im Abschnitt 11.2 auf folgendes Bild : Aufträge strömen in das System ein und durchwandern es, wobei jeder Besuch an einer Funktionseinheit der Bedingung eines Teilauftrages entspricht; die Aufträge können sich auch in nebenläufige Teilaufträge aufspalten. Das zeitliche Verhalten wird durch die Verweilzeiten in den Funktionseinheiten bestimmt. Bevor wir diesen Auftrags„verkehr" im System quantitativ analysieren, werden wir sehr wichtige Zusammenhänge zwischen den drei grundlegenden Beschreibungsgrößen Füllung, Durchsatz und Verweilzeit (vgl. Def. 11.6 und 11.7) studieren, denen schon das Beispiel am Ende des Abschnitts 11.2 gegolten hat : *Littlesche Formeln*.

Dazu betrachten wir eine Funktionseinheit F, der Aufträge zugehen, in einem Intervall $(t_1\,;\,t_2)$ $t_2 > t_1$. In einem Zeitpunkt t_0 $(t_0 \leq t_1)$ sei die Funktionseinheit leer. Es sei für $t \geq t_0$

- $n_z(t)$: die Zahl der Aufträge, die in $(t_0;t($ ihre Verweilzeit in F begonnen haben.
- $n_a(t)$: die Zahl der Aufträge, die in $(t_0;t($ ihre Verweilzeit in F beendet haben.

Die Füllung ist

$$f(t) = n_z(t) - n_a(t) \geq 0 \tag{11.6}$$

Durch Rückgriff auf die Verweilzeitdefinition haben wir Abbruch oder Verdrängung von Aufträgen ausgeschlossen. Wir können allerdings die Gültigkeit unserer Überlegung jederzeit auf diese Fälle ausdehnen, wenn wir im Fall von Abbruch oder Verdrängung den Auftrag durch den bis zum Abbruch bzw. bis zur Verdrängung erledigten Teilauftrag ersetzen können.

Definition 11.10 *Als* Zugang *von Aufträgen an die Funktionseinheit bezeichnen wir den Quotienten*

$$z(t_1, t_2) = \frac{n_z(t_2) - n_z(t_1)}{t_2 - t_1} \tag{11.7}$$

wobei $t_2 > t_1$ gilt. Als Angebot *im Intervall $(t_1 ; t_2)$, $(t_2 > t_1)$ bezeichnen wir die Bedienzeiten der in $(t_1 ; t_2)$ zugehende Aufträge, geteilt durch die Dauer des Intervalls.*

Als Durchsatz von F bezeichnen wir natürlich

$$d(t_1, t_2) = \frac{n_a(t_2) - n_a(t_1)}{t_2 - t_1} \tag{11.8}$$

Wir betrachten nun ein einfaches operationales Modell und nachfolgend ein stochastisches Modell.

Einfaches operationales Modell

Wir nehmen zusätzlich an, dass die Funktionseinheit in den Zeitpunkten t_1 und t_2 leer ist. Dann ist

$$n_z(t_2) - n_z(t_1) = n_a(t_2) - n_a(t_1) \tag{11.9}$$

d.h. alle Aufträge, die der Funktionseinheit in $(t_1 ; t_2)$ zugegangen sind, sind vor ihr in $(t_1 ; t_2)$ auch erledigt worden.

Allgemein bezeichnet man den Umstand, dass in einem Betrachtungsinterval ebenso viele Aufträge zugehen wie abgehen, als *Flußgleichgewicht* (flow balance). Es ist dann

$$z(t_1, t_2) = d(t_1, t_2) \tag{11.10}$$

was natürlich insbesondere in unserem noch spezielleren Fall gilt. Abbildung 11.13 zeigt beispielhaft den Verlauf von $n_z(t)$ und $n_a(t)$ unter diesen Voraussetzungen. Der Mittelwert der Füllung in$(t_1\ ;\ t_2)$ ist

$$\overline{f}(t_1, t_2) = \frac{1}{t_2 - t_1} \int_{t_1}^{t_2} f(t)dt \tag{11.11}$$

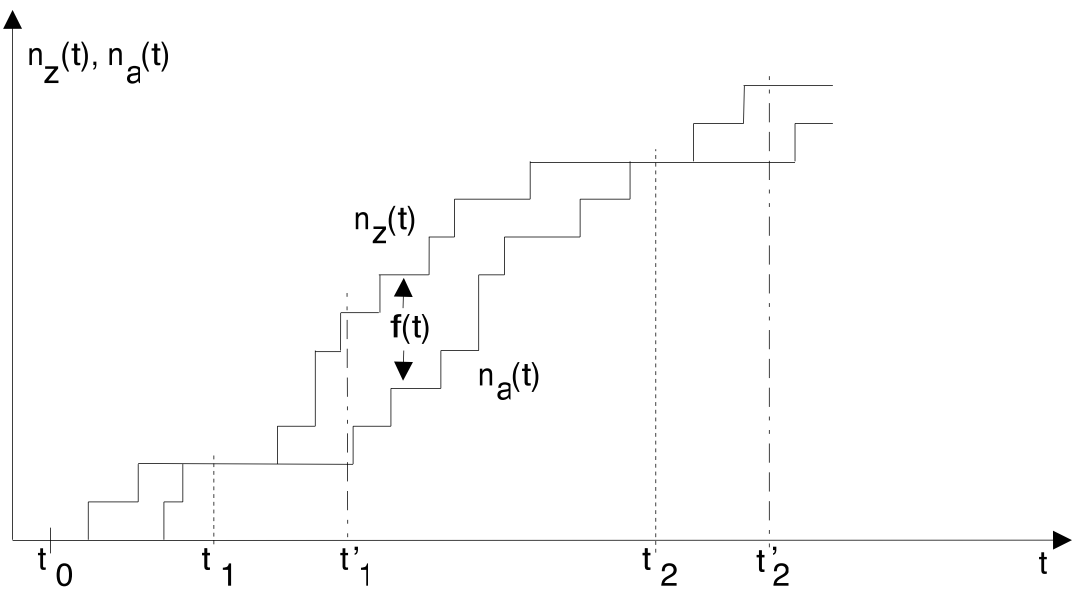


Abbildung 11.13: zur Erläuterung der Littelschen Formel : einfaches operationales Modell

Der Zugang und Durchsatz ist

$$z(t_1, t_2) = d(t_1, t_2) = \frac{n_a(t_2) - n_a(t_1)}{t_2 - t_1} \tag{11.12}$$

Schließlich ist die mittlere Verweilzeit ebenfalls in den benutzten Größen definierbar, da der i-te seit t_0 eintreffende Auftrag mit seiner Verweilzeit y_i einen Beitrag $1 \cdot y_i$ zum Füllungs-Zeit-Integral leistet :

$$\overline{y}(t_1, t_2) = \frac{1}{n_a(t_2) - n_a(t_1)} \cdot \sum_{i=n_z(t_1)+1}^{n_z(t_2)} y_i = \frac{1}{n_a(t_2) - n_a(t_1)} \int_{t_1}^{t_2} f(t)dt \tag{11.13}$$

Also gilt in dem einfachen operationalen Modell (unter Verwendung von 11.11, 11.12, 11.13)

$$\overline{f}(t_1, t_2) = z(t_1, t_2) \cdot \overline{y}(t_1, t_2) \tag{11.14}$$

$$\overline{f}(t_1, t_2) = d(t_1, t_2) \cdot \overline{y}(t_1, t_2) \tag{11.15}$$

d.h. das zeitliche Mittel der Füllung in $(t_1 ; t_2)$ ist gleich dem Produkt des Zugangs bzw. Durchsatzes in $(t_1 ; t_2)$ und der mittleren Verweilzeit der in $(t_1 ; t_2)$ in der Funktionseinheit verweilenden Aufträge. 11.14 und 11.15 sind Littlesche Formeln unseres Modells.

Es sei noch folgendes bemerkt :

- Die Begriffe ,,Aufträge", ,,erledigen" und ,,Funktionseinheit" sind nicht erforderlich zur Ableitung der Formeln 11.14 und 11.15. Sie beschreiben die für unsere Problematik bedeutsame Deutung des Modells und der in ihm bestehenden Gesetze.
- Wenn man die Voraussetzung $f(t_1) = f(t_2) = 0$ fallen läßt, (vgl. die Punkte t_1' und t_2' in Abbildung 11.13) entsteht i.a. ein Fehler in 11.12 (Zugang nicht notwendigerweise so groß wie der Durchsatz) und in 11.13 (in $(t_1 ; t_2)$ können Aufträge fertig werden, deren Verweilzeit im Füllungs-Zeit-Integral nur teilweise erfaßt ist, da sie schon vor t_1' eingetroffen sind; andererseits werden in t_2' vorhandene Aufträge erst nach t_2' fertig, liefern Beiträge zum Füllungs-Zeit-Integral, nicht aber zum Durchsatz). Beide Fehler wirken gegensinnig. Ist der resultierende Fehler klein relativ zum Zugang / Durchsatz bzw. zum Füllungs-Zeit-Integral, dann gelten auch die Formeln mit guter Näherung. Zur exakten Analyse siehe [Kow80].
- Betrachten wir insbesondere eine ,,einfache" Funktionseinheit (Def. 11.6) : hier ist die Füllung eine binäre Größe und die mittlere Füllung ist gleich dem Zeitanteil, in dem diese Funktionseinheit belegt ist. Dieser ist aber nach 11.4 gleich der Auslastung. Also gilt für eine einfache Funktionseinheit 11.14 bzw. 11.15 in der Form

$$\rho(t_1, t_2) = z(t_1, t_2) \cdot \overline{y}(t_1, t_2) = d(t_1, t_2) \cdot \overline{y}(t_1, t_2) \tag{11.16}$$

Stochastisches Modell

Wir setzen jetzt die Füllung $F(t)$, die Verweilzeit des i-ten Auftrages Y_i, Zugang $Z(t_1, t_2)$ und Durchsatz $D(t_1, t_2)$ als Zufallsvariable mit beschränkten, von t, t_1, t_2, i unabhängigen Erwartungswerten voraus : $E[F]$, $E[Y]$, $E[D]$, $E[Z]$. Daraus folgt, dass im Mittel so viele Aufträge zugehen wie erledigt werden,

$$E[Z] = E[D] \tag{11.17}$$

Wegen der Zeitunabhängigkeit der Erwartungswerte muß eine Beziehung zwischen ihnen für *beliebige* Betrachtungsintervalle $(t_1^{'} ; t_2^{'})$ gelten. Es sei nun in $t_1^{'}$ und $t_2^{'}$ die Füllung nicht notwendigerweise null. Dann enthält 11.13 den bereits beschriebenen Fehler, dessen Betrag am oberen Rand jeweils

$$\sum_{j=1}^{F} RY_j$$

ist (RY Restverweilzeit, F Füllung am oberen bzw. unteren Rand).

Als *beliebiges* Betrachtungsintervall untersuchen wir nun $t_2 - t_1 \rightarrow \inf$. Da mit $E[F]$ und $E[Y]$ auch

$$E[\sum_{j=1}^{F} RY_j]$$

beschränkt ist, verschwinden mit $t_2 - t_1 \rightarrow \inf$ auch die mittleren Randfehler in 11.13, und es gilt

$$E[F] = E[Z] \cdot E[Y] \qquad (11.18)$$

$$E[F] = E[D] \cdot E[Y] \qquad (11.19)$$

nach unseren Voraussetzungen für *jedes* Intervall. 11.18 und 11.19 sind Littlesche Formeln.

Bemerkung :
Wieder gilt für einfache Funktioneinheiten

$$\rho = E[Z] \cdot E[Y] = E[D] \cdot E[Y] \qquad (11.20)$$

Dabei ist ρ nun als

$$\rho = \frac{E[D]}{C} \qquad (11.21)$$

zu definieren.
Fur den – sehr speziellen – Fall, dass die Funktionseinheit die Aufträge in derselben Reihenfolge abgehen läßt, weil sie zugegegangen sind, bietet 11.18 eine sehr anschauliche Deutung : in der mittleren Verweilzeit E[Y] eines Auftrags regeneriert sich die mittlere Füllung (es treffen nämlich im Mittel $R[Z] \cdot E[Y]$ Aufträge ein.

11.4 Systeme mit großer Füllung

Wir entwickeln nun eine wichtige Gesetzmäßigkeit, nämlich das allgemeine Durchsatzgesetz. Es bestimmt die Durchsätze der Funktionseinheiten bzw. Funktionseinheitentypen des Systems abhängig von der Belastung, und die Existenz von Grenzdurchsätzen erlaubt es, auf die Auslastung der Funktionseinheiten(typen) zu schließen und auf den Systemgrenzdurchsatz. Dabei betrachten wir zunächst ein *operationales* Modell. Entsprechende Gesetze gelten auch für ein *stationäres stochstisches Modell* bei dem (wie beim Littleschen Gesetz) die Mittelwerte durch Erwartungswerte ersetzt sind (siehe [JV87]).

Für offene Systeme dürfen wir verlangen, dass es in t_1 und t_2 ($t_2 > t_1$) leer ist. Wir betrachten den Auftragsverkehr im Intervall (t_1 ; t_2). Das Fuktionseinheitenystem FES umfasse m Funktionseinheiten. In (t_1 ; t_2) mögen ihm n Aufträge zugehen und erledigt werden, von denen der Auftrag k in der Funktionseinheit i zu v_{ik} Teilaufträgen führt ($k \in \{1 \ldots n\}, i \in \{1 \ldots m\}$). Dabei ist v_{i_k} natürlich die schon aus Def. 11.9 bekannte *Besuchszahl* des Auftrages k in der Funktionseinheit i. Dort verbringt er *Verweilzeiten* $y_{ikl}(l \in \{1 \ldots v_{ik}\}$ und erzeugt einen Teildurchsatz

$$d_{ik}(t_1, t_2) = \frac{v_{ik}}{t_2 - t_1} \tag{11.22}$$

.

Für diese Feststellung müssen wir nicht voraussetzen, dass der Auftrag sequentiell abgearbeitet wird, vielmehr können Teilaufträge kollateral abgearbeitet werden. Ebenfalls kann nicht erwartet werden, dass je Teilauftrag nur eine Funktionseinheit beauftragt wird, vielmehr darf der Teilauftrag gleichzeitig in mehreren Funktionseinheiten verweilen! Natürlich ist der Gesamtdurchsatz der Funktionseinheit i

$$d_i(t_1, t_2) = \sum_{k=1}^{n} d_{ik}(t_1, t_2) = \frac{1}{t_2 - t_1} \sum_{k=1}^{n} v_{ik} \tag{11.23}$$

oder, indem wir die mittlere Besuchszahl $\overline{v}_i$ je Auftrag und den Systemdurchsatz $d_{FES}(t_1, t_2)$ einführen, den die n Aufträge hervorbringen

$$d_i(t_1, t_2) = \frac{n}{t_2 - t_1} \cdot \frac{1}{n} \sum_{k=1}^{n} v_{ik} \tag{11.24}$$

$$d_i(t_1, t_2) = d_{FES}(t_1, t_2) \cdot \overline{v}_i \tag{11.25}$$

11.23 bzw. 11.25 ist das *allgemeine Durchsatzgesetz*. Besteht nun in (t_1 ; t_2) für die Funktionseinheit i ein Grenzduchsatz c_i, definiert in den in (t_1 ; t_2) anfallenden Teilaufträgen (Besuchen), so ist die Auslastung von i nach Def. 11.7

$$\rho_i = \frac{d_i(t_1, t_2)}{c_i} = \frac{1}{c_i \cdot (t_2 - t_1)} \sum_{k=1}^{n} v_{ik} = \frac{d_{FES}(t_1, t_2) \cdot \overline{v}_i}{c_i} \tag{11.26}$$

Definition 11.11 *In einem System aus m Funktionseinheiten heißen die Funktionseinheiten mit der Auslastung*

$$\rho_{max} = max\{\rho_i\}, i \in \{1 \dots m\}$$

(Verkehrs-)Engpässe *(bottlenecks) des Systems.*

Sollen alle in (t_1 ; t_2) eingehenden Aufträgen bis t_2 erledigt sein, dann muß nach Definition des Grenzdurchsatzes

$$\rho_{max} \leq 1 \tag{11.27}$$

sein.

Sind die Funktionseinheiten ,,einfach" (Def 11.6), dann reduziert sich 11.26 zu

$$\rho_i = \frac{1}{t_2 - t_1} \sum_{k=1}^{n} x_{ik} = d_{FES}(t_1, t_2) \cdot \overline{x}_i = \frac{d_i(t_1, t_2)}{\overline{v}_i} \cdot \overline{v}_i \cdot \overline{b}_i = d_i(t_1, t_2) \cdot \overline{b}_i \tag{11.28}$$

wobei $\overline{x}_i$ die mittlere Belegtzeit je Auftrag ist und – gültig bei einfachen Instanzen – $\overline{b}_i$ die mittlere Bedienzeit je Teilauftrag; 11.27 reduziert sich zu

$$max\{\sum_{k=1}^{n} x_{ik}\} \leq t_2 - t_1, \tag{11.29}$$

wobei x_{ik} nach Def. 11.8 die Belegtzeit des Knotens i durch den Auftrag k ist.

Beispiel 11.12

Im Zeitintervall (0s; 50s) werde ein System von Funktionseinheiten $F_1, \dots, F_4$ mit 4 passenden Auftragssystemen $AS_1, \dots, AS_4$ beschickt, wobei AS_1 vom Typ 1 und AS_2, AS_3, AS_4 vom Typ 2 in Abb. 11.14 sind.

Von den Funktionseinheiten nehmen wir an, dass sie elementare Wartesysteme mit *einer* Bedieneinheit sind; entsprechend ist ein Auftragstyp zusätzlich durch die an den Teilaufträgen in Klammern angegebenen Bedienzeiten gekennzeichnet. Die Ankunftszeit des Auftragssystems vom Typ 1 sei $g_1 = 0s$, derjenigen vom Typ 2 seien $g_2 = 0s$, $g_2^{'} = 3s$, $g_2^{''} = 6s$. Damit ergeben sich die folgenden Belegtzeiten (vgl. Def. 11.8) der Bedieneinheit unter Berücksichtigung der Zahl der Teilaufträge je Typ :

Bedieneinheit	Belegtzeit Typ 1	Belegtzeit Typ 2	gesamte Belegtzeit	Auslastung
F_1	$5 + 2$	4	$7 + 3 \cdot 4 = 19s$	$0,38$
F_2	$3 + 2$	$6 + 5$	$5 + 3 \cdot 11 = 38s$	$0,76$
F_3	7	7	$7 + 3 \cdot 7 = 28s$	$0,56$
F_4	8	1	$8 + 3 \cdot 1 = 11s$	$0,22$

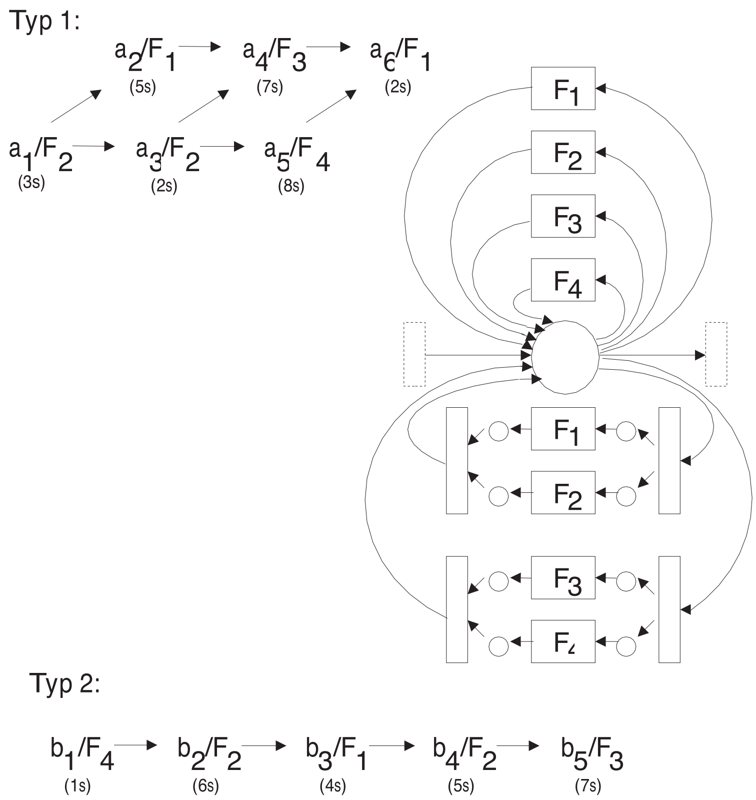


Abbildung 11.14: Auftragssysteme und System von Funktionseinheiten in den Beispielen 11.12, 11.14

Dabei berechnet sich die Auslastung als gesamte Belegtzeit pro Zeitintervall (50s), vgl. 11.4. Die Funktionseinheit F_2 ist Verkehrsengpass. Mit Hilfe der Besuchszahlen ergeben sich folgende Durchsätze für die Teilaufträge :

Bedieneinheit	Besuchszahlen Typ 1	v_i Typ 2	Durchsatz (in Teilaufträgen)	Grenzdurchsatz (in Teilaufträgen)
F_1	2	1	$5/50 = 0,1/s$	$5/19 = 0,26/s$
F_2	2	2	$8/50 = 0,16/s$	$8/38 = 0,21/s$
F_3	1	1	$4/50 = 0,08/s$	$4/28 = 0,14/s$
F_4	1	1	$4/50 = 0,08/s$	$4/11 = 0,36/s$

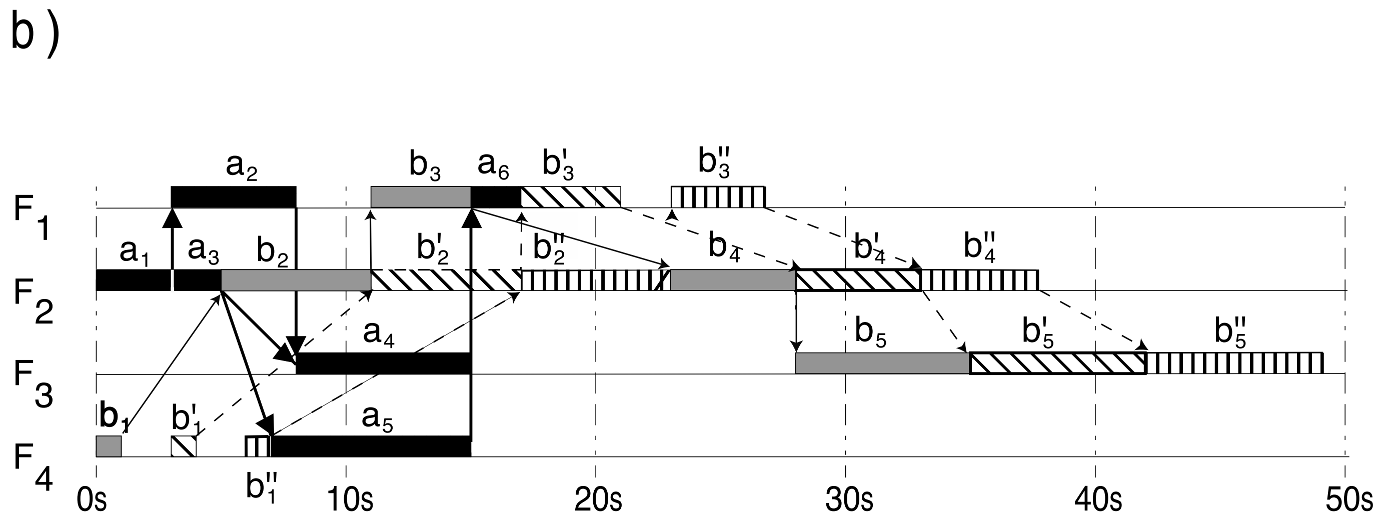


Abbildung 11.15: Synchroner Prozeß in a) Kausalnetz– und in b) Balkendiagramm–Darstellung

Abb. 11.15 zeigt einen möglichen Verlauf der Auftragsbearbeitung als synchronen Prozess (Def. 1.3.10 c) in einer Darstellung als Kausalnetz und als Balkendarstellung (vgl. Abb. 1.3.12) (Die zu AS_1 gehörenden Teilaufträge sind schwarz hervorgehoben; die gestrichelte Linie zeigt AS_2).

Der Prozess illustriert, weshalb 11.27 / 11.29 keine hinreichende Bedingung für die Erledigbarkeit aller Aufträge im Intervall (t_1 ; t_2) ist; man vergrößere in Gedanken die Bedienzeiten von b_5, $b_5^{'}$, $b_5^{''}$.

Offenbar ist nur

$$\max_{1 \le k \le n} \{g_k + y\} \le t_2 \qquad (11.30)$$

(g_k Ankunftszeit, y_k Verweilzeit) für den Auftrag hinreichend. Bei dem gewählten Prozess treten Wartezeiten z. B. vor den Bedienzeiten von a_5, b_2, b_4 und $b_5^{'}$ auf (siehe insbesondere die Zeiten in Abb. 11.15 a)).

Aus der Definition 11.11 mit den Aussagen 11.23 und 11.27 folgt eine Schranke für den Durchsatz des Systems FES. Es gilt nmlich nach 11.23/ 11.27 für den Verkehrsengpaß VE

$$\rho_{VE} = \frac{1}{c_{VE}(t_2 - t_1)} \sum_{k=1}^{n} v_{VEk} = max(\rho_i) \leq 1$$

bzw. entsprechend

$$\frac{d_{FES}(t_1, t_2) \cdot \overline{v}_{VE}}{c_{VE}} \leq 1 \tag{11.31}$$

$$d_{FES}(t_1, t_2) \leq \frac{c_{VE}}{\overline{v}_{VE}} \tag{11.32}$$

und für eine einfache Instanz als Verkehrsengpaß

$$d_{FES}(t_1, t_2) \leq \frac{1}{\overline{v}_{VE} \cdot \overline{b}_{VE}} \tag{11.33}$$

Das Beispiel 11.12 hat gezeigt, dass es eine i.a. niedrigere Schranke für den Systemdurchsatz als 11.27 (und damit 11.32/11.33) gibt, nämlich 11.30, was allerdings weit schwerer auszuwerten ist, da es die Kenntnis der Verweilzeit voraussetzt. Das Beispiel gibt zugleich einen Hinweis, wann 11.27 eine zu optimistische Abschätzung ist, nämlich wenn Freizeiten der Engpässe anfallen. Das ist der Fall

- wenn die Systemfüllung f_{FES} zu klein ist
- wenn – auch bei großer Systemfüllung – am Anfang oder Ende des Intervalls (t_1 ; t_2) von den Aufträgen überwiegend eine andere Funktionseinheit als ein Engpaß beaufschlagt wird. Typisches Beispiel : Durchführung eines „Benchmarktests“, in dem auf ein zu Beginn und zu Ende des Tests leeres System eine Menge von Aufträgen gegeben wird, für die – üblicherweise – der Zentralprozessor Engpaß ist, die aber mit einer Ausgabephase enden, die den sonst wenig ausgelasteten Drucker beaufschlagt.

11.27 (und damit 11.32/11.33) sind gute Abschätzungen, wenn (t_2 - t_1) groß gegen die Verweilzeit eines einzelnen Auftrags allein im System ist, und die Füllung groß ist. Wann eine Füllung „groß“ ist, greifen wir weiter unten auf. Zunächst definieren wir

Definition 11.13 *Als* Grenzdurchsatz *(troughput limit)* c_{FES} eines Systems FES *aus m Funktionseinheiten mit den Grenzdurchsatzsätzen c_i, mittleren Besuchszahlen $\overline{v}_i(i \in \{1 \dots m\})$ bezeichnen wir*

$$c_{FES} = min\{\frac{c_i}{\overline{v}_i}\} = \frac{c_{VE}}{\overline{v}_{VE}} \tag{11.34}$$

Im Falle mehrerer Auftragstypen läßt sich der Grenzdurchsatz je Auftragstyp angeben, wahrscheinlich unter Hemmung der anderern Auftragstypen, oder auch für eine Mischung, etwa dem vorgegebenen Durchsatz entsprechend. Dann ist für c_{VE} und $\overline{v}_{VE}$ ein mit den Durchsatzanteilen gewichtetes Mittel einzusetzen.

Beispiel 11.14 Wir nehmen Beispiel 11.12 wieder auf.

Funktionseinheit 2 ist Engpaß VE, $\rho_2 = \rho_{VE} = 0,76$. Die (hier ja nicht wirksame) Schranke c_{FES} (vgl. 11.34) ergibt

$$c_{FES} = \frac{c_{VE}}{\overline{v}_{VE}} = \frac{0,21 \text{ Teilaufträge / s}}{2 \text{ Teilaufträge / Auftrag}} = 0,105 \text{ Aufträge / s}$$

da die mittlere Besuchszahl der Funktionseinheit

$$\overline{v}_{VE} = \frac{1}{n}\sum_{k=1}^{n} v_{VEk} = \frac{1}{4}(2+2+2+2) = 2$$

und ihr Grenzdurchsatz, wie schon ermittelt,

$$c_{VE} = 0,21 \text{ Teilaufträge / s}$$

ist.

Also verbietet der Engpaß, Funktionseinheit 2, mehr als 5,25 Aufträge in 50s durch das System „durchzusetzen". Abb. 11.15 zeigt aber, dass im kurzen Intervall t_2 - $t_1 = 50s$ die Füllung so niedrig bleibt, dass bereits 4 Aufträge die Obergrenze bilden.

Die bisherige Überlegungen lassen sich unmittelbar auf den Fall übertragen, dass anstelle einer Funktionseinheit i ein Element eines Typs i von Funktionseinheiten beauftragbar ist. Damit läßt sich der Typendurchsatz, die Typauslastung und der Engpaß nach obigem Vorbild ermitteln, worauf wir verzichten.

Aufgabe 11.15 (Verkehrsengpass, Grenzdurchsatz, Durchsatzgesetz)

Gegeben seien ein Auftragssystem AS_1 vom Typ A (Ankunftszeit $t = 0\,$s) und zwei Auftragssysteme AS_2 und AS_3 vom Typ B (Ankunftszeiten $t = 0\,$s und $t = 3\,$s).

a) Berechnen Sie die Auslastung jeder der vier Funktionseinheiten F_1 bis F_4 und damit den Verkehrsengpass. Berechnen Sie die Grenzdurchsätze. Das betrachtete Zeitintervall sei $[t_1, t_2] = [0\,\mathrm{s}, 30\,\mathrm{s}]$.

b) Geben Sie einen möglichen Ablauf in Balkendarstellung an.

c) Wie groß ist der Grenzdurchsatz des Systems?

d) Berechnen Sie die Mittelwerte der Füllung, der Verweilzeit und den Durchsatz in Teilaufträgen für die vier Funktionseinheiten.

e) Zeigen Sie für alle vier Funktionseinheiten das allgemeine Durchsatzgesetz.

Zwei Grenzfälle des Verhaltens eines Systems unter Belastung sind verhältnismäßig einfach zu berechnen und zu messen, und diese sind die mittlere Bedienzeit und der Grenzdurchsatz.

Zur Untersuchung betrachten wir ein System S in einem Zeitraum, der so groß ist, dass die in 11.3 beschriebenen Randeffekte vernachläßigt werden können. In diesem Zeitraum möge die mittlere Verweilzeit $\overline{y}_S$ betragen. Im stochastischen Modell mögen zeitunabhängige Erwartungswerte $E[F_S]$, $E[D_S]$, $E[Y_S]$ bestehen. Es gilt Littels Formel (11.22). Die Aufträge sind als Auftragssysteme beschrieben. Das System kann offen oder geschlossen sein. Die mittlere Verweilzeit eines Auftrages im System allein, nach Definition seine Bedienzeit, sei $\overline{b}_S$ bzw. $E[B_S]$; sie ist nach den Formeln in 5.3.4 berechenbar, sehr oft auch einfach meßbar. Der Grenzdurchsatz ergibt sich nach 11.34 als

$$c_s = \frac{c_{VE}}{\overline{v}_{VE}} \; bzw. \; \frac{c_{VE}}{E[V_{VE}]}$$

Im allgemeinen wird die Anwesenheit eines weiteren Auftrags im System die Verweilzeit erhöhen (wir sehen von dem anomalen Beispiel ab). Als ,,klein" bezeichnen wir eine Füllung des Systems, bei der die Verweilzeit nicht wesentlich von der Bedienzeit verschieden ist. Nach Littles Formel 11.34 gilt dann

$$d_S = \frac{\overline{f}_S}{\overline{y}_S} \approx \frac{\overline{f}_S}{\overline{b}_S}, \tag{11.35}$$

d.h. der Durchsatz wächst bei kleinen Füllungen proportional dem Mittelwert der Füllung. Es sei darauf hingewiesen, dass sehr wohl die Füllung einen kleinen Mittelwert haben kann, ohne dass sich im System je die Verhältnisse von kleiner Füllung ergeben, nämlich wenn in einen langen Zeitraum, in dem das System frei ist, kurze Intervalle großer Füllung fallen. Natürlich gilt dann unsere Überlegung nicht, weil die Aufträge sich erheblich behindern können. Beschränkt man sich auf geschlossene Systeme, so braucht man diese Einschränkung nicht auszusprechen; insbesondere tritt hier bei $f_S = 1$ keine Behinderung auf und 11.34 gilt exakt.

Bei großer Füllung des Systems wird der Durchsatz durch den Verkehrsengpaß begrenzt, und wenn dieser voll ausgelastet wird, folgt

$$\overline{y}_S = \frac{\overline{f}_S}{d_S} \approx \frac{\overline{f}_S}{c_S}, \tag{11.36}$$

d.h. die mittlere Verweilzeit wächst bei großen Füllung proportional dem Mittelwert der Füllung.

Ein konsistentes Bild ergibt sich nur, wenn der Charakter der Aufträge durch die Füllung unbeeinflußt bleibt. Wir verlangen daher faire Bedienstrategien, um zu verhindern, dass im stark gefüllten System eine Veränderung der Auftragsmischung durch unbegrenztes Warten eintritt.

Offenbar zeigt das System den folgenden asymptotischen Verlauf von Durchsatz und mittlerer Verweilzeit als Funktion der mittleren Füllung (Abb. ??). Der Schnittpunkt der beiden Asymptoten liegt bei der mittleren Füllung

$$\overline{f}^*_S = c_S \cdot \bar{b}_S. \tag{11.37}$$

Dieser Wert heißt *Sättigungsfüllung*. Er erlaubt abzuschätzen, was ,,kleine" und ,,große" Füllung ist, und ist zugleich, da im geschlossenen System der Durchsatz für einen Auftrag allein

$$d_S(f_S = 1) = \frac{1}{b_S}$$

ist, gleich dem Verhältnis von Grenzdurchsatz zum Durchsatz im mit nur einem Auftrag belasteten geschlossenen System.

Das graphische Bild der beiden Funktionen $E[D_S]$, $E[Y_S]$ folgt umso besser den Asymptoten, je geringer die Streuung der Bedienzeiten ist. Für konstante Bedienzeiten fällt es mit den Asymptoten zusammen. Bei streuenden Bedienzeiten fällt der Durchsatz niedriger und die Verweilzeit höher aus, wenn das Verhalten des Systems durch mehrere Engpässe bestimmt wird oder mehrere Instanzen fast so hoch wie der Engpaß ausgelastet sind.

Wir haben bisher angenommen, dass die Verweilzeit – solange keine Behinderung eintritt – und der Grenzdurchsatz von der mittleren Füllung unabhängig sind. Das ist nicht notwendigerweise so, wenn ein Overhead vom System aufzubringen ist. Wir unterscheiden verschiedene Fälle.

Der Overhead nehme einen füllungsabhängigen Anteil der Leistung des Systems ein. So wird etwa eine periodische Unterbrechungsbehandlung im Zentralprozessor die mittleren Bedienzeiten um einen konstanten Faktor strecken und – ist er bereits Engpaß – den Grenzdurchsatz verkleinern; er kann aber auch erst durch die Bürde des Overheads Engpaß werden.

Nimmt der Overhead proportional oder überproportional zur mittleren Füllung zu, etwa durch periodischen Aufruf der Planungskomponente des Betriebssystems mit einer der Füllung proportionalen oder überproportionalen Laufzeit, so nimmt der Grenzdurchsatz der beanspruchten Funktionseinheit für die Aufträge des Systems hyperbolisch ab; ab einer gewissen mittleren Füllung wird diese Funktionseinheit Engpaß und bestimmt das asymptotische Verhalten.

Abbildung 11.17 zeigt die beiden Overheadfälle (die sich natürlich überlagern können). Man beachte, dass die d_S und $\overline{y}_S$-Kurven über Littlesche Formel miteinander zusammenhängen.

Aufgabe 11.16 (Sättigungsfüllung bei konstanter Bedienzeit)

In einem System von 3 Funktionseinheiten sei die Bedienzeit aller Aufträge gleichmäßig gleich 0.5. Sind mehr als 3 Aufträge zu- aber noch nicht abgegangen, dann warten diese in einem Wartepool, dessen Füllung zur Systemfüllung beiträgt. Bei konstanter Füllung f enthält dieser also $f - 3$ Aufträge.

Zeichnen Sie die Diagramme von Abbildung 11.16 für diesen Fall. Wie groß ist die Sättigungsfüllung? Welche Werte nimmt die Formel $\overline{f}_S^* = c_S \cdot \overline{b}_S$ in diesem Fall an?

Besonders durchsichtige Verhältnisse ergeben sich für das Grenzverhalten von geschlossenen Systemen. Hier ist die Füllung eine konstante Größe, und die Unterscheidung von Füllung und mittlerer Füllung entfällt. Wir engen unsere Untersuchung auf den folgenden Fall ein.

Beispiel 11.17 Ein System S bestehe aus m Instanzen i ($i \in \{1 \ldots m\}$), gekennzeichnet durch Bedienzeiten $E[b_i]$ und m_i Bedieneinheiten. Die Bedienstrategien seien produktiv und fair. Die Grenzdurchsätze sind folglich $c_i = \dfrac{m_i}{E[B_i]}$. Die Belastung bestehe aus f_S seriell ausgeführten Auftragssystemen mit (von den Bedienstrategien unabhängigen) Besuchszahlen $E[V_i]$. Da die Bedienstrategien produktiv sind, folgt

$$E[B_S] = E[Y(f_S = 1)] = \sum_{i=1}^{m} E[V_i] \cdot E[B_i]$$

Weiter ist

$$c_S = \frac{c_{VE}}{E[V_{VE}]} = \frac{m_{VE}}{E[V_{VE}] \cdot E[B_{VE}]} \tag{11.38}$$

und nach 11.37

$$f_S^* = c_S \cdot E[B_S] = \frac{m_{VE} \cdot \sum E[V_i] \cdot E[B_i]}{E[V_{VE}] \cdot E[B_{VE}]} \tag{11.39}$$

Die Auslastung des Knotens i ist wegen $E[D_i] = E[V_i] \cdot E[D_S]$

$$\rho_i = \frac{E[V_i] \cdot E[D_S]}{c_i} = \frac{E[V_i] \cdot E[B_i]}{m_i} \cdot E[D_S] \tag{11.40}$$

Damit läßt sich 11.39 auch schreiben als

$$f_S^* = \frac{\sum\limits_{i=1}^{m} m_i \dfrac{E[V_i] \cdot E[B_i]}{m_i}}{\dfrac{E[V_{VE}] \cdot E[B_{VE}]}{m_{VE}}} = \frac{\sum\limits_{i=1}^{m} m_i \rho_i}{\rho_{VE}} \tag{11.41}$$

Also läßt sich die Sättigungsfüllung auf die Auslastung der Instanzen zurückführen ! Im Sonderfall, dass alle Instanzen einfach sind ($m_i \equiv 1$), ist

$$f_S^* = \frac{\sum\limits_{i=1}^{m} \rho_i}{\rho_{VE}} \tag{11.42}$$

Falls alle Instanzen gleich ausgelastet sind, besagt unser Ergebnis, dass eine Füllung als klein anzusehen ist, die klein gegen die Zahl der unabhängig beauftragbaren Instanzen ist.

Im Sonderfall, dass alle Instanzen einfach sind, folgt auch ein einfaches Meßverfahren, mit dem sich die Kennwerte der Diagramme ermitteln lassen. Man schicke eine Menge von Aufträgen, die man für repräsentativ hält, seriell durch das System. Die mittlere Bedienzeit ergibt sich durch Messung der individuellen Bedienzeiten der Aufträge und Mittellung oder – bei n Aufträgen – durch Mittellung über die Summe aller Belegtzeiten x_i

$$\overline{b}_S = \frac{1}{n} \sum_{i=1}^{m} x_i$$

Die Belegtzeit der Instanz i gibt an, wie lange die Instanz i während des Meßintervalls, in dem n Aufträge bedient wurden, belegt war. Der Engpaß VE folgt aus $x_{VE} = max\ x_i$. Der Grenzdurchsatz beträgt

$$c_S = \frac{n}{x_{VE}} \tag{11.43}$$

und die Sättigungsfüllung

$$f_S^* = \frac{\sum\limits_{i=1}^{m} x_i}{x_{VE}} \tag{11.44}$$

Wir benutzen die Ergebnisse diese Beispiels zur Analyse einer Mehrprozessoranlage im Vergleich zu einer Einprozessoranlage gleiche Prozessordurchsatzes.

Beispiel 11.18 Der Auftragsverkehr in einer Rechenanlage sei im wesentlichen durch Rechen- und Transportphasen bestimmt. Ein Auftrag verlange im Mittel $E[V_{ZP}] = 100$ Rechenphasen, $E[B_{ZP}] = 30ms$, und 100 Transportphasen, die sich auf 4 Platten gleichmäßig verteilen,

$E[V_P] = 100$, $E[B_P] = 20ms$, $m_P = 4$. Es wird erwogen, anstelle des einen Prozessors 4 Prozessoren von je einem Viertel der Leistung ($E[B_{P'}] = 120ms$) einzusetzen.

Mit einem Prozessor ist $E[B_S] = E[V_{ZP}] \cdot E[B_{ZP}] + E[V_P] \cdot E[B_P]$

$$E[B_S] = 100 \cdot 30ms + 100 \cdot 20ms = 5s$$

Die Auslastung ist nach 11.40

$$\rho_i = \frac{E[V_i] \cdot E[B_i]}{m_i} \cdot E[D_S]$$

Sie ist maximal für den Prozessor, der also den Verkehrsengpaß VE ist.

$$c_S = \frac{c_{ZP}}{E[V_{ZP}]} = \frac{1}{30ms \cdot 100} = 0,33/s$$

$$f_S^* = 0,33 \cdot 5 = 1,65$$

(um diesen Faktor kann Mehrprogrammbetrieb den Durchsatz verbessern). In Abb. 11.18 ist der hier (geschlossenes System!) exakt bestimmbare Wert für $f_S = 1$ eingetragen ($\otimes$) und zusätzlich Schätzpunkte für größere Füllungen ($\circ$).

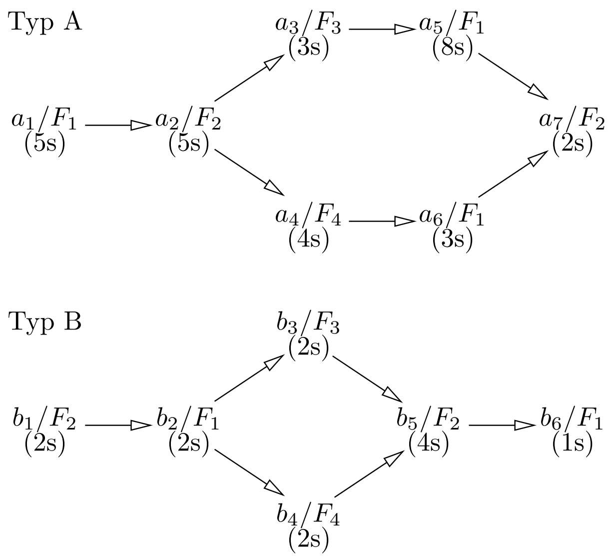
Bei Übergang auf 4 Prozessoren von je einem Viertel der Leistung folgt

$$E[B_{S'}] = 100 \cdot 120ms + 100 \cdot 20ms = 14s$$

c_{ZP} ist unverändert, ebenfalls bleibt der Prozessor Engpaß. Es folgt

$$f_S^{*}\text{'} = 0,33 \cdot 14 = 4,62$$

Das Ergebnis ist in Abb 11.18 eingetragen und zeigt, dass die Aufteilung auf vier Prozessoren bezüglich Durchsatz und Verweilzeit deutlich unterlegen ist. Ein Ausgleich ist bei großer Füllung möglich, allerdings um den Preis eines vergrößerten Hauptspeichers. Wenn der Betrieb der 4 Prozessoren einen vergrößerten Overhead erbringt, sieht der Vergleich noch ungünstiger aus.


Typ A
a_1/F_1 (5s)
a_2/F_2 (5s)
a_3/F_3 (3s)
a_4/F_4 (4s)
a_5/F_1 (8s)
a_6/F_1 (3s)
a_7/F_2 (2s)
Typ B
b_1/F_2 (2s)
b_2/F_1 (2s)
b_3/F_3 (2s)
b_4/F_4 (2s)
b_5/F_2 (4s)
b_6/F_1 (1s)

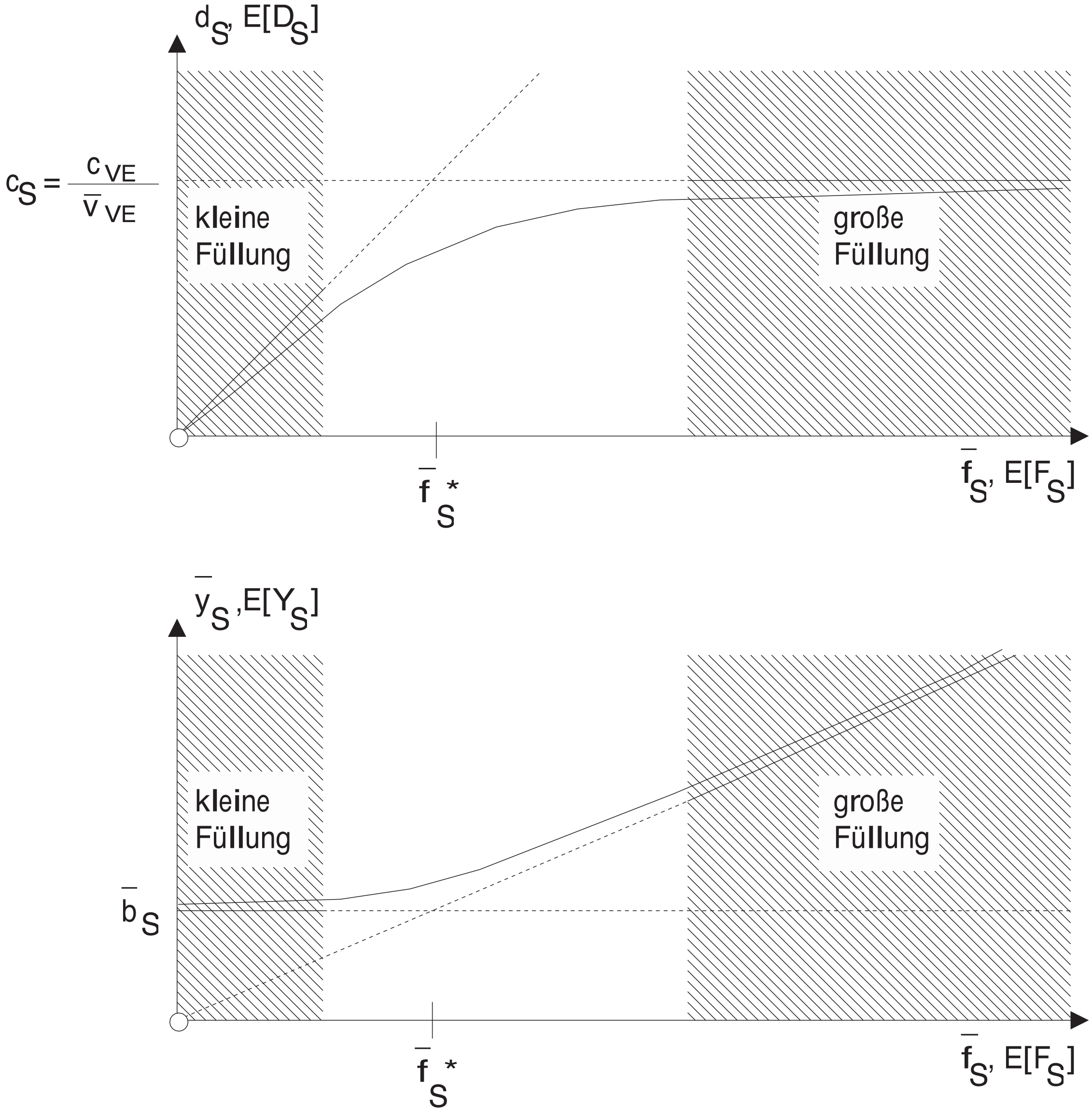


Abbildung 11.16: **Verhalten eines Systems mit Bedienzeit $\overline{b}_S(E[B_S])$ und Grenzdurchsatz $c_S \cdot \overline{f}_S^*$ ist die Sättigungsfüllung**

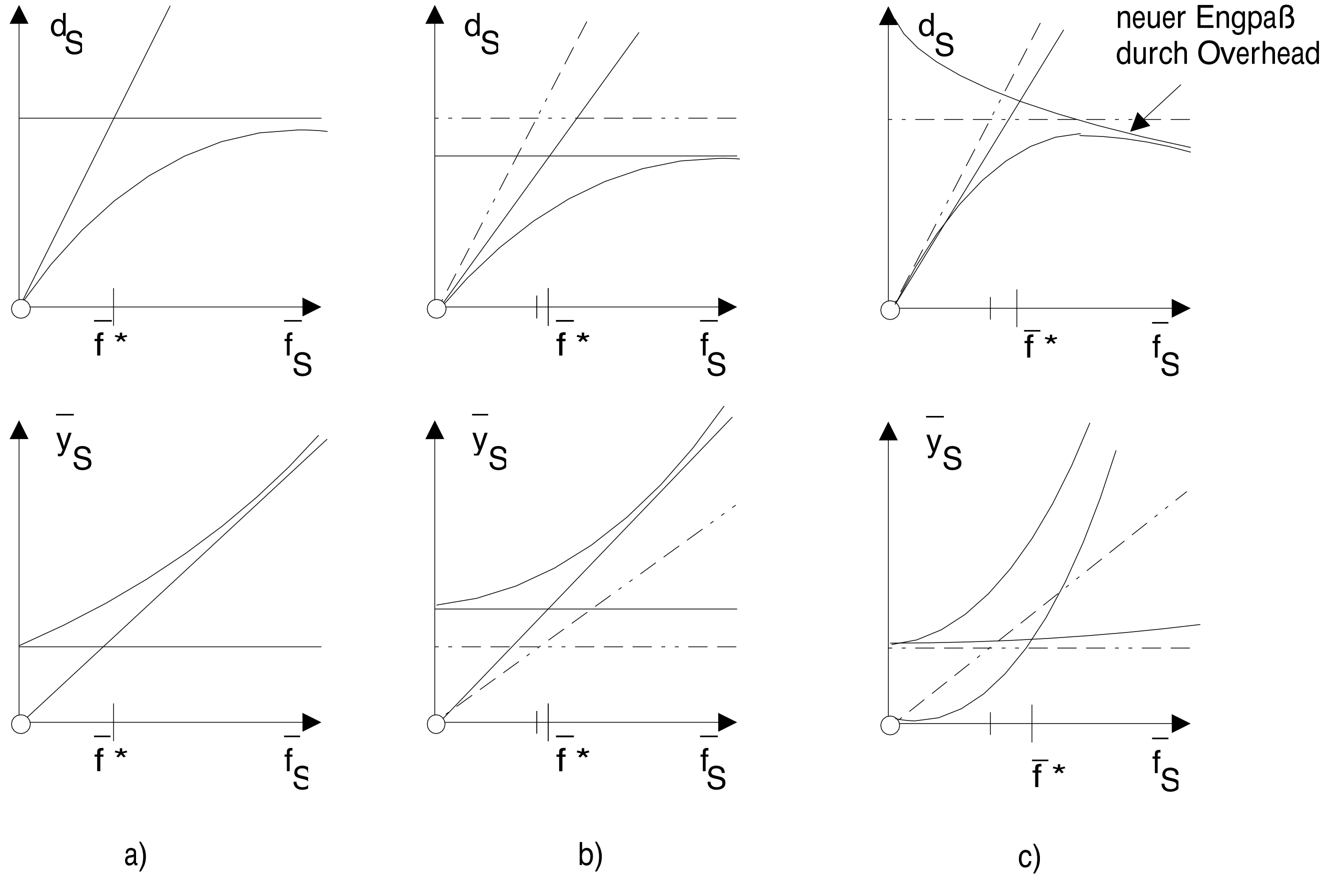


Abbildung 11.17: Idealer Verlauf von Durchsatz d_S und mittlerer Verweilzeit $\overline{y}_S$ (a), füllungsunabhängiger Overhead in Verweilzeit und Grenzdurchsatz (b), mir der Füllung wachsender Overhead, durch den bei großer Füllung ein anderer Engpaß entsteht (c). Entsprechendes gilt im stochastischen Fall für $E[B_S]$, $E[Y_S]$, $E[F_S]$.

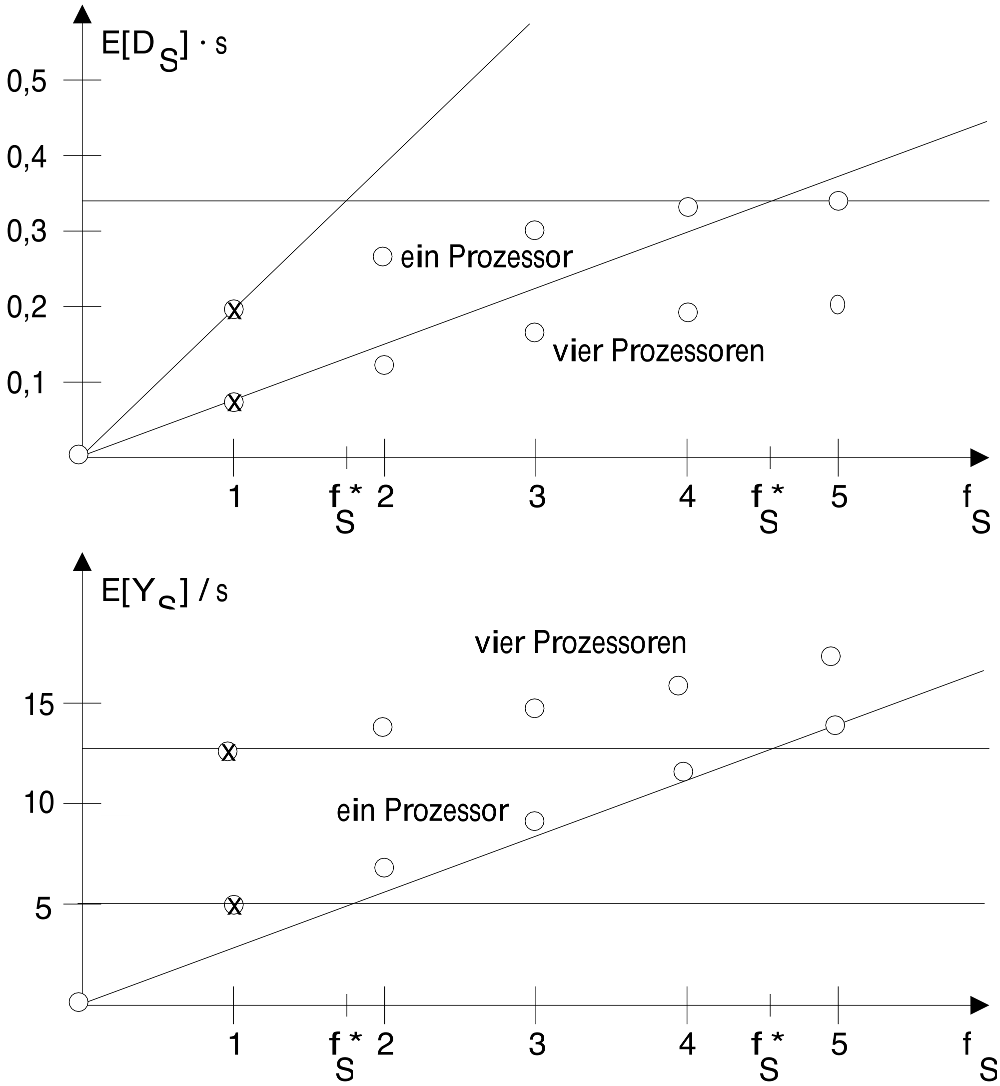


Abbildung 11.18: Erbringung der Prozessorleistung durch einen und durch vier Prozessoren; Verweilzeit $E[Y_S]$, Durchsatz $E[D_S]$, Füllung (grad des Mehrprogrammbetriebs) f_S. Nur die Punkte $\otimes$ sind berechnet; die Punkte $\circ$ sind geschätzt.