4. Mathematische Leistungsmodelle

Problemfeld

Erinnern wir uns an Kap. 1: Wir betrachten eine gewisse Menge von Rechensystemen (MASCHINEN) sowie eine gewisse Menge von Rechenlasten (LASTEN), die von diesen Rechensystemen bearbeitet werden können. Wir einigen uns ferner darauf, wie wir die "Güte der Bearbeitung einer Last durch ein System" bewerten wollen, und legen gewisse (dieser Absicht angepaßte) Leistungsmaße samt einer zugeordneten Menge von Leistungswerten (L-WERTE) fest. Unser Ziel ist es, den Zusammenhang zwischen den verschiedenen (Maschine, Last)-Paaren und den jeweils zugehörigen Leistungswerten zu ermitteln. Wir hatten diesen Zusammenhang gekennzeichnet durch die "Leistungsfunktion" (1.18)

L: MASCHINEN x LASTEN —> L-WERTE

deren genaue Kenntnis uns die Beantwortung aller im Leistungsbereich relevanten Fragen ermöglichen würde.

Objekt-experimente

Wir haben inzwischen gelernt (Kap. 2 und 3), daß wir mittels Objektexperimenten die Funktion L punktweise abtasten können. Zu diesem Zweck mußten wir interessierende Rechensysteme (bzw. Rechensystem-Versionen) jeweils einzeln de facto installieren/implementieren, sie mit interessierenden Lasten (bzw. Last-Versionen) jeweils einzeln de facto beschicken, den Vorgang der Bearbeitung einer Last durch ein System mittels geeigneter Meßinstrumente jeweils einzeln beobachten und schließlich aus den Beobachtungen den gemäß gewählter Leistungsmaße jedem (System,Last)-Paar zugeordneten Leistungswert jeweils einzeln errechnen. Methodisch gesehen erhöht sich mit jedem durchgeführten Experiment der Grad unserer Kenntnis der Leistungsfunktion; andererseits ist jede neu auftretende konkrete Frage wieder nur über neue Objektexperimente zu klären.

Die Leistungsbewertung von RS mittels Objektexperimenten ist ganz offensichtlich eine sehr aufwendige Technik. Sie besitzt darüber hinaus den zusätzlichen Nachteil, daß sie während Entwurfsphasen prinzipiell nicht einsetzbar ist (Entwurf bedeutet ja, daß das Objekt noch nicht existiert). Wir besinnen uns daher auf die in Kap. 1 bereits skizzierte, hoffnungsvolle Alternative: Anstatt ein Objekt-System direkt zu beobachten, versuchen wir, ein Ersatz-System zu schaffen, das einerseits dem zu untersuchenden Objekt in den wesentlichen Eigenschaften "stark ähnelt", andererseits aber "leichter manipulierbar" ist als das Objekt selbst. Jedes derartige Ersatzsystem hatten wir "Modell" getauft; uns geht es speziell um Modelle, die in der Lage sind, unsere Leistungsfunktion (s. oben) in konkreten Fällen anstelle der Objektwelt und in hinreichender Ähnlichkeit zu ihr zu repräsentieren. Solche Modelle nennen wir "Leistungsmodelle".

Modelle

mathematische Modelle

Wir wollen uns in diesem und dem folgenden Kapitel mit mathematischen Modellen befassen. Bei diesem Modelltyp geht es darum, aus Angaben über Struktur und Parameterwerte von System und Last, (zugehörige) Leistungswerte mittels mathematischer Techniken zu "errechnen", bzw. einen mathematischen Zusammenhang der diversen involvierten Größen zu beschreiben. Als Idealziel eines solchen Unterfangens wären zweifellos geschlossene Formeln anzusehen, die ein leichtes Hantieren in einer Art erlauben, wie sie etwa bei dem bekannten mathematischen Modell des freien Falls $s = g/2 \cdot t^2$ vorliegt. Wir werden dieses Ziel nicht in dieser idealen Form und auch nicht für alle denkbaren Problemfälle erreichen. Wir werden aber für eine praktisch relevante Klasse von Problemen implizite mathematische Zusammenhänge zwischen Maschinenparametern, Last-

parametern und Leistungswerten ableiten, die es (im Rahmen dieser Problemklasse) erlauben, Leistungswerte auf mathematischem Wege zu ermitteln.

Lernziele

Dazu werden wir im vorliegenden Kapitel (als Lernziel)

Verkehrsnetze

- eine strukturierte Vorstellungswelt für mathematische Leistungsmodelle, die sog. Verkehrsnetze, kennenlernen sowie die Fähigkeit erwerben, Probleme der realen Welt in diese Modelwelt zu übersetzen bzw. zu erkennen, wo eine solche Übertragung scheitert;

Betriebsanalyse

- in einer ersten quantitativen Interpretation der Verkehrsnetze eine auf der Vorstellung von Messungen aufbauende Analysetechnik, die sog. Betriebsanalyse, betrachten; wir werden dabei eine Reihe, auch für zukünftige Überlegungen wesentlicher, Größen und Begriffe definieren (so: Ankunfts- und Abgangsrate; Bedienbedarf, Bediengeschwindigkeit und Bedienzeit; arbeitserhaltende Station; Übergangsrate und Übergangshäufigkeit; Auslastung und Flaschenhals) und mit ihnen umzugehen lernen; wir werden erkennen, daß diverse Meßgrößen formelmäßig zusammenhängen und werden diese Zusammenhänge (konkret: Auslastungsgesetz, Verkehrsflußgleichgewicht, Flaschenhals beim offenen und geschlossenen Modell) einzusetzen lernen;

stochastische Verkehrsnetze

- in einer weiteren, stochastischen Interpretation der Verkehrsnetze diese soweit konkretisieren, daß sie als Generatoren stochastischer Betriebsabläufe erkennbar sind; wir werden verstehen lernen, daß die mathematische Modellklasse "stochastischer Prozeß" zur Analyse der Zusammenhänge zwischen den Eigenschaften eines stochastischen Verkehrsnetzes einerseits und den Merkmalen der Menge generierbarer Betriebsabläufe andererseits geeignet ist; wir werden lernen, mit dem Begriff "Zustand" umzugehen, daraus einen prinzipiellen Plan für die effektive Analyse stochastischer Verkehrsnetze ableiten und diesem Pfad zumindest ein Stück weit folgen.

analytische Modelle

In der Literatur stößt man häufig auf den Begriff "analytische Modelle". Dabei sind meist die im letzten Absatz skizzierten mathematischen Modelle auf der Basis stochastischer Verkehrsnetze gemeint.

4.1 Modellbildung: Verkehrsnetze

Vielen gebräuchlichen Leistungsmodellen des mathematischen Typs liegt eine Abstraktion struktureller Natur zugrunde, nach der ein Rechensystem in der Form eines Verkehrsnetzes dargestellt und untersucht wird. Ein Verkehrsnetz besteht (in der hier verwendeten Terminologie) aus zwei Komponenten, der "Maschine" und der "Last". Die Maschine wird dabei durch einen (zeitlich festen) gerichteten Graphen repräsentiert, die Last durch eine (oft zeitlich variierende) Menge von Prozessen.

Verkehrsnetz

Werden wir konkreter: Die Maschine besteht aus einer Menge von "Betriebsmitteln" (auch: "Ressourcen", "Funktionseinheiten", "Stationen") sowie einer Menge von "Übergangsmöglichkeiten" (auch: "Verbindungen", "Wegen") zwischen gewissen Paaren dieser Betriebsmittel. In graphischer Darstellung erhalten wir somit den erwähnten Graphen, ein Netz aus Stationen:

Maschine

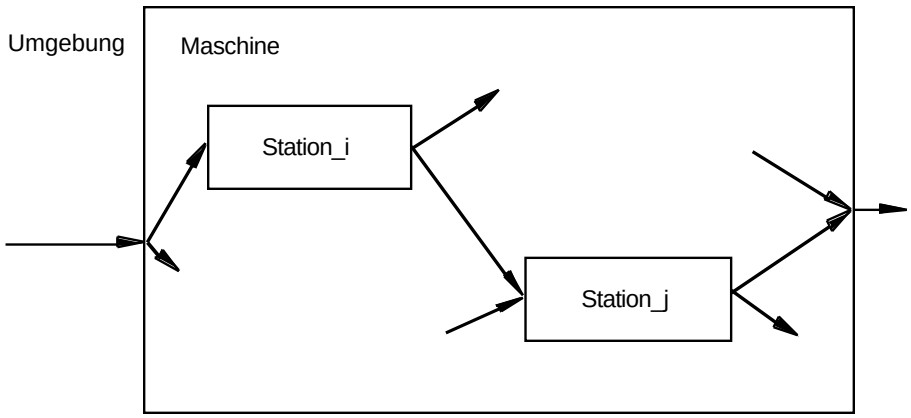


Abbildung 4.1.1: Maschine eines Verkehrsnetzes

Die obige schematische Darstellung gibt exemplarisch nur zwei Betriebsmittel wieder (genannt "Station_i" und "Station_j") - es könnten natürlich mehr sein; auch ist exemplarisch nur eine Übergangsmöglichkeit (zwischen Station_i und Station_j) eingezeichnet - es könnten natürlich mehr sein, so z.B. von Station_i nach anderen Betriebsmitteln, von anderen Betriebsmitteln nach Station_j, u.s.w. Des weiteren unterscheidet die Darstellung explizit zwischen der Maschine und ihrer Umgebung (dem "Rest der Welt") und weist Übergangsmöglichkeiten aus der Umgebung in die Maschine (zu gewissen Betriebsmitteln) und aus der Maschine (von gewissen Betriebsmitteln) in die Umgebung aus.

Die zweite Komponente, die Last, besteht aus einer Menge von "Prozessen" (je nach Auffassung auch: "Lasteinheiten", "Aufträgen", "Tasks", "Jobs", "Kunden"). Jeden Prozeß stelle man sich zunächst statisch vor als eine zusammengehörige Menge von Betriebsmittelanforderungen (Einzelanforderungen an einzelne Betriebsmittel der Maschine) und eine Reihenfolgevorschrift, die eine zeitliche (u.U. nur partielle) Ordnung über den Betriebsmittelanforderungen festlegt. Wir werden zur Erhöhung der Verständlichkeit die Begriffe Betriebsmittelanforderung und Reihenfolgevorschrift später detaillierter diskutieren. Zunächst aber: Wir haben in dieser statischen Charakterisierung eines Prozesses eine Verhaltensvorschrift vor uns, die wir "Prozeßmuster" nennen wollen. Ein Prozeß im dynamischen Sinne entsteht, wenn die Maschine dieses Prozeßmuster als Arbeitsauftrag versteht, dem sie Schritt für Schritt nachkommt, indem sie Betriebsmittelanforderung nach Betriebsmittelanforderung in der vom Prozeßmuster geforder-

Last

ten Abfolge erfüllt. Bleibt zu klären, wie Prozesse (in unserer Modellwelt) entstehen, wofür wir zwei unterschiedliche Fälle vorsehen:

temporäre Prozesse

- Im einen Fall generiert eine nicht näher betrachtete Umgebung den Prozeß und übergibt ihn der Maschine, von welchem Zeitpunkt ab er existiert; wir können uns z.B. den "Benutzer" vorstellen, der sein Programm irgendwie "startet"; in diesem Fall ist es auch sinnvoll, anzunehmen, daß der Prozeß irgendwann (nach vollständiger Abarbeitung) wieder terminiert; wir nennen einen solchen Prozeß "temporär" und ein System/Modell, das nur temporäre Prozesse vorsieht, ein "offenes".

permanente Prozesse

- Im anderen Fall wird der Prozeß weder generiert noch terminiert; er existiert "dauernd" (und muß daher in seiner Anforderungsreihenfolge wohl irgendwie zyklisch sein); wir nennen einen solchen Prozeß "permanent" und ein System/Modell, das nur permanente Prozesse vorsieht, ein "geschlossenes". Diese (auf den ersten Blick sicher wirklichkeitsfremd anmutende) Vorstellung permanenter Prozesse wird sich schnell als äußerst hilfreiche Fiktion erweisen.

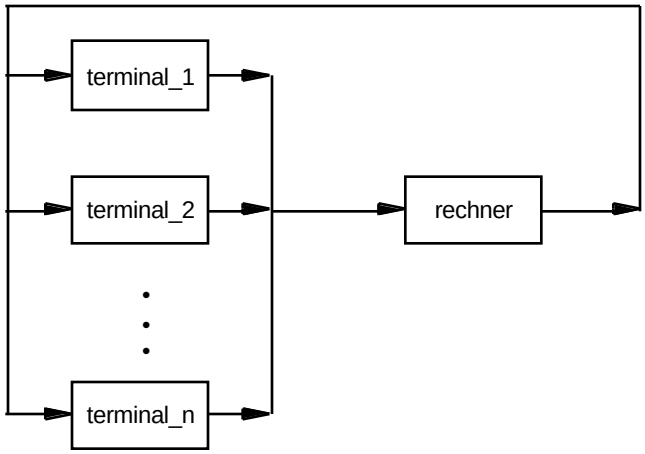
Testfrage

Testfrage 4.1.2: Warum ist es i.allg. sinnvoll, anzunehmen, daß ein explizit generierter ("der Maschine übergebener") Prozeß auch wieder terminiert?

Schieben wir hier einige veranschaulichende Überlegungen ein. Die skizzierte Vorstellung von Verkehrsnetzen wird häufig ohne detailliertere Erklärung in Vorlesungen und Büchern über Rechnerarchitektur, Betriebssysteme u.ä. eingesetzt. Wir greifen willkürlich zwei typische Beispiele heraus.

Beispiel

Beispiel 4.1.3: Häufig stößt man auf folgendes Verkehrsnetz als simples Modell eines Teilnehmerrechensystems. Dabei wird die Maschine (in unserer jetzigen Terminologie) durch einen Graphen repräsentiert:



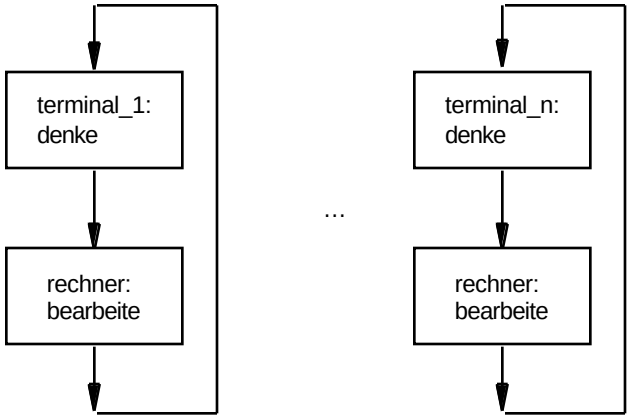
Wir haben es demnach mit n+1 Stationen zu tun: Den Endgeräten "terminal_1" bis "terminal_n" und dem eigentlichen "rechner". Als Wege sind vorgesehen die Übergangsmöglichkeiten von terminal_i nach rechner und von rechner nach terminal_i, (i=1,2,…,n). Das Modell hat keine Ein/Ausgänge von/nach irgendeiner Umgebung, ist also geschlossen.

Von der Last besteht ungefähr folgende Vorstellung: An ("in") jeder Station terminal_i (i=1,2,…,n) wird eine Weile nachgedacht (und dabei wohl ein Auftrag vorbereitet), dann wird ein Auftrag an die Station rechner gesandt (die diesen dann wohl bearbeitet), dann wird wieder nachgedacht, dann bearbeitet u.s.w.

Notation für Prozeßmuster

Diese verbal geschilderten Verhaltensregelmäßigkeiten sind in unserem Sinne "Prozeßmuster" (s. Kap. 1). Über eine formalere Notation für Prozeßmuster haben wir uns noch keine Gedanken gemacht; naheliegend und einleuchtend sind folgende Formulierungen,

• eine graphische Notation:



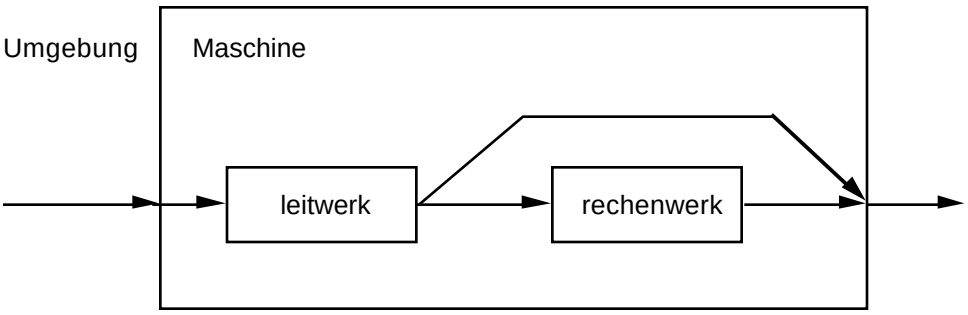
• eine sprachliche Notation:

```
LOOP                          LOOP
   terminal_1.denke;             terminal_n.denke;
   rechner.bearbeite     ...     rechner.bearbeite
ENDLOOP                       ENDLOOP
```

beide das Wesentliche ausdrückend: In fortwährender Wiederholung wird von einem terminal_i eine "denke"-Tätigkeit verlangt, anschließend von rechner eine "bearbeite"-Tätigkeit u.s.w. Wir haben es insgesamt mit n Prozeßmustern zu tun (jedes ist den Vorgängen bzgl. genau einer Station_i zugeordnet); um die Beschreibung zu komplettieren, können wir sinnvollerweise (an jedem Terminal sitzt wohl nur ein Benutzer) festlegen, daß es zu jedem Prozeßmuster genau einen Prozeß gibt. Alle Prozesse sind permanent (wir machen uns ganz einfach keinerlei Gedanken darüber, wie der Betrieb beginnt oder endet), wie auch die Prozeßmuster ausweisen.

Beispiel

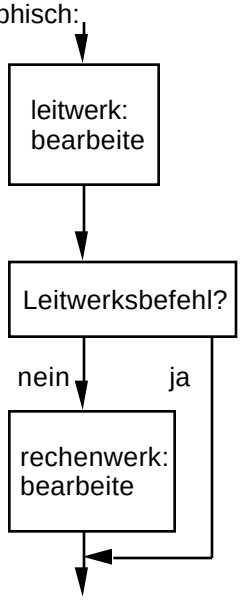
Beispiel 4.1.4: Betrachten wir folgendes Verkehrsnetz als simples Modell eines Zentralprozessors. Als Maschine diene der Graph



Zwei Stationen also: "leitwerk" und "rechenwerk"; dazu die Wege von Umgebung nach leitwerk, von leitwerk nach rechenwerk und Umgebung, von rechenwerk nach Umgebung. Das Modell hat Ein- und Ausgang, ist also offen.

Die Last unterliegt folgender Vorstellung: Aufträge ("Befehle") betreten das Netz, leitwerk muß zunächst für jeden Auftrag tätig werden, danach ist entweder die Bearbeitung bereits abgeschlossen (im Falle eines "Leitwerksbefehls") oder rechenwerk muß tätig werden (falls kein reiner Leitwerksbefehl vorliegt). Die Entscheidung Leitwerksbefehl bzw. kein Leitwerksbefehl erfolge (in unserem Modell!) zufällig, z.B. mit Wahrscheinlichkeiten 1-r bzw.r. Dabei machen wir uns im jetzigen Kontext über die Darstellung des "Würfelns" Leitwerksbefehl/kein Leitwerksbefehl keine Gedanken. Als Prozeßmuster erhalten wir:

Graphisch:



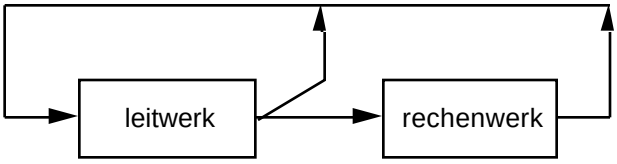
Sprachlich:

leitwerk.bearbeite;
IF "kein Leitwerksbefehl"
THEN rechenwerk.bearbeite

Wir haben es mit genau einem Prozeßmuster zu tun, dem jeder generierte Prozeß folgt; es kann (im Lauf der Zeit) beliebig viele solche Prozesse geben, alle sind temporär. Um die Beschreibung zu komplettieren, müßten wir festlegen, wann die diversen einzelnen Prozesse generiert werden sollen (in der Terminologie des Beispiels: wann die einzelnen Befehle dem Zentralprozessor zur Ausführung übergeben werden). Bevor wir uns in der Erfindung eines diesbezüglichen "Generator-Prozesses" verlieren (i.allg. ist ein solcher sicher vonnöten und könnte im einfachsten Fall auf einer Tafel mit Generierungszeitpunkten basieren), sollten wir beachten, daß beliebig festgelegte Prozeß-Generierungszeitpunkte dem aktuellen Beispiel nicht gerecht werden: Der Zentralprozessor bearbeitet ja ein Programm, das im Normalfalle sofort nach Abarbeitung eines Befehls einen nächsten zur Abarbeitung bereithält. Wenn wir den Fall keines verfügbaren Programms ganz einfach ignorieren, dürfte folgendes Verkehrsnetz die Sachlage besser treffen:

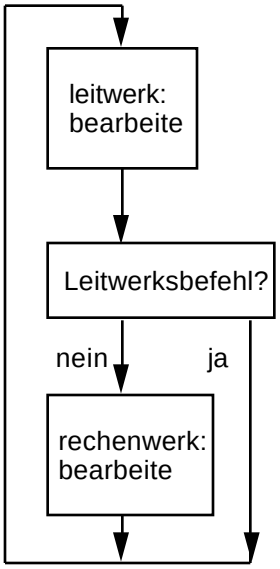
Beispiel in alternativer Form

Maschine:



Bemerke: Das Modell ist geschlossen

Last (Prozeßmuster):




```
LOOP
leitwerk.bearbeite
  IF "kein Leitwerkbefehl"
  THEN rechenwerk.bearbeite
ENDLOOP
```

Last (Prozesse):
Es gibt genau einen permanenten Prozeß zu obigem Prozeßmuster, der ein ganzes Programm, mehr noch: eine Folge von Programmen, präziser: die Aufträge einer Folge von Programmen an einen Zentralprozessor, repräsentiert.

Kehren wir nach diesen Beispielen zur Hauptlinie zurück, für die eine detailliertere Diskussion der Begriffe Betriebsmittelanforderung und Reihenfolgevorschrift noch ausstand.

Zu den Betriebsmittelanforderungen:

Betriebsmittelanforderungen

Ohne dies explizit zu erwähnen, hatten wir in den Beispielen Anforderungen an die Stationen (Betriebsmittel) gestellt, die funktionaler Natur waren. So sollte in Bsp. 4.1.3 "terminal_i" die Tätigkeit "denke", "rechner" die Tätigkeit "bearbeite" vollziehen u.s.w. Etwas konkreter ausgedrückt war unsere implizite Vorstellung, daß

- Stationen gewisse benannte "Dienste" zu erbringen fähig waren,

- Prozesse die Erbringung dieser Dienste von den Stationen fordern konnten.

Diese Darstellung erinnert Sie sicher (und soll dies auch tun) an Wissensbereiche, die Ihnen aus anderen Veranstaltungen geläufig sind, wie z.B. "Abstrakte Datentypen" oder "Objekt-orientierte Programmierung". Im Unterschied zu diesen Bereichen sind wir aber hier (ich erinnere an Kap. 1)

- an den informationsorientierten Wirkungen der Erbringung von Diensten (also den Resultaten von Berechnungen oder den Veränderungen von Speicherinhalten) wenig interessiert;

- an den physikalischen Wirkungen der Erbringung von Diensten (also der Dauer der Beanspruchung von Prozessoren oder des Umfangs der Beanspruchung von Speichern) stark interessiert.

Es ist daher für Leistungsmodelle (insbesondere auch für die hier betrachteten Verkehrsnetze) bequemer und allgemein gebräuchlich, Betriebsmittelansprüche nicht funktional, sondern in physikalischen Begriffen (Zeit und Raum) anzugeben.

So etwa in Bsp. 4.1.3
nicht: terminal_i.denke
sondern: terminal_i."widme mir (z.B.) 20 sec"

Oder in Bsp. 4.1.4
nicht: leitwerk.bearbeite
sondern: leitwerk."widme mir (z.B.) 10 nsec"

Der Usus, die Anforderungen an Betriebsmittel direkt durch deren physikalische Beanspruchung auszudrücken, bringt allerdings auch Nachteile mit sich. Der wohl gravierendste besteht in der Tatsache, daß eine Angabe der physikalischen Beanspruchung eines Betriebsmittels nur möglich ist, wenn Interna der Bearbeitung im Betriebsmittel (und damit in der modellierenden Station) bekannt sind. Um dies wieder am Beispiel klarzumachen: Die Übersetzung (s. Bsp. 4.1.3)

von: rechner.bearbeite
nach: rechner."widme mir ??"

ist angesichts der anzunehmenden internen Komplexität von "rechner" nicht ohne weiteres möglich; zumindest nicht ohne Interna von "rechner" zu kennen; und auch dann voraussichtlich nicht als simple eindimensionale Größe (wie sich das etwa bei "terminal_i" und "leitwerk" anbot).

Stationen

Als Folge dieser Beobachtung sehen wir uns genötigt, in unseren Verkehrsnetzen die Modellkomponente "Station" nicht als Abbild eines beliebig komplexen "Subsystems" der realen Welt zu verwenden; vielmehr sehen wir Stationen nur als Repräsentanten relativ einfacher Subsysteme, wobei "relativ einfach" dadurch definiert sei, daß wir berechtigt sind (oder uns berechtigt fühlen), in etwa folgende interne Struktur einer Station zu postulieren

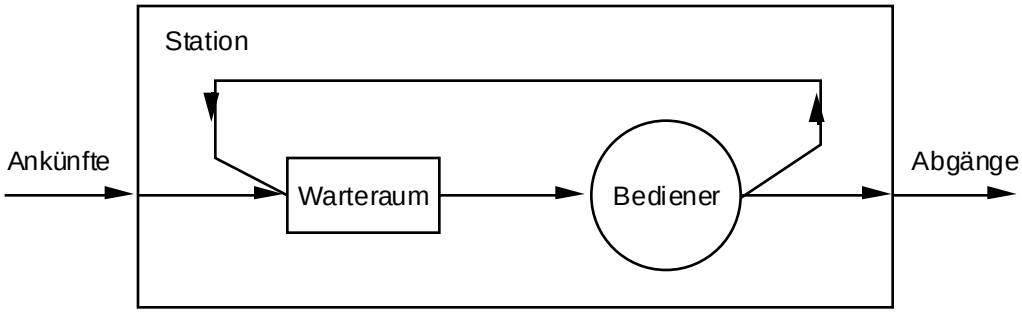


Abbildung 4.1.5: Struktur einer Bedienstation

Ankommende Kunden (Ankünfte) betreten die Station mit einem Bedienwunsch, der eine (eindimensionale) physikalische Beanspruchung des Bedieners ausdrückt. Die Bedienstation widmet sich diesem Bedienwunsch nach eigenen internen Regeln: Sie kann den Kunden z.B. warten lassen, seine Bedienung unterbrechen u.ä. mehr. Schließlich, nach Abarbeitung seines Bedienwunsches, wird der Kunde entlassen (Abgänge). Bei den Stationen und den für sie möglichen Bedienwünschen unterscheidet man zwei wesensverschiedene Arten:

- die "aktive" Bedienstation, die die zeitliche Belegung eines "Prozessors" verwaltet und mit Bedienwünschen der Dimension "Zeiteinheit" konfrontiert wird; Beispiele sind Stationen, die eine Zentraleinheit, einen Kanal, eine Übertragungsleitung u.ä. repräsentieren; *aktive Station*

- die "passive" Bedienstation, die die räumliche Belegung eines "Speichers" verwaltet und mit Bedienwünschen der Dimension "Raumeinheit" konfrontiert wird; Beispiele sind Stationen, die einen Arbeitsspeicher, einen Pufferbereich u.ä. repräsentieren. *passive Station*

Wir werden allerdings die Berücksichtigung räumlicher Bedienwünsche in mathematischen Modellen bald wieder fallenlassen müssen ("leider" und "zunächst").

Zu den Reihenfolgevorschriften:
Wir hatten akzeptiert, daß zwischen den einzelnen Betriebsmittelanforderungen eines Last-Prozesses Reihenfolgevorschriften bestehen können etwa der Art: "Erst muß Tätigkeit a abgeschlossen sein, dann kann Tätigkeit b begonnen werden". Wir hatten in den Beispielen 4.1.3 und 4.1.5 bei den versuchsweisen Notationen für Prozeßmuster solche Reihenfolgevorschriften auch bereits freizügig ausgedrückt (s. dort). Auch tritt Ihnen die Notwendigkeit, Reihenfolgen zwischen einzelnen Arbeitsschritten eines Prozesses zu berücksichtigen, nicht zum ersten Mal entgegen. So werden im Bereich Betriebssysteme u.a. Präzedenzgraphen zur Festlegung der Ordnung über Teilschritten nicht-zyklischer Prozesse eingesetzt. Ich erwähne dies deshalb, weil bei diesen Präzedenzgraphen besonders deutlich wird, daß die Ordnung i.allg. nicht total, sondern nur partiell ist. Ein Lastprozeß kann durchaus (von sich aus) zwischen gewissen seiner Teilschritte keine strikte Reihenfolge vorsehen:

Reihenfolgevorschriften

Präzedenzgraph

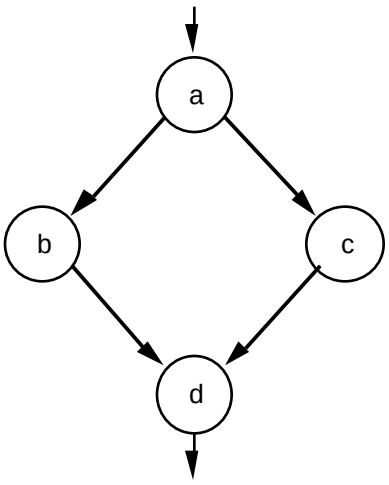


Abbildung 4.1.6: Präzedenzgraph mit partieller Ordnung

Das einfache Beispiel der Abb. 4.1.6 drückt aus: a muß vor b und c erledigt werden, d nach b und c, implizit damit auch d nach a; andererseits ist zwischen b und c keine Reihenfolge festgelegt: b könnte (vom Prozeß aus gesehen) vor oder nach c "drankommen"; b und c könnten auch gleichzeitig "laufen" oder irgendwie "zeitlich überlappend" u.s.f.

gleichzeitige Belegung von Betriebsmitteln

Die Schwierigkeiten der (in dieser Vorlesung nur ansatzweise behandelten) mathematischen Behandlung stochastischer Leistungsmodelle bringen es mit sich, daß der Fall der gleichzeitigen Belegung mehrerer Betriebsmittel durch einen Prozeß nicht immer in voller Allgemeinheit berücksichtigt werden kann. Auch wir bescheiden uns daher und fordern für die Reihenfolgevorschriften bis auf weiteres eine totale Ordnung - wodurch aus den Lastprozessen sequentielle Prozesse werden, aus den Prozeßmustern so etwas wie sequentielle Programme. Natürlich reduziert das unsere Möglichkeiten, die Realität adäquat zu modellieren: Die schmerzlichste Folgerung ist die, daß wir die passiven Bedienstationen wieder vergessen müssen: "Speicherplatz" wird nahezu immer gleichzeitig mit "Prozessorzeit" benötigt - gleichzeitige Belegung von Betriebsmitteln mußten wir aber ausschließen.

Zusammenfassung

Fassen wir zusammen: Strukturell gesehen, stellen Verkehrsnetze eine geeignete Modellklasse zur Beschreibung von Leistungsmodellen dar. Ein Verkehrsnetz setzt sich zusammen aus einer Maschine und einer Last. Die Maschine besteht aus einer Menge von Stationen samt Übergangsmöglichkeiten zwischen den Stationen, wobei die Stationen typmäßig auf den Bereich der aktiven Bedienstationen beschränkt sind. Die Last wird beschrieben durch eine Menge von Prozeßmustern sowie Regeln, welche die Generierung bzw. Zahl von Prozessen bezüglich einzelner Prozeßmuster festlegen. Jedes Prozeßmuster notiert alle Bedienwünsche zugehöriger Prozesse und ihre Reihenfolge, wobei die Bedienwünsche als zeitliche Beanspruchung jeweils eines Bedieners (einer Station) angegeben werden und die Bedienwünsche streng sequentiell aufeinanderfolgen. Der Verkehr im Netz (als Abbild des Betriebs des Rechensystems) entsteht durch die Bewältigung der Bedienwünsche aller Lastprozesse seitens der Stationen der Maschine, wobei dank vorgegebener Restriktionen jeder Lastprozeß dynamisch als "Kunde" gesehen werden kann, der von Station zu Station wandert (insbesondere zu jedem Zeitpunkt in genau einer Station verweilt).

4.2 Betriebsanalyse (Operational Analysis)

Einordnung

Das Konzept der Betriebsanalyse (oft auch als "operationale Analyse" ins Deutsche übersetzt) ist in in einer ganzen Reihe von Arbeiten eingeführt, ausgebaut und angewendet worden (guter Übersichtsartikel: DeBu78; wichtige Erweiterung: Rood79, durchgängige Verwendung im Lehrbuch LZGS84). Obzwar im allgemeinen nicht zu den analytischen Modellen gerechnet, verwenden die Modelle der Betriebsanalyse unzweifelhaft eine Technik, die auf mathematischem Weg Zusammenhänge zwischen Kenngrößen von Maschine, Last und Leistung aufdeckt: Die Betriebsanalyse zählt daher in unserem Sinne zu den mathematischen Modellierungstechniken. Vor allem aber ist die Betriebsanalyse eine Technik, die den Vorgang der Modellbildung im Bereich analytischer Leistungsmodelle auf sehr leicht faßbare Weise unterstützt.

Meßbarkeit, Überprüfbarkeit

Es zählt zu den Prinzipien der Betriebsanalyse, daß alle verwendeten Größen während des Betriebs eines Rechensystems meßbar sind (bzw. aus solchen meßbaren Größen mittels klar definierter Formeln errechnet werden können) sowie daß sämtliche einfließenden Annahmen während des Betriebs eines Rechensystems mittels Messungen überprüft werden können (daher auch der Name der Technik: "Betriebs"-Analyse, "operational" analysis). Um evtl. Mißverständnissen von vornherein vorzubeugen: Wir setzen nicht voraus, daß alle verwendeten Größen tatsächlich während des Realbetriebs gemessen werden - sie sollten nur "im Prinzip" meßbar sein.

Meßpunkte

Zur Definition der Meßpunkte bemühen wir die vorgestellte Maschinenstruktur nach Abb. 4.1.1, sehen das Rechensystem also von vornherein als Verkehrsnetz. Die Meßpunkte liegen jeweils an Stations-Eingängen und -Ausgängen, so daß wir den Verkehr der Kunden in die Stationen hinein und aus den Stationen heraus beobachten können. Nicht unmittelbar beobachtbar sind damit natürlich die Prozeßmuster (die "treibende Kraft" des Betriebs).

Station

Konzentrieren wir uns zunächst auf eine beliebige einzelne Station des Verkehrsnetzes:

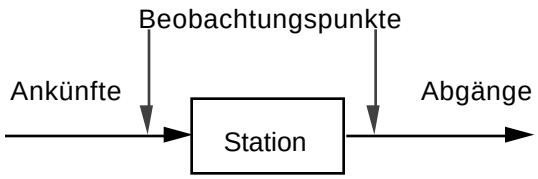


Abbildung 4.2.1: Einzelstation

Wir halten folgende (im Prinzip sicher meßbaren) Größen fest:

Basisgrößen

Definition 4.2.2 Basisgrößen

T Gesamtdauer der Beobachtung

sowie die während des Beobachtungsintervalls beobachteten

A Zahl der Kundenankünfte ("arrivals")
C Zahl der Kundenabgänge ("completions")
B Belegtzeit der Station ("busy time"),
definiert als Gesamtzeit (Summe aller Zeitintervalle) mit mindestens einem Kunden an der Station

Als Anmerkung: Zur Messung von B ist es erforderlich, zu jedem Zeitpunkt den Belegungszustand der Station zu kennen ("wieviele Kunden sind drin?") - zu-

mindest zu erkennen, ob überhaupt ein Kunde anwesend ist. Der Belegungszustand läßt sich aber bei Kenntnis des Anfangszustands (zu Beginn der Beobachtung, z.B. "leer") aus der Zahl der Ankünfte und Abgänge "von außen" ermitteln.

Aus den Meßgrößen lassen sich leicht folgende Größen ableiten:

Abgeleitete Größen

Definition 4.2.3: Abgeleitete Größen

$a := A/T$
$c := C/T$
$\underline{b} := B/T$
$\overline{B} := B/C$

Benennungen

Für die rein formal abgeleiteten Größen der Def. 4.2.3 haben sich bestimmte Benennungen eingebürgert. Diese legen allerdings Interpretationen nahe, die nicht in allen Fällen korrekt sind und deshalb genau diskutiert werden sollen. Zunächst spielt bei allen Benennungen die Vorstellung einer relativ großen Beobachtungszeit T und einer gewissen statistischen Regelmäßigkeit eine Rolle. Unter diesen Voraussetzungen sind die Benennungen

a (=A/T) "Ankunftsrate"
c (=C/T) "Abgangsrate"
b (=B/T) "Belegungsgrad", "Belegungsanteil"

einleuchtend und hilfreich.

Testfrage

Testfrage 4.2.4: Inwiefern sind die Vorstellungen einer "relativ großen Beobachtungszeit" und einer "gewissen statistischen Regelmäßigkeit" wesentlich für die Angemessenheit der Benennungen Ankunfts-, Abgangs-"Rate"?

Mittlere Bedienzeit

Wesentlich vorsichtiger zu genießen ist dagegen die Benennung

$\overline{B}$ (=B/C) "mittlere Bedienzeit"

Erinnern wir uns an unsere prinzipiellen Annahmen über Verkehrsnetze: Kunden betreten eine Station mit einem Bedienwunsch, dessen Größe wir durch die Dauer der (bevorstehenden) Beanspruchung des Bedieners angeben wollten. Über eine gewisse Menge von Kunden (mit in der Regel unterschiedlichen Bedienwünschen) gemittelt, könnten wir einen mittleren Bedienwunsch errechnen, der nach unserer Konvention der mittleren zeitlichen Beanspruchung des Bedieners pro Kunde entspricht. Nun sind uns in der Betriebsanalyse (aufgrund der festgelegten Meßtechnik) die Bedienwünsche der Kunden nicht direkt zugänglich, so daß wir versucht sein werden, Kenngrößen für Bedienwünsche (z.B. deren Mittel) aus den (gemessenen) Basisgrößen zu ermitteln. Die Schlußfolgerung aber

Mittlerer Bedienwunsch pro Kunde
= mittlere zeitliche Beanspruchung des Bedieners pro Kunde
(aufgrund der Konvention der Angabe)
= "mittlere Bedienzeit"
(auch diese Definition wäre sinnvoll - und wird auch benutzt!)
= $\overline{B}$

Doppeldeutigkeit

ist nur unter zusätzlichen Annahmen über die Funktionsweise der Bedienstation korrekt (und deckt damit die Gefährlichkeit der in zweifachem Sinne gebrauchten Benennung "mittlere Bedienzeit" auf).

Bedienungsmechanismus

Gehen wir dieser Frage etwas weiter nach. Habe unsere aktive Bedienstation die einfache Struktur der Abb. 4.1.5. Wie "bedient" die Station die ankommenden Kunden? Nehmen wir zunächst an, die Station sei bei Eintreffen eines Kunden unbeschäftigt (englisch: "idle") und habe auch genügend Zeit, seine Bedienung

vor Eintreffen des nächsten Kunden zu beenden. Wir lockern unsere Konvention der Angabe von Bedienwünschen etwas, indem wir akzeptieren, daß der Bedienwunsch des Ankömmlings zunächst in irgendwelchen eindimensionalen "Arbeitseinheiten" (AE) angegeben wird (Zahl auszuführender Instruktionen, Länge zu übertragender Nachricht o.ä.), wobei der konkrete Bedienwunsch des betrachteten Neuankömmlings die Größe W habe. Der Bediener in der Station wende sich unmittelbar der Bedienung des Neuankömmlings zu und führe diese ohne Unterbrechung zu Ende. Er benötigt dazu ("aktive" Station!) eine gewisse Zeit S (die "Bedienzeit"), die bei sinnvoller Wahl der Maßeinheit AE monoton wachsend mit W zusammenhängt. Wir lassen ausschließlich eine direkte Proportionalität zwischen W und S zu, in Zeichen

(4.2.5) $S = W / r$

so daß r auch als "Bearbeitungsgeschwindigkeit" interpretiert werden kann: Die Station "schafft" r AE je ZE (Zeiteinheit). Der einfachste Fall ist offensichtlich r=1; kein spezieller Sonderfall, sondern in unserem Belieben, wenn wir uns entschließen, Bedienwünsche in Bezug auf die speziellen Stationsfähigkeiten zu messen: "Der Bedienwunsch entspricht einer Bedienzeit von S Zeiteinheiten" bezeichnet genau unsere Konvention, Bedienwünsche durch Beanspruchungen des Bedieners zu kennzeichnen.

Bearbeitungsgeschwindigkeit

Wie bedient die Station in komplizierteren Fällen? Stellen wir uns die Organisation des Betriebs innerhalb der Station geregelt vor durch eine "Bediendisziplin". Diese hält einen Teil der anwesenden Kunden im Warteraum (wo sie einfach warten, also nicht bedient werden), den Rest der anwesenden Kunden in der Bedieneinrichtung (wo sie alle gleichzeitig einen Fortschritt in ihrer Bedienung erfahren). Der einfachste, leicht vorstellbare Fall ist offensichtlich die Bedienung genau eines Kunden zu jedem Zeitpunkt (wenn die Station nicht "leer" ist); andere Fälle werden wir als bequeme Modell-Fiktionen später kennenlernen. Nehmen wir nun zusätzlich an, die Bearbeitungsgeschwindigkeit des Bedieners sei konstant gleich r, gleichgültig wieviele Kunden in der Station weilen (in diesem Kontext verwendet man gerne auch die Bezeichnung "Arbeitskapazität" statt "Bearbeitungsgeschwindigkeit") und diese Arbeitskapazität komme voll den anwesenden Kunden zugute (der Bediener legt also insbesondere keine "Pausen" ein, wenn Arbeit vorhanden ist und er hat auch keine sonstigen "Verwaltungstätigkeiten", englisch: "overhead activities", zu leisten). Abb. 4.2.6 stellt einen möglichen Verlauf des Stationszustandes über der Zeit dar (ein "Belegungsgebirge").

Bediendisziplin

Arbeitskapazität

Wir beobachten C (in der Abbildung: 5) Abgänge. Seien die Bedienwünsche der zu diesen C Abgängen gehörigen Kunden (in irgendeiner Reihenfolge) mit $W_1, W_2, \ldots, W_C$ bezeichnet. Die gesamte zu leistende Arbeit betrug also

(4.2.7a) $$W_{total} = \sum_{i=1}^{C} W_i$$

Jedem Bedienwunsch W_i entspricht (Annahme: isolierte, ungestörte Bearbeitung) nach (4.2.5) eine Bedienzeit

(4.2.7b) $S_i = W_i / r \qquad i=1,2,\ldots,C$

Belegungs-gebirge

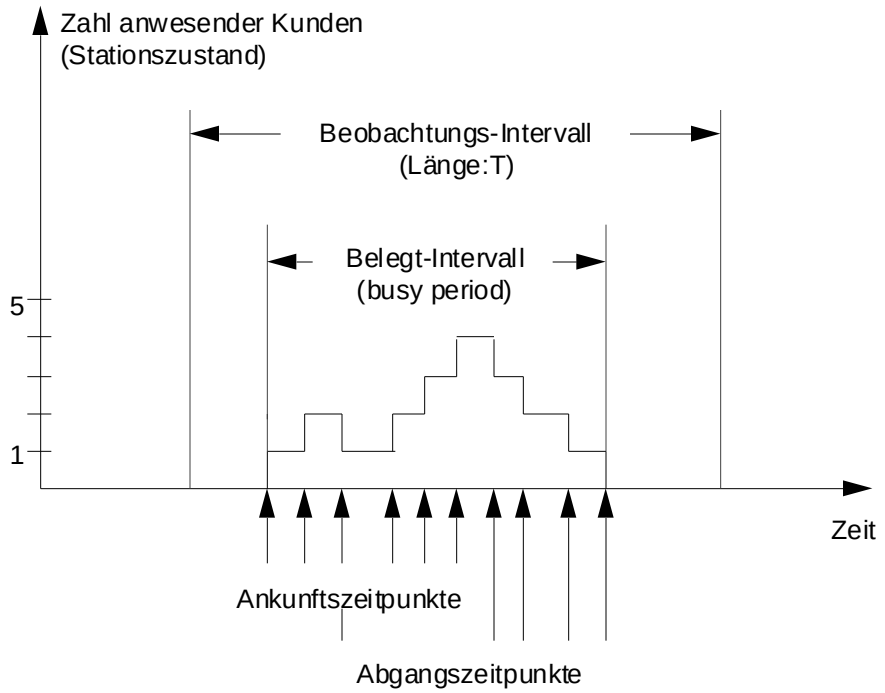


Abbildung 4.2.6: Belegungsgebirge

Als mittlere Bedienzeit (im Sinne der mittleren Belastung des Bedieners pro Kunde) erhalten wir für diese C Kunden

(4.2.7c) $$\bar{S} = \frac{1}{C} \sum_{i=1}^{C} S_i = \frac{1}{Cr} \sum_{i=1}^{C} W_i$$

und bei lückenlosem "Aneinanderlegen" aller Bedienzeiten eine totale Bedienzeit

(4.2.7d) $$S_{total} = \sum_{i=1}^{C} S_i = \frac{1}{r} \sum_{i=1}^{C} W_i = C\ \bar{S}$$

Andererseits benötigt der Bediener zur Abarbeitung der gesamten Arbeit W_{total} (wegen der konstanten Geschwindigkeit r) die Zeit

(4.2.7e) $$B = W_{total} / r$$

die wir schon mit B bezeichnet haben, da sie genau der Länge des Belegt-Intervalls (und damit der Belegtzeit gemäß Def. 4.2.2) entspricht: Der Bediener ist ja so lange tätig, wie Arbeit vorhanden ist - ein Belegt-Intervall beginnt mit der Ankunft eines "ersten" Kunden (bei leerer Station) und endet mit der erfolgten Bewältigung der gesamten Arbeit dieses ersten Kunden und derer, die nachfolgend (bei nicht-leerer Station) eintrafen. Aus (4.2.3, 4.2.7d, 4.2.7e) erhält man leicht

(4.2.7f) $$S_{total} = B \quad \text{bzw.:} \quad \bar{S} = \bar{B}$$

womit in diesem Fall die "mittlere Bedienzeit" $\bar{B}$ im Sinne der Def. 4.2.3 und die "mittlere Bedienzeit" $\bar{S}$ im Sinne von (4.2.7c) tatsächlich zusammenfallen. Wie man sich leicht überlegen kann, bleibt die Gleichheit erhalten, solange die Bedienstation "arbeitserhaltend" und "von konstanter Bedienkapazität" ist:

arbeits-erhaltende Station

Definition 4.2.8: Eine arbeitserhaltende Bedienstation (vom Englischen "work-conservative", kurz "conservative", mit der üblichen Übersetzung "konservativ" aber nicht treffend bezeichnet) ist immer tätig, solange zu bearbeitende Tätigkei-

ten anstehen; sie erledigt genau die Tätigkeiten, die durch die Bedienwünsche der Kunden bezeichnet sind. Eine Station konstanter Bedienkapazität arbeitet, solange es zu bearbeitende Tätigkeiten gibt, mit konstanter Bearbeitungsgeschwindigkeit (Dimension: Arbeitseinheiten / Zeiteinheit).

konstante Bedienkapazität

Testfrage 4.2.9: Betrachten Sie Stationen mit den Bediendisziplinen

- FCFS (First Come First Served);
- HOL (Head Of the Line);
- RR (Round Robin).

Inwiefern, bzw. unter welchen Bedingungen, treffen die Gleichungen (4.2.7) für diese Stationen zu?

Testfrage

Wie man sich ebenfalls leicht überlegen kann, bleibt die Gleichheit $\overline{S} = \overline{B}$ auch erhalten, wenn das Beobachtungsintervall nicht (wie in Abb. 4.2.6) nur ein, sondern mehrere Belegt-Intervalle (voll) umfaßt. Sie geht allerdings verloren, wenn zu Beginn oder Ende des Beobachtungsintervalls die Station nicht leer ist. Mit "genügend großem" T wird die Abweichung dennoch klein, so daß immer noch $\overline{S} \approx \overline{B}$.

erweiterte Gültigkeit von (4.2.7f)

Um Ihnen zu zeigen, daß es durchaus Fälle gibt, in denen $\overline{S} \neq \overline{B}$, betrachten wir einen anderen (aber voll im Rahmen der Bedienstation nach Abb. 4.1.4 liegenden) Stationstyp (vgl. Abb. 4.2.10). Der Bediener bestehe aus beliebig vielen (untereinander identischen) Sub-Bedienern; ankommende Kunden finden jedenfalls unmittelbar einen freien Sub-Bediener, der ihre (jeweils alleinige, völlige) Bedienung übernimmt; der Warteraum ist demzufolge immer leer. Die Station sei arbeitserhaltend. Sie ist aber offensichtlich nicht von konstanter Bedienkapazität: Bei einer angenommenen Arbeitsgeschwindigkeit von r je Sub-Bediener ist vielmehr die Bedienkapazität der Gesamtstation bei Anwesenheit von einem Kunden gleich r, bei zwei anwesenden Kunden 2·r (es werden ja beide gleichzeitig mit Geschwindigkeit r bedient), allgemein bei n>0 anwesenden Kunden n r. Hier ist sicher die für eine Station konstanter Bediengeschwindigkeit gültige Schlüsselbeziehung (4.2.7e) nicht gegeben: $\overline{B}$ hat zu $\overline{S}$ keinen einfachen Bezug. Wir werden diesen Stationstyp später wieder aufnehmen; für hier soll lediglich noch darauf hingewiesen werden, daß die Station (neben anderen Benennungen) deshalb den Namen "Verzögerungsstation" trägt, weil sie jeden Kunden genau nach Ablauf seiner Bedienzeit S_i entläßt, S_i also als reine Verzögerung im Ablauf des verursachenden Prozesses gesehen werden kann.

Verzögerungsstation

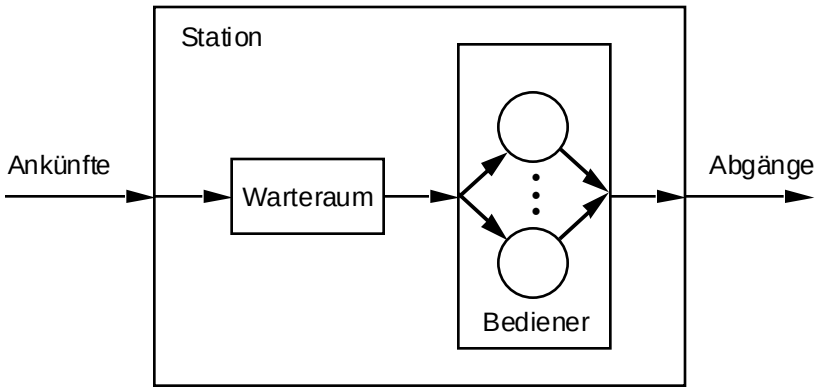


Abbildung 4.2.10: Verzögerungsstation

Testfrage 4.2.11: Nehmen Sie sich Beispiel 4.1.3 nochmals vor. Können Sie unter Einsatz einer Verzögerungsstation zu einer einfacheren Maschinen- und Last-

Testfrage

(Modell-)Vorstellung kommen als dort verwendet? Hinweis: "Verschmelzen" Sie die Endgeräte.

Auslastung

Kehren wir nach diesen längeren Diskussionen zur Hauptlinie im Anschluß an Def. 4.2.3 samt zugehöriger Benennungen zurück. Sei zunächst erwähnt, daß für b auch die Benennung "Auslastung" gebräuchlich ist, was in der Interpretation "relativer Zeitanteil der Benutzung während des Beobachtungsintervalls" eine immer korrekte Benennung ist (ergibt sich direkt aus der Definition), in der für "Auslastung" ebenfalls gebräuchlichen Interpretation "relative Ausnutzung der Arbeitskapazität der Station" aber wieder nur für arbeitserhaltende Stationen konstanter Bedienkapazität zutrifft.

Aus Def. 4.2.3 erhält man unmittelbar

$$b = \frac{B}{T} = \frac{B}{C}\,\frac{C}{T} = \overline{B}\,c$$

Auslastungsgesetz

kurz das sog. Auslastungsgesetz

(4.2.12) $b = c\ \overline{B}$

Betriebsgesetz

(4.2.12) ist unser erstes Beispiel für ein sog. Betriebsgesetz (operational law): Ein solches legt für jede mögliche Messung einen exakten Zusammenhang zwischen abgeleiteten Größen fest. Die Interpretation

Auslastung = Abgangsrate mittlere Bedienzeit

die ja nach kurzem Nachdenken durchaus einleuchtet, ist mit der jetzt schon gewohnten Vorsicht zu genießen (als mittlere Bedienzeit kann in (4.2.12) nur in den bekannten Fällen $\overline{S}$ substituiert werden).

Betriebsprinzipien

Neben meßbaren Basisgrößen, daraus abgeleiteten Größen und verschiedenen, diese Größen verbindenden Betriebsgesetzen (wir werden weitere kennenlernen) verwendet die Betriebsanalyse letztlich sog. Betriebsprinzipien (operational principles): Diese stellen zusätzliche Hypothesen (Annahmen) über Zusammenhänge zwischen den verwendeten Größen dar, die in manchen Fällen exakt, in vielen Fällen in guter Approximation gelten, immer aber aus den Messungen überprüfbar sind.

Unser erstes Beispiel für ein Betriebsprinzip beruht auf der plausiblen Annahme, daß für großes T (große Beobachtungs-Intervalle) die Zahl der Kundenankünfte an einer Station und die der Kundenabgänge von einer Station sich, relativ gesehen, wenig voneinander unterscheiden werden

Verkehrsflußgleichgewicht

(4.2.13a) $A \approx C$ Prinzip des
(4.2.13b) $a \approx c$ Verkehrsflußgleichgewichts

da ja (A-C)/C mit wachsendem T kleiner werden sollte. Unter Benutzung von (4.2.13) erhalten wir sofort eine zweite (approximative) Form unseres Auslastungsgesetzes (4.2.12)

(4.2.14) $b \approx a\ \overline{B}$

Wir werden im folgenden von der exakten Erfüllung der Betriebsprinzipien ausgehen und daher die "≈"-Zeichen nicht mehr explizit verwenden (obwohl dies ei-

gentlich korrekt wäre und die substituierten "="-Zeichen oft nur approximativ - aber nachprüfbar! - gelten). Das als gültig vorausgesetzte Prinzip des Verkehrsflußgleichgewichts postuliert Gleichheit der Ankunftsrate a und der Abgangsrate c; als Benennung (für beide, als gleichwertig angenommenen Größen) hat sich die Bezeichnung "Durchsatz" eingebürgert.

Durchsatz

Verfeinern wir jetzt unsere Beobachtungen des Stationszustandes. Anstatt die gesamte Belegtzeit B zu messen, also die Summe der Zeiten mit mindestens einem anwesenden Kunden, beobachten wir die Gesamtdauer der Stationszustände bezüglich der Zahl anwesender Kunden, also Größen B(n), n=0,1,2,…

Definition 4.2.15: Basisgrößen Belegtzeiten

Belegtzeiten

B(n) n=0,1,2,…	Summe aller Zeitintervalle (während T), in denen die Station genau n Kunden beherbergte (vgl. auch Anmerkung im Anschluß an Def. 4.2.2).

Aus den Belegtzeiten ergeben sich Belegungsgrade (mit n Kunden) gemäß

Belegungsgrade

(4.2.16) $b(n) := B(n)/T \quad n=0,1,2,...$

Offensichtlich ist

(4.2.17a) $B = \sum_{n>0} B(n)$

(4.2.17b) $T = \sum_{n\geq 0} B(n)$

Eine auf den Belegtzeiten basierende abgeleitete Größe ist

(4.2.18) $J := \sum_{n>0} n\,B(n) \qquad (= \sum_{n\geq 0} n\,B(n))$

welche die Summe aller (ins Beobachtungsintervall fallenden) Verweilzeiten aller im Beobachtungsintervall anwesenden Kunden mißt, und die wir "kumulative Verweilzeit" taufen wollen. Wir können uns die Größe J auch anschaulich machen anhand eines möglichen Geschehnisablaufs an der betrachteten Station:

kumulative Verweilzeit

Belegungsgebirge

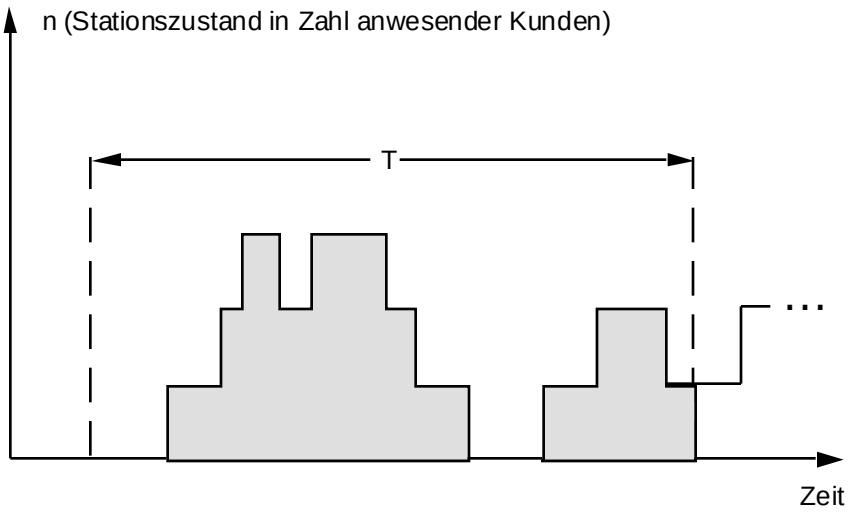


Abbildung 4.2.19: Belegungsgebirge

J entspricht genau der (schraffierten) Fläche unter der Treppenkurve. Die mittlere Höhe der Treppenkurve erhält man als

(4.2.20) $n = J/T$

mittlere Kundenzahl

Für n ist die Benennung "mittlere Kundenzahl" an der Station angemessen, wenn man bedenkt, daß man ungefähr denselben Wert erhält, wenn man die Kundenzahl während T hochfrequent äquidistant mißt und das Mittel bildet.

Wir können andererseits die kumulative Verweilzeit J auf alle beteiligten Kunden aufteilen und so die mittlere Verweilzeit pro Kunden ermitteln. Mitteln wir über die Zahl der Abgänge,

(4.2.21a) $\bar{J} = J/C$

mittlere Verweilzeit

dann ist diese "mittlere Verweilzeit" im Falle einer zu Beginn und zu Ende der Beobachtung leeren Station exakt gleich der "mittleren Kundenverweilzeit", die man auf der Basis individueller Kundenverweilzeiten V_i aller beteiligten Kunden zu

(4.2.21b) $$\overline{V} = \frac{1}{C} \sum_{i=1}^{C} V_i$$

errechnen würde. Ist die Station zu Beginn oder Ende der Beobachtung nicht leer, dann ist $\bar{J}$ nur approximativ gleich $\overline{V}$, aber wieder mit der Tendenz, daß bei wachsendem T der relative Unterschied der beiden Größen sinkt. In jedem Fall erhalten wir aus (4.2.20, 4.2.21a, 4.2.3)

Little's Gesetz

(4.2.22a) $$n = \frac{J}{T} = \frac{J}{C}\,\frac{C}{T} = \bar{J}\,c$$ Little´s Gesetz

ein wesentliches Betriebsgesetz, das eine Sonderform des Ihnen sicher wohlbekannten "Little'schen Theorems" darstellt und das bei Anwendung des Prinzips des Verkehrsflußgleichgewichts (4.2.13) - mit dadurch implizierter Annäherung von $\bar{J}$ und $\overline{V}$ - in die übliche Form

(4.2.22b) $n = \overline{V}\,a$

übergeht.

Gesamtnetz

Nach diesen Überlegungen bzgl. einer beliebigen Einzelstation des zu betrachtenden Verkehrsnetzes wenden wir uns jetzt dem Gesamtnetz zu. Ich erinnere an die weiterhin gültige Maschinenstruktur nach Abb. 4.1.1. Die Anzahl der Stationen sei M. Unsere (potentiell möglichen) Messungen während des Betriebs umfassen folgende Basisgrößen:

Basisgrößen

Definition 4.2.23 Basisgrößen

T Gesamtdauer der Beobachtung

sowie die während des Beobachtungsintervalls an den Stationen gemessenen

- A_j Zahl der Kundenankünfte an Station j; j=1,2,…,M
- C_j Zahl der Kundenabgänge von Station j; j=1,2,…,M
- B_j Belegtzeit der Station j (Gesamtzeit mit mindestens einem Kunden); j=1,2,…,M
- C_{ij} Zahl der Abgänge von Station i, die nach Station j gehen ("Übergangszahl"); i,j=1,2,…,M
- C_{0j} Zahl der Kundeneintritte ins Netz, die nach Station j gehen; j=1,2,…,M
- C_{i0} Zahl der Abgänge von Station i, die das Netz verlassen; i=1,2,…,M

Umgebung

Bei den Benennungen C_{0j}, C_{i0} haben wir die Umgebung als eine Art zusätzliche Station mit Index 0 behandelt; sie tritt natürlich nur bei offenem Netz in Erschei-

nung. Der Vollständigkeit halber können wir aber in diesem Fall mit C_0 noch die Gesamtzahl der "Abgänge aus der Umgebung" (d.h. der Ankünfte am Netz), mit A_0 die Zahl der "Ankünfte in der Umgebung" (d.h. der Abgänge vom Netz) bezeichnen und sinnvollerweise $C_{00}=0$ festsetzen (wir betrachten ja niemanden, der das Netz nicht berührt).

Analog zur Ableitung der Größen 4.2.3 und etwas darüber hinausgehend ergeben sich:

Definition 4.2.24: Abgeleitete Größen

abgeleitete Größen

a_j	:= A_j/T	Ankunftsrate Station j
c_j	:= C_j/T	Abgangsrate Station j
$\underline{b}_j$	:= B_j/T	Belegungsgrad Station j
$\overline{B}_j$	:= B_j/C_j	Mittlere Bedienzeit Station j
c_{ij}	:= C_{ij}/T	Übergangsrate Station i - Station j
h_{ij}	:= C_{ij}/C_i	Relative Übergangshäufigkeit Station i -Station j (relativer Anteil der Übergänge i-j bezüglich der Gesamtabgänge i)

wobei die Benennungen mit der bei der Einzelstation eingeübten Sorgfalt zu verwenden sind.

Im Verhältnis der Stationen zueinander bemerken wir zunächst, daß jede Kundenbewegung einerseits als Abgang von einer Station, andererseits als Zugang zu einer Station gezählt wird, so daß

(4.2.25) $$A_j = \sum_{i=0}^{M} C_{ij} \qquad j=0,1,..,M$$

Setzen wir nun Verkehrsflußgleichgewicht gemäß (4.2.13) voraus, also

(4.2.26) $$A_i = C_i \qquad i=0,1,..,M$$

dann wird aus (4.2.25)

$$C_j = \sum_{i=0}^{M} C_{ij} = \sum_{i=0}^{M} \frac{C_{ij}}{C_i}\ C_i$$

und mit Def. 4.2.24:

(4.2.27a) $$C_j = \sum_{i=0}^{M} h_{ij}\ C_i \qquad j=0,1,..,M$$

bzw.

(4.2.27b) $$c_j = \sum_{i=0}^{M} h_{ij}\ c_i \qquad j=0,1,..,M$$

(4.2.27) trägt den Namen "Formel des Verkehrsflußgleichgewichts" im Netz. Die Formel drückt, im Sinne ihrer Herleitung, eine Beziehung zwischen den beobachteten bzw. aus Beobachtungen abgeleiteten Größen c., h. aus, die mit wachsendem Beobachtungsintervall immer genauer eingehalten wird. Im Sinne zwingender wechselseitiger Abhängigkeiten zwischen den Größen c., h. interpretiert, führt sie zu interessanten, praktisch relevanten Folgerungen.

Verkehrsflußgleichgewicht

Nehmen wir einmal an, wir wüßten über die relativen Übergangshäufigkeiten h. Bescheid, ohne sie aus den Abgangszahlen gemäß Def. 4.2.23 / 4.2.24 bestimmen zu müssen. Dies ist eine durchaus realistische Annahme, liegen doch den Übergangshäufigkeiten die Bearbeitungswünsche der treibenden Lastprozesse zugrunde, die wir in Abschnitt 4.1 durch geeignete Prozeßmuster charakterisiert hatten. Konkreter gesprochen, sollten sich die h. direkt aus den dort ausgedrückten Reihenfolgevorschriften ergeben; so etwa

in Bsp. 4.1.3 zu $h_{terminal\_i,rechner} = 1$
oder in Bsp. 4.1.4 zu $h_{leitwerk,rechenwerk} = r$

Eine naheliegende Frage ist, ob nun (mit dieser Kenntnis der h.) die Abgangszahlen C. aus Gl. (4.2.27) direkt, also ohne Messungen bestimmbar sind; bzw. besser noch die Abgangsraten c., da sie von der Länge T des Beobachtungsintervalls unabhängig sind.

Zusammenhang der Abgangsraten

Das Gleichungssystem (4.2.27b) hat, leicht umgeschrieben, die Form

(4.2.27c) $$\sum_{i=0,\neq j}^{M} h_{ij}\, c_i + (h_{jj}-1)\, c_j = 0 \qquad j=0,1,..,M$$

oder in Vektor-/Matrix-Schreibweise, mit $\underline{c} = (c_0,c_1,\ldots,c_M)^T$ als Vektor der Abgangsraten, $H = (h_{ij};\ i,j=0,1,\ldots,M)$ als Matrix der relativen Übergangshäufigkeiten, $\underline{0}$ als Nullvektor und E als Einheitsmatrix passender Dimension

$$\underline{c}^T(H-E) = \underline{0}^T$$

bzw. letztlich, mit der Abkürzung $H^* = (h^*_{ij}) = H-E$

(4.2.27d) $$\underline{c}^T \underline{H}^* = \underline{0}^T$$

Dies ist ein homogenes lineares Gleichungssystem, das nur dann eine nichttriviale Lösung besitzt, wenn die Determinante von H* verschwindet - also det(H*) = 0 - oder, anders ausgedrückt, der Rang von H* kleiner ist als die Dimension von H* - also rang(H*) < dim(H*) = M+1. Diese Bedingung ist tatsächlich immer erfüllt, da nach Def. 4.2.24

$$\begin{aligned}\sum_{j=0}^{M} h^*_{ij} &= \sum_{j=0}^{M} h_{ij} - 1 \qquad i=0,1,..,M\\ &= \sum_{j=0}^{M} C_{ij}/C_i - 1\\ &= C_i/C_i - 1\\ &= 0 \qquad\qquad i=0,1,..,M\end{aligned}$$

alle Zeilensummen von H* identisch verschwinden. Gehen wir der Lösbarkeit von (4.2.27d) weiter nach: Die Lösung eines homogenen linearen Gleichungssystems mit Koeffizientenmatrix H* und Dimension M+1 besitzt

$$s = (M+1) - \text{rang}(H^*)$$

Freiheitsgrade, d.h. s der Unbekannten (hier der c.) sind frei wählbar, die restlichen M+1-s sind aus diesen s errechenbar. Man kann sich leicht überlegen, daß i.allg. rang(H*)=M: Wäre dem nicht so, dann müßte jede Unterdeterminante der Dimension M aus Matrix H* verschwinden. Nehmen wir beispielsweise die ersten M Spalten von H*, bedeutete dies, daß eine Linearform L existieren müßte derart daß

$$L(h^*_{ij};j=0,1,\ldots,M-1) = 0 \qquad i=0,1,\ldots,M$$

Dies ist aber nicht generell möglich: In L waren die Elemente h^*_{iM} nicht berücksichtigt, d.h. ohne Belang. Diese Elemente sind aber "frei setzbar" in dem Sinne, daß sie für jedes Problem einen anderen Wert haben können (sehen Sie die Definition der h*.. bzw. h.. an) - dies für jede Zeile i je unabhängig. Die Setzung der h^*_{iM} führt über den erkannten Zusammenhang $\Sigma h^*_{ij} = 0$ zu (dem Problem entsprechenden) Änderungen an den h^*_{ij},j=0,1,…,M-1; je zeilenweise unabhängig läßt sich damit obige Forderung an die Linearform L verletzen - es kann ein solches L nicht geben. rang(H*) kann demnach höchstens für einzelne, speziell gelagerte Probleme kleiner als M sein, nicht aber im allgemeinen.

Treffen wir nach diesen Überlegungen eine Fallunterscheidung zwischen offenen und geschlossenen Systemen:

Fallunterscheidung

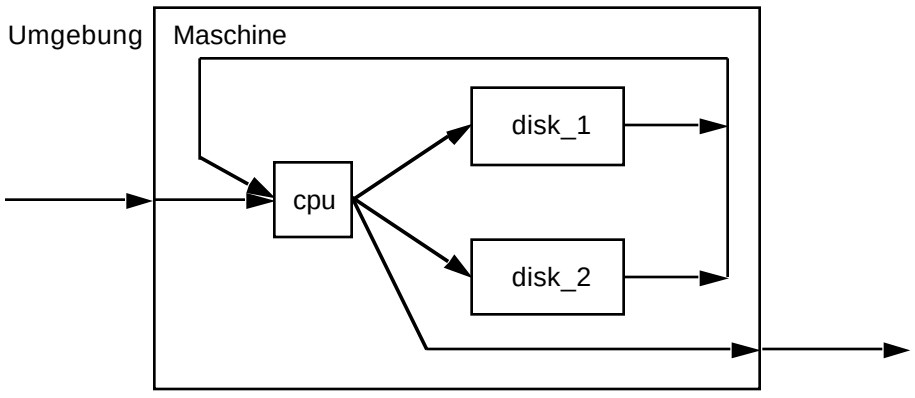
(i) Offenes System:

offenes System

Hier gibt es laut Definition Übergänge zwischen Umgebung und System und meßbare, nicht verschwindende Systemankunfts- bzw. Systemabgangsraten c_0 (Verkehrsflußgleichgewicht!). (4.2.27c) hat (i.allg.) genau einen Freiheitsgrad. Was liegt näher, als c_0 (die von der Umgebung erzwungene, treibende Zugangsrate) als Systemparameter zu betrachten! Damit liegen aber gemäß (4.2.27c) alle anderen Durchsätze (Abgangsrate = Ankunftsrate = "Durchsatz") c_i, i=1,2,…,M, als Funktionen von c_0 fest: Aus Kenntnis von c_0 lassen sich alle Verkehrsflüsse im Netz errechnen. Dies setzte allerdings Kenntnis der relativen Übergangshäufigkeiten h. voraus und implizit ein bisher unerwähntes Betriebsprinzip, das der "Homogenität des Verkehrsflusses": Die relativen Übergangshäufigkeiten h. werden als konstant angesehen, insbesondere unbeeinflußt von der absoluten Verkehrsstärke im Netz. Solange aber, wie angenommen, die h. unmittelbare Folge zwingender Prozeßmuster sind (nach denen sich die Maschine gefälligst zu richten hat!), ist daran sicher nichts auszusetzen.

Testfrage 4.2.28: Ein idealisiertes Rechensystem bestehe aus 3 Stationen: einer Zentraleinheit cpu und zwei Ein / Ausgabeeinheiten disk_1 und disk_2. Die Maschinenstruktur habe inklusive aller Übergangsmöglichkeiten folgendes Aussehen

Testfrage



Die Maschine werde von einem Strom von Stapelaufträgen belastet, deren jeder eine Folge von Anforderungen der Form cpu-disk_i-cpu-disk_j-… stellt. Jeder Auftrag berührt cpu genau 31-mal. Die E / A-Aufträge gelten in zwei Drittel aller Fälle der Einheit disk_1, die restlichen der disk_2. Ermitteln Sie die Durchsätze an cpu, disk_1, disk_2 in Abhängigkeit von der Ankunftsrate der Stapelaufträge.

geschlossenes System

(ii) Geschlossenes System:
Hier gibt es laut Definition keine Übergänge Umgebung / System, d.h. $c_0=a_0=0$. (4.2.27c) erhält, entsprechend reduziert, die Form

$$\sum_{i=1,\neq j}^{M} h_{ij}\, c_i + (h_{jj}-1)\, c_j = 0 \qquad j=0,1,..,M$$

Analog den Überlegungen zu (4.2.27d) ist dies ein homogenes lineares Gleichungssystem der Dimension M mit Rang M-1. Zur Ausfüllung des verbleibenden Freiheitsgrades bietet sich kein "natürlicher" Parameter an (wie das im Fall des offenen Systems die Systemzugangsrate c_0 war). Somit ist auch keine eindeutige Lösung des Gleichungssystems verfügbar. Wie können wir die obigen Beziehungen dennoch nutzen?

Betrachten wir offene und geschlossene Systeme wieder gemeinsam, also Gleichungssysteme (4.2.27). Greifen wir willkürlich eine Station j* heraus und definieren die Größen

(4.2.29a) $$X_j = \frac{C_j}{C_{j^*}} \qquad j=0,1,..,M$$

Wegen der Definitionen 4.2.24 und wegen des angenommenen Verkehrsflußgleichgewichts gilt offenbar auch

(4.2.29b) $$X_j = \frac{A_j}{A_{j^*}} = \frac{c_j}{c_{j^*}} = \frac{a_j}{a_{j^*}}$$

relative Besuchszahlen

Nach Def. (4.2.29a) ist X_j die "Zahl der Abgänge von Station j pro Abgang von Station j*" und trägt den Namen "relative Besuchszahl" (relative visit count). Substituiert man in (4.2.27c) die c_j entsprechend (4.2.29b), also gemäß

(4.2.30) $$c_j = X_j\ c_{j^*} \qquad j=0,1,..,M$$

dann ergibt sich

$$\sum_{i=0,\neq j}^{M} h_{ij}\, X_i\, c_{j^*} + (h_{jj}-1)\, X_j\, c_{j^*} = 0 \qquad j=0,1,..,M$$

bzw. ($c_{j^*} \neq 0$ vorausgesetzt)

(4.2.31a) $$\sum_{i=0,\neq j}^{M} h_{ij}\, X_i + (h_{jj}-1)\, X_j = 0 \qquad j=0,1,..,M$$

wozu noch das offensichtlich richtige

(4.2.31b) $$X_{j^*} = \frac{C_{j^*}}{C_{j^*}} = 1$$

tritt. (4.2.31a) ist mit (4.2.27c) identisch bis auf die zu errechnenden Unbekannten (dort Durchsätze c., hier relative Besuchszahlen X.). Auf (4.2.31a) treffen alle bzgl. (4.2.27c) gemachten Bemerkungen zu. Insbesondere läßt sich wegen der linearen Abhängigkeit der Gleichungen (4.2.31a) eine beliebige davon streichen - und durch (4.2.31b) ersetzen, wodurch (4.2.31) insgesamt ein nicht-homogenes lineares Gleichungssystem darstellt, das i.allg. (bis auf numerische Sonderfälle) eine eindeutige Lösung besitzt.

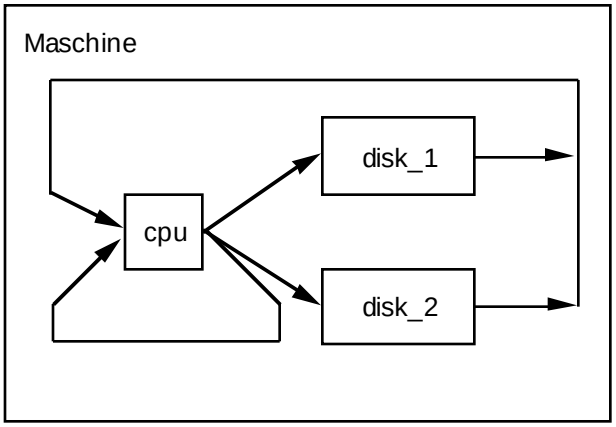
Wir können also bei Kenntnis der Matrix H der Übergangshäufigkeiten alle relativen Besuchszahlen X. (bezüglich einer beliebigen Station j*) explizit bestimmen.

Bestimmung der relativen Besuchszahlen

In einer meßtechnischen Anwendung der Beziehungen bedeutet dies, daß die Messung der Abgangsrate an einer beliebigen Station j* genügt, um alle Durchsätze mittels (4.2.30) zu erhalten (die X_j konnten wir ja errechnen!).

Testfrage

Testfrage 4.2.32 Betrachten Sie erneut Testfrage 4.2.28. In Abänderung der Aufgabe nehmen wir an, daß das System (vom Betriebssystem her) über einen vorgegebenen maximalen Multiprogrammierungsgrad MP in der Kundenaufnahme beschränkt ist und "unter Vollast" betrieben werde. Der Modellierungs-"Trick", der hierfür üblicherweise eingesetzt wird, besteht in folgender Abänderung der Maschine



und der Annahme, daß ein "abgearbeiteter" Stapelauftrag (der eigentlich das System verläßt) wegen bestehender Vollast sofort durch einen neuen gleicher Art ersetzt wird (im Modell also "wieder von vorn anfängt", im speziellen Beispiel von cpu-Ausgang nach cpu-Eingang geht). Alle sonstigen Annahmen von Testfrage 4.2.28 bleiben bestehen. Sie messen 40 Abgänge / ZE an disk_1. Wie sind die Durchsätze aller Stationen? Wieviel Stapelaufträge werden je ZE abgewickelt? Sei bei dieser Messung MP=5 festgehalten. Wie lange dauert im Mittel die gesamte Bearbeitung eines Stapelauftrags (denken Sie an Little!)?

Belastbarkeit

Das Konzept der relativen Besuchszahlen führt in seiner praktisch wohl wesentlichsten Anwendung zu Erkenntnissen über die maximale Belastbarkeit des Netzes und über jene Stationen, welche die Belastbarkeit begrenzen. Kehren wir zum Auslastungsgesetz (4.2.12) zurück, das ja für jede Station des Netzes gilt, so daß also

$$b_j = \bar{B}_j c_j \qquad j=1,2,\ldots,M$$

Gemäß (4.2.30) ist damit auch

$$b_j = \bar{B}_j X_j c_j^* \qquad j=1,2,\ldots,M$$

bzw.

(4.2.33a) $$\frac{b_j}{\bar{B}_j X_j} = c_j^* \qquad j=1,2,..,M$$

Das bedeutet zum einen, daß der linksstehende Ausdruck systemweit invariant ist:

systemweite Invarianz

(4.2.33b) $$\frac{b_1}{\bar{B}_1 X_1} = \frac{b_2}{\bar{B}_2 X_2} = \ldots = \frac{b_M}{\bar{B}_M X_M}$$

Berücksichtigen wir weiter, daß laut Definition 4.2.3 kein b_j den Wert 1 übersteigen kann, dann bedeutet (4.2.33a) zum anderen, daß

(4.2.33c) $$c_{j^*} \le \frac{1}{\overline{B}_j X_j} \qquad j=1,2,..,M$$

physikalische Durchsatzgrenze

das heißt

$$c_{j^*} \le \min_{j=1,2,..,M} \frac{1}{\overline{B}_j X_j}$$

gelten muß im Sinne einer physikalischen Grenze des Durchsatzes.

Voraussetzungen für Flaschenhalsbetrachtungen

Für die folgenden Überlegungen setzen wir voraus, daß alle Stationen des Netzes arbeitserhaltend und von konstanter Bedienkapazität im Sinne der Def. 4.2.8 sind. Damit dürfen wir für alle Stationen

$$\overline{B}_i = \overline{S}_i \qquad i=1,2,\ldots,M$$

setzen, da ja (wie vorne diskutiert) die (aus Messungen ermittelten) Größen $\overline{B}_.$ mit den (auf den Bedienwünschen $\overline{W}_.$ basierenden) mittleren Bedienzeiten $\overline{S}_.$ übereinstimmen. In ähnlicher Weise wie anläßlich der Entwicklung der Gleichungen (4.2.27) für die Übergangshäufigkeiten $h_.$ diskutiert, sind jetzt auch die mittleren Bedienzeiten $\overline{S}_.$ in vielen Fällen bereits aus der Beschreibung der treibenden Lastprozesse bekannt, u.U. in Prozeßmustern genau festgelegt. Selbst wenn die Größe der $\overline{S}_.$ nicht bekannt ist, dürfen wir doch davon ausgehen, daß sie allein auf den Kundenwünschen beruhen, daher von Gegebenheiten des Betriebs unbeeinflußt sind und von einer Messung zur anderen (bei u.U. unterschiedlichen Verkehrsstärken im Netz) unverändert bleiben.

Trennen wir nach diesen Vorbemerkungen wieder unsere Diskussion offener von der geschlossener Systeme:

offenes System

(i) Offenes System:
Wir hatten erkannt, daß der gesamte Verkehr im Netz durch den (von außen erzwungenen) Zugangsstrom der Rate c_0 bestimmt wird. Wir tragen dieser Sonderrolle der Umgebung dadurch Rechnung, daß wir für alle Gleichungen ab (4.2.29) $j^*=0$ wählen. Insbesondere erkennen wir aus (4.2.30), daß alle Stationsdurchsätze gemäß

$$c_i = X_i \cdot c_0 \qquad i=1,2,\ldots,M$$

linear mit c_0 wachsen (die X_i sind auf $j^*=0$ bezogen und damit $X_0=1$), solange dies physikalisch möglich ist. Wir entnehmen jetzt der Gl. (4.2.33c), daß

$$c_0 \le \min_{i=1,2,..,M} \frac{1}{\overline{S}_i X_i}$$

wobei wir $\overline{S}_.$ für $\overline{B}_.$ substituiert haben wegen der Annahme der Identität der beiden Größen sowie der des fehlenden Einflusses der Verkehrsstärke auf die $\overline{S}_.$ Die physikalisch maximale Zugangsrate c_0 (der maximale Systemdurchsatz) ist damit aus den mittleren Bedienzeiten $\overline{S}_.$ und den relativen Besuchszahlen $X_.$ direkt bestimmbar (wobei die Bestimmung der $X_.$ ihrerseits die Kenntnis der Übergangshäufigkeiten $h_.$ voraussetzte).

maximaler Systemdurchsatz

Testfrage

Testfrage 4.2.34: Was geschieht, wenn die Zugangsrate c_0 über den ermittelten Maximalwert hinaus gesteigert wird?

Stations-durchsätze

Führen wir uns die Abhängigkeit der Stationsdurchsätze c_i, i=1,2,…,M, von der Systemzugangsrate c_0 der Deutlichkeit halber graphisch vor Augen, wobei als Beispiel M=3 gewählt ist und $\bar{S}_1X_1 > \bar{S}_2X_2 > \bar{S}_3X_3$, sowie $X_2 > X_1$, $X_3 < X_1$ gelte (s. Abb. 4.2.35). Alle Stationsdurchsätze steigen linear mit c_0, alle besitzen (entsprechend der Begrenzung von c_0) einen spezifischen, maximal erreichbaren Wert.

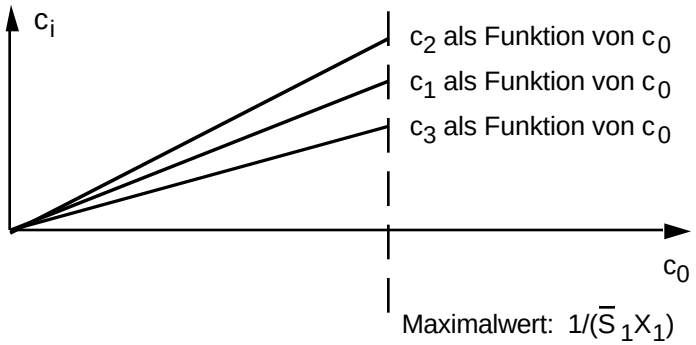


Abbildung 4.2.35: Durchsätze als Funktionen der Zugangsrate

Stations-auslastungen

Die Darstellung wird noch aufschlußreicher, wenn wir statt der Durchsätze c. die Auslastungen b. in ihrer Abhängigkeit von c_0 betrachten. Laut (4.2.33a) gilt ja

$$b_i = \bar{S}_i X_i c_0 \qquad i=1,2,\ldots,M$$

was wegen der Beschränkung von c_0 durch die Station mit maximalem $\bar{S}.X.$ die Auslastungen aller anderen Stationen auf Werte

(4.2.36)
$$b_j \le \dot{\bar{S}}_i X_i c_{0,max} = \frac{\dot{\bar{S}}_i X_i}{(\bar{S}.X.)_{max}} \le 1$$

beschränkt. Graphisch mit obigen Beispielsannahmen:

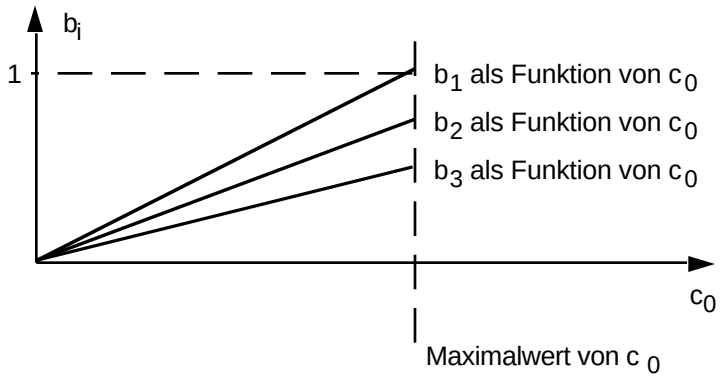


Abbildung 4.2.37: Auslastungen als Funktionen der Zugangsrate

Interpretation

Flaschenhals

Wir interpretieren: Alle Stationsauslastungen b. steigen linear mit dem Systemdurchsatz c_0 bis zu einem Maximalwert. Für (i.allg. genau:) eine Station k wird b_k=1 erreicht. Diese Station (sie weist das maximale $\bar{S}X$ aller Stationen auf) begrenzt eine weitere Erhöhung von c_0: Sie wird daher als "Flaschenhals" (englisch: "bottleneck") des Systems bezeichnet. Alle anderen Stationen sind zwingend auf Auslastungswerte $b_i < 1$, i=1,2,…,M, i≠k, festgelegt. Diese Interpretation hat offensichtlich zentrale praktische Bedeutung, wie in folgender Testfrage überdeutlich werden sollte.

Testfrage

Testfrage 4.2.38: Eine Leistungsverbesserung eines Rechensystems läßt sich in vielen Fällen durch Ersatz einer Maschinenkomponente durch eine funktional gleichwertige, aber schnellere, andere erreichen. In unserer Modellwelt heißt dies, daß die Arbeitskapazität r_i einer Station i vergrößert wird, so daß (bei gleichbleibendem mittlerem Bedienwunsch $\overline{W}_i$) die mittlere Bedienzeit $\overline{S}_i$ sinkt. Können Sie durch eine derartige Maßnahme immer den maximalen Systemdurchsatz c_0 erhöhen?

geschlossenes System

(ii) Geschlossenes System:
Wir hatten schon diskutiert, daß sich bei geschlossenen Systemen unter den betrachteten Größen keine als natürlicher Parameter anbietet, um den verbleibenden Freiheitsgrad des Gleichungssystems (4.2.27) auszufüllen und es einer Lösung zuzuführen. Andererseits wissen wir (machen uns dies aber erst jetzt richtig bewußt), daß in einem geschlossenen System aufgrund der fehlenden Zu- und Abgänge eine konstante Anzahl von Kunden zirkuliert. Sei N die Zahl dieser Kunden. So wie beim offenen System der als unabhängige Größe angenommene Systemdurchsatz c_0 die Verkehrsdichte im Netz bestimmte, sollte man annehmen, daß beim geschlossenen System die Zahl N insgesamt anwesender Kunden (die "Gesamtpopulation") ihren Einfluß auf die Verkehrsdichte im Netz hat, sie evt. sogar völlig bestimmt. Fassen wir also im geschlossenen System die Gesamtpopulation N als zentralen Parameter auf und überlegen uns ihren Einfluß auf das Verkehrsgeschehen. Dazu folgender

Gesamtpopulation

Satz 4.2.39: Sei mit $b_i(N)$ die Auslastung der Station i, i=1,2,…,M, eines geschlossenen Systems mit einer Gesamtpopulation N, N=0,1,2,…, bezeichnet (und mögen die für den jetzigen Kontext getroffenen Annahmen zutreffen!). Dann gilt

$$b_i(N) \le b_i(N+1) \qquad i=1,2,\ldots,M;\ N=0,1,2,\ldots$$

und für mindestens eine Station k

$$b_k(N) < b_k(N+1)$$

In Worten: Bei Einbringen eines zusätzlichen Kunden in das (ansonsten unveränderte) System wird die Auslastung keiner Station (im Vergleich zu den vor Einbringen beobachteten Werten) sinken, für mindestens eine Station steigen.

Der Satz entspricht sicher der Intuition. Der interessierte Leser findet einen Beweis (allerdings in stochastischem Kontext) z.B. in ChLa74. Der Beweis ist länglich und in seiner Beweisidee hier uninteressant, so daß wir den Satz ohne Beweis übernehmen.

Erinnern wir uns, daß nach (4.2.33b) die Größen $b_j / (\overline{B}_j X_j)$ systemweit invariant waren, z.B. den Wert K hatten. Dabei hatte (unausgesprochen) die Gesamtpopulation irgendeinen festen Wert. Bei Variation der Population N muß sich, wenn wir laut Satz 4.2.39 von durch diese Variation veränderten Durchsätzen $b_j(N)$ ausgehen (und bei entsprechend getroffener Annahmen unveränderlicher $\overline{B}_\cdot, X_\cdot$), die Systeminvariante K ebenfalls ändern, so daß aus (4.2.33b) die Beziehungen

$$\textbf{(4.2.40)} \qquad \frac{b_j(N)}{\overline{B}_j X_j} = K(N) \qquad j=1,2,\ldots,M;\ N=0,1,2,\ldots$$

folgen. Da nach wie vor (siehe Diskussion des offenen Systems) keine Stationsauslastung den Wert 1 übersteigen kann, ist K(N) in seiner Größe beschränkt.

Wenn wir entsprechend Satz 4.2.39 und (4.2.33b) von monoton mit N steigenden Auslastungen ausgehen, sollte auch K(N) gemäß (4.2.40) monoton mit N steigen und sein Maximum bei über alle Grenzen wachsendem N erreichen. Schreiben wir kurz K(∞) für lim(K(N);N->∞), dann erhalten wir zusammengefaßt

(4.2.41) $$K_\infty \le \min_{j=1(1)M} \frac{1}{\overline{B}_j X_j}$$

und daraus mit (4.2.40) und der Abkürzung

$$b_i(\infty) = \lim(b_i(N);N\text{->}\infty)$$

die Beziehung:

(4.2.42) $$b_i(\infty) = \overline{B}_i X_i K(\infty) \qquad i = 1,2,..,M$$
$$\le \overline{B}_i X_i \min_{j=1(1)M} \frac{1}{\overline{B}_j X_j}$$

Unter der weiteren Annahme, daß für "unendlich großes" N die Grenzwerte tatsächlich erreicht werden, sollten wir die folgende Skizze als prinzipielle Abhängigkeit der Stationsauslastungen von der Gesamtpopulation festhalten, wobei (wie für Abb. 4.2.35 / 4.2.37) M=3, $\overline{S}_1X_1 > \overline{S}_2X_2 > \overline{S}_3X_3$ gelte. *Auslastungen*

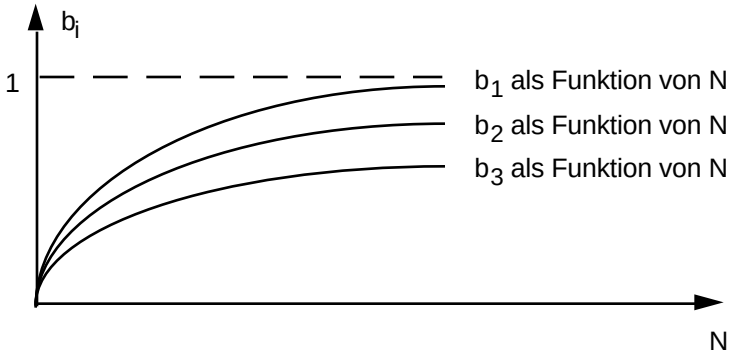


Abbildung 4.2.43:Auslastungen als Funktionen der Population

Diese Skizze ist analog der Abb. 4.2.37 des offenen Systems zu verstehen. Beide Abbildungen zeigen die Abhängigkeit der Stationsauslastungen vom zentralen, verkehrsbestimmenden Systemparameter (dort Systemdurchsatz c_0, hier Systempopulation N) auf, wobei allerdings die Funktionsverläufe der Abb. 4.2.43 nur als Prinzipskizzen zu verstehen sind (wir wissen nur, daß monotones Steigen mit endlicher Asymptote vorliegt). Auch die Interpretation der Skizze ist analog der von Abb. 4.2.37 zu sehen: Alle Stationsauslastungen steigen monoton mit der Systempopulation N gegen einen Grenzwert. Dieser liegt für (i.allg. genau:) eine Station k bei $b_k(\infty)=1$. Diese Station (sie weist das maximale $\overline{S}\,V$ aller Stationen auf) begrenzt die Auslastungen aller Stationen im Netz und wirkt, wie beim offenen System, als Flaschenhals; insbesondere sind die Auslastungen der Stationen i ($i \neq k$) damit auf Werte $b_i<1$ festgelegt. Die zentrale praktische Bedeutung dieser Aussage ist offensichtlich. *Interpretation*

Folgefragen

Wir haben mit diesen Ergebnissen ein beträchtliches Verständnis für die Vorgänge in einem (als Verkehrsnetz modellierten) Rechensystem gewonnen. Gleichzeitig eröffnen sich bisher unbeantwortete Folgefragen: Wie ist der genaue Verlauf der $b_i(N)$ im geschlossenen Netz? Und (ein Fragenkomplex, den wir bisher ganz ausgespart haben) wie bestimmen wir die Verweilzeiten (d.h. die Abwicklungszeiten für eine Anforderung) der Kunden an den Stationen? Ferner: Was ist, wenn es mehrere "Sorten" von Kunden (d.h. mehrere Prozeßmuster) gibt? Auch für diese Fragen stellt die Betriebsanalyse Antworten bereit, die über zusätzlich eingeführte Betriebsprinzipien gewonnen werden. Die wesentlichen weiterführenden Konzepte sind die des "Zustands" des dynamischen Systems Verkehrsnetz (grob gesprochen die Frage "wieviele Kunden sind an welcher Station"?) und die der Verfolgung der Zustandsänderungen. Diese Konzepte werden wir in den folgenden Abschnitten kennenlernen.