

Wissenschaftliche Berichte des
Institutes für Fördertechnik und Logistiksysteme
der Universität Karlsruhe (TH)
Prof. Dr.-Ing. Dr. h.c. D. Arnold (Hrsg.)

Habilitationsschrift

Bedientheoretische Methoden als Hilfsmittel der Materialflußplanung

Eingereicht von
Dr.-Ing. Dipl. Wirtsch.-Ing. Kai Furmans
Kantstr. 7
76137 Karlsruhe

28. März 1999

Vorwort

Die vorliegende Arbeit entstand überwiegend während meiner Tätigkeit als Abteilungsleiter des Forschungsbereichs Materialfluß am Institut für Fördertechnik und Logistiksysteme der Universität Karlsruhe (IFL).

Unter der Institutsleitung von Prof. Dr.-Ing. Dr. h.c. D. Arnold liegt ein Schwerpunkt der Forschungsarbeit am IFL auf der Anwendung bedientheoretischer Modelle für Fragestellungen aus der Materialflußplanung. Diese Forschungsarbeit und der spätere Transfer der dabei erzielten Forschungsergebnisse in die industrielle Praxis haben gezeigt, daß mit Hilfe stochastischer bedientheoretischer Modelle viele Phänomene im industriellen Materialfluß analysiert und erklärt werden können.

Es hat sich jedoch auch gezeigt, daß viele potentielle Anwender nicht über Verständnis und Grundkenntnisse verfügen, um stochastische und insbesondere bedientheoretische Modelle anwenden zu können. Gelingt es, die Verbindung zwischen mathematischen Modellen und deren Anwendung in der Materialflußlehre herzustellen, so ist das Interesse bei Praktikern und Studenten gleichermaßen hoch. Die weitere Auseinandersetzung mit bedientheoretischen Modellen führt zu einem wesentlich verbesserten Systemverständnis für stochastische Modelle.

Deshalb wird seit dem Wintersemester 1994/1995 eine Wahlveranstaltung „Analytische Methoden in der Materialflußplanung“ im Rahmen des Vorlesungsangebotes des Institutes für Fördertechnik und Logistiksysteme der Universität Karlsruhe für Hörer aus den Fakultäten Maschinenbau und Wirtschaftsingenieurwesen angeboten.

Der stetig zunehmende Zahl der Hörer dieser Vorlesung und Professor Arnold haben mich ermutigt, den Inhalt der Vorlesung lehrbuchartig in der vorliegenden Form aufzubereiten. Ziel ist eine Darstellung der wesentlichen Grundlagen von bedientheoretischen Modellen die für die Materialflußlehre relevant sind. Damit soll für den Leser ein Einstieg in die weiterführende bedientheoretische Literatur und ein grundsätzliches Verständnis bedientheoretischer Effekte möglich werden.

Karlsruhe, im Winter 1998

Kai Furmans



Inhaltsverzeichnis

1	Einführung	1
1.1	Die Rolle bedientheoretischer Modelle in der Materialflußplanung	2
1.2	Einsatz bedientheoretischer Modelle bei der Optimierung von Materialflußsystemen	4
1.3	Das Zusammenwirken von bedientheoretische Methoden und computergestützter Simulation	5
1.4	Struktur der weiteren Vorgehensweise	5
2	Einzelemente von Materialflußsystemen	7
2.1	Grundbestandteile eines bedientheoretischen Modells	8
2.2	Deterministisches Modell eines Materialflußelementes	10
2.3	Operationelle Analyse	12
2.4	Exkurs: Stochastische Prozesse	16
2.5	Das Bediensystem $M M 1$ -FCFS	19
2.6	Das Bediensystem $M M m$ -FCFS	22
2.7	Bediensysteme mit mehrphasigen Verteilungen	23
2.7.1	Mehrphasige Verteilungen	24
2.7.2	Das Bediensystem $\lambda(k) C_n 1$	31
2.7.3	Modellierung des Ausfalls- und Reparaturverhaltens eines FTS mit Hilfe des Bediensystems $\lambda(k) C_n 1$	34
2.8	Little's Gesetz	36
2.9	Das Bediensystem $M G 1$ -FCFS	38
2.9.1	Modellierung von Verzweigungselementen mit Hilfe von $M G 1$ -Bediensystemen	41
2.9.2	Losgrößenoptimierung mit Hilfe eines $M G 1$ -Bediensystems	42
2.10	Das $M G 1$ -Bediensystem mit Prioritäten	49

2.10.1	$M G 1$ -Bediensystem mit nicht unterbrechenden Prioritäten	53
2.10.2	Modellierung von Zusammenführungselementen mit Hilfe von $M G 1$ -Bediensystemen mit nicht unterbrechenden Prioritäten	56
2.10.3	$M G 1$ -Bediensystem mit unterbrechenden Prioritäten	58
2.10.4	Modellierung von Störungen mit Hilfe unterbrechender Prioritäten	59
2.11	Approximationen für $M G m$ -Bediensysteme	61
2.12	Näherungen und Grenzen für $G G 1$ -FCFS Bediensysteme	62
2.13	Approximationen für $G G m$ -Bediensysteme	69
3	Bediensystemnetzwerke	71
3.1	Beschreibung von Bediensystemnetzwerken	72
3.2	Offene Bediensystemnetzwerke mit exponential verteilten Zeiten	74
3.2.1	Die Produktformlösung	74
3.2.2	Beispiel: Analyse eines Bediensystemnetzwerkes	81
3.2.3	BCMP-Netzwerke	82
3.3	Näherungsverfahren für offene Bediensystemnetzwerke	83
3.3.1	Exkurs: Erneuerungsprozesse und Punktprozesse	85
3.3.2	„QNA“, das Berechnungsverfahren von Whitt	86
3.4	Geschlossene Bediensystemnetzwerke mit exponentialverteilten Bedienzeiten	94
3.4.1	Produktformlösung für geschlossene Bediensystemnetzwerke	96
3.4.2	Mittelwertanalyse, ein Berechnungsverfahren für geschlossene Bediensystemnetzwerke	102
3.5	Geschlossene Bediensystemnetzwerke mit generell verteilten Bedienzeiten	105
3.5.1	Das Verfahren der Antwortzeiterhaltung	106
3.5.2	Das Verfahren von Marie	108
3.6	Modellierung von Produktionssystemen mit bedientheoretischen Modellen	111
3.6.1	Hinweise zur Gestaltung von Materialflußsystemen	111
3.6.2	Optimierung der Losgrößen in Werkstattfertigungen	112
3.6.3	Weitere Anwendungsbereiche	119

4 Zusammenfassung und Ausblick	127
Symbolverzeichnis	131
Literaturverzeichnis	133

1 Einführung

Die ersten bedientheoretische Modelle wurden Anfangs des zwanzigsten Jahrhunderts durch den dänischen Ingenieur Erlang entwickelt, um die Zahl der in einer Vermittlungsstelle zu installierenden Telefonleitungen festzulegen. Ausgehend von der Beobachtung, daß eine Dimensionierung der Leitungen alleine auf der Basis der mittleren Zahl gleichzeitiger Gespräche nicht ausreichend ist, wurde durch ihn eine Methode entwickelt, die es erlaubt, die stochastischen Schwankungen des Gesprächsaufkommens zu erfassen und die Auswirkungen der Investitionsentscheidungen (Anzahl Leitungen) auf die Wartezeit der Kunden und damit den Kundenservice abzubilden.

Die Anwendung bedientheoretische Methoden blieb lange auf das Gebiet der Telekommunikation begrenzt, die Einsatzmöglichkeit bei der Planung von Materialflußsystemen wurde in Deutschland erst Mitte der siebziger Jahre durch Gudehus erkannt. Durch zunehmende Ansprüchen der Kunden in Bezug auf Reaktionsfähigkeit und Variantenvielfalt auch im Bereich der industriellen Produktion und Distribution spielen externe stochastische Einflüsse, eine größere Rolle als je zuvor. Die Beantwortung der Frage „wie schnell nach Auftragserteilung erhalte ich meine Ware?“ wird in Zukunft häufiger erfolgen müssen. Gleichzeitig wird die Beherrschung interner stochastischer Einflüsse (z.B. Ausfallsverhalten von Anlagen) durch die engere Vernetzung der Prozeßkette in der Produktion und Distribution immer wichtiger.

Vor diesem Hintergrund ist der Umgang mit stochastischen Modellen auch für Ingenieurstätigkeiten im Bereich Materialfluß immer wichtiger. Die in den letzten zwanzig Jahren weiterentwickelten bedientheoretischer Methoden können für den Planungsingenieur im Materialflußbereich ein Werkzeug zur Beantwortung von Fragen in Bezug auf die Leistungsfähigkeit der durch ihn geplanten Systeme werden. Aus diesem Grund werden in dieser Arbeit die wichtigsten Basismethoden und Modelle vorgestellt und die Anwendung dieser Modelle auf Fragestellungen aus dem Materialflußbereich erläutert, um damit die Grundlage für einen erweiterbaren Werkzeugkasten stochastischer Modelle zu legen.

1.1 Die Rolle bedientheoretischer Modelle in der Materialflußplanung

Die Produktion, der Vertrieb und die Entsorgung von Gütern geschieht durch die Verknüpfung unterschiedlichster Tätigkeiten zu einem Gesamtprozeß. Sofern diese Schritte an mehr als einem Ort durchgeführt werden, erfordert die Verkettung der an den Gütern durchzuführenden Tätigkeiten, wie z.B. Bearbeiten, Montieren, Lagern und Prüfen, die Bereitstellung und den Betrieb von Materialflußsystemen.

Die Aufgabe der Materialflußsysteme ist die Verknüpfung der räumlich getrennten Orte an denen oben genannte Tätigkeiten durchgeführt werden und nötigenfalls die Überbrückung der Zeitlücke zwischen Beendigung einer Tätigkeit und Beginn der nachfolgenden Aktivität am gleichen Objekt.

Werden alle Tätigkeiten zu einer Zeit, an einem Ort und von einer Person oder Einrichtung durchgeführt, so ist das zugehörige Materialflußsystem einfach.

Mit zunehmender Spezialisierung von Unternehmen oder Unternehmensteilen auf einzelne Schritte, insbesondere des Produktionsprozesses, wächst die Komplexität des notwendigen Materialflußsystems und damit auch die Bedeutung des Materialflußsystems für den effiziente Ablauf des Gesamtprozesses. Eine notwendige Voraussetzung für einen wirtschaftlichen Betrieb von Materialflußsystemen ist eine aufgabengerechte Gestaltung des Materialflußsystems.

Ein aufgabengerecht gestaltetes Materialflußsystem kann durch eine fundierte Planung des Systemes erreicht werden. Ausgehend von einer quantitativen (z.B. Kosten, Durchlaufzeiten, Durchsatz, Verfügbarkeit der Betriebsmittel) und qualitativen Zielsetzung (Flexibilität hinsichtlich Produktvarianten, Standardisierung von Verfahren etc.), werden in der Materialflußplanung mehreren Planungsstufen in zunehmender Konkretisierung Alternativen entwickelt, die eine bestmögliche Erreichung dieser Ziele versprechen.

Zur Erzielung guter Planungsergebnisse ist eine frühzeitige Fokussierung auf eine Lösungsalternative zu vermeiden. Da der Planungsaufwand und damit die Planungskosten mit zunehmendem Detaillierungsgrad ansteigen, ist es jedoch im allgemeinen nicht ökonomisch, alle Alternativen bis zur Stufe der Detailplanung auszuarbeiten. Aus diesem Grund sind nach den vorgelagerten Planungsstufen (z.B. Grobplanung oder Realplanung) alle Alternativen im Hinblick auf die Zielerreichung und den damit verbundenen Aufwand zu bewerten, um im nächsten Schritt die weitere Planung auf erfolgversprechende Varianten zu konzentrieren.

Bei der Bewertung von Planungsalternativen, insbesondere in einem frühen Planungsstadium, zu dem viele Varianten auf der Basis einer noch nicht detaillierten Planungsdatenbasis zu beurteilen sind, ist eine Prognose der mit dieser Planungsalternative zu erzielenden Leistungskennwerte notwendig. Bedientheoretische Mo-

delle können hier einen wesentlichen Beitrag bei der Prognose von Leistungskennwerten der Planungsalternativen leisten.

Dies ist zum Einen darin begründet, daß die Eingangsparameter statistische Größen sind, die in frühen Planungsstadien einfacher erhoben oder geschätzt werden können. Diese Parameter können auch mit geringem Aufwand manipuliert werden, wenn es darum geht, das Verhalten eines Materialflußsystems unter verschiedenen Umgebungsbedingungen (z.B. wachsender Auslastung, wechselndem Produktmix) zu untersuchen.

Zum Anderen erlauben bedientheoretische Modelle, die Verfügbarkeit geeigneter Auswertungswerkzeuge vorausgesetzt, eine sehr schnelle Ermittlung von Leistungskennwerten für eine Planungsalternative. Diese Kennwerte können dann monetär bewertet werden, um das Erreichen von Kostenzielen (z.B. für Umlaufbestände) zu messen oder können dazu dienen, die Erreichung von Leistungszielen wie z.B. Systemdurchsatz oder die Einhaltung von wesentlichen Randbedingungen, wie z.B. Durchlaufzeiten oder Flächenverbrauch, zu kontrollieren.

Die Kennwertermittlung für ein Materialflußsystem mit Hilfe bedientheoretischer Methoden erfordert die Modellierung der wesentlichen Systemmerkmale der Planungsalternative in einem bedientheoretischen Modell. Für den Modellierungsschritt stehen Bausteine mit aus der Literatur bekannten Eigenschaften zur Verfügung. Die Kennwertberechnung für diese Bausteine kann bei bekannten Eingangsparametern mit Hilfe analytischer Funktionen erfolgen. Durch die Untersuchung dieser Funktionen ist es auch möglich, allgemeingültige Vorhersagen über das Verhalten dieser Bausteine zu treffen. Auf diese Weise entsteht bei den Planern, die sich mit bedientheoretischen Methoden und Modellen beschäftigen, Basiswissen und Systemverständnis, das es ermöglicht, abschätzende Aussagen über das Verhalten eines bedientheoretisch modellierbaren Systemes zutreffen, ohne Berechnung anzustellen.

Die Gestaltung eines Materialflußsystems ist häufig eine Teilaufgabe einer umfassenderen Planungsaufgabe. Aus diesem Grund sind während der Entwicklung der Planungsvarianten Informationen als Eingangsinformation für andere Teilplanungsaufgaben zu liefern oder Informationen anderer Teilplanungsaufgaben zu verarbeiten.

Vorgehensweisen, die eine solche enge Kooperation von Fachleuten verschiedener Ausrichtungen vorsehen, wie z.B. Simultaneous Engineering, können vom Einsatz bedientheoretischer Methoden bei der Analyse von Materialflußsystemen profitieren, da mit Hilfe dieser Methoden kurzfristige Tendenzaussagen möglich werden und so Einflüsse des Materialflußsystems auf das Gesamtsystem frühzeitig in die Planungsüberlegungen einbezogen werden können. Ebenso verhält es sich mit Anregungen zur Gestaltung des Gesamtsystems, die aus einer Abschätzung der Leistungsfähigkeit des Materialflußsystems kommen können. Neben der Auswahl und Bewertung von Planungsvarianten ist auf diese Weise auch eine Lenkung

hinsichtlich Variantengestaltung möglich.

Auf diese Weise kann der interaktive Ansatz, der der oben genannten simultanen Vorgehensweisen zugrundeliegt, durch die Verwendung bedientheoretischer Modelle gestärkt und effizienter gestaltet werden.

1.2 Einsatz bedientheoretischer Modelle bei der Optimierung von Materialflußsystemen

Neben der Nutzung bedientheoretischer Modelle zur Analyse und Bewertung von Planungsvarianten können diese Modelle auch zur Struktur- und Parameteroptimierung von Materialflußsystemen herangezogen werden. Während die Strukturoptimierung für die Gestaltung des Materialflußsystemes genutzt wird, kann die Parameteroptimierung sowohl in der Planungsphase (z.B. für die Bestimmung der Zahl der notwendigen Pufferplätze) als auch in der Betriebsphase eines Materialflußsystems (z.B. bei der Bestimmung von Fertigungslosgrößen) genutzt werden.

Kann der Bezug zwischen Leistungskennwerten wie die Durchlaufzeit und systemgestaltenden Parametern, wie die Zahl der für eine Tätigkeit zu installierenden Maschinen mit Hilfe von invertierbaren Funktionen ausgedrückt werden, so kann durch Umkehrung dieser Funktionen eine Optimierung vorgenommen werden.

Die Komplexität der Planungsaufgaben und die Zahl der in die Zielsetzung eingehenden Faktoren läßt jedoch eine solche einfache Optimierung meist nicht zu. Häufiger sind Optimierungen in mehrdimensionalen Räumen mit nichtlinearen Zielfunktionen durchzuführen. Hierfür können Heuristiken und Suchverfahren herangezogen werden, die durch planvolle Vorgehensweise ausgehend von einer Anfangslösung eine Verbesserung des Zielfunktionswertes vornehmen.

Zur Erzielung eines guten Ergebnisses ist im Rahmen solcher Verfahren bei jedem Schritt der gegenwärtige Wert der Zielfunktion und damit die Verbesserung im Vergleich zum vorherigen Schritt sowie in Abhängigkeit vom Verfahren die nächstfolgende Parametervariation zu bestimmen.

In einem solchen Umfeld kann die Geschwindigkeit, mit der Beiträge zur Zielfunktion ermittelt werden können, entscheidend für die Einsetzbarkeit eines Optimierungsverfahrens sein. Für diese Optimierungsverfahren können bedientheoretische Berechnungsverfahren eingesetzt werden, sofern in die Zielfunktion bewertete Kennwerte für Durchlaufzeiten und Bestand eingehen.

1.3 Das Zusammenwirken von bedientheoretische Methoden und computergestützter Simulation

Bedientheoretische Methoden und computergestützte Simulationsmodelle ergänzen sich in Ihrer Anwendung im Rahmen der Materialflußplanung. Grundsätzlich können alle bedientheoretischen Modelle auf geeigneten computergestützten Simulationssystemen abgebildet werden. Mit Hilfe von Simulationsexperimenten sind für diese Modelle dann Kennwerte im Rahmen der hierbei zu erreichenden statistischen Genauigkeit zu bestimmen.

Zur Interpretation dieser Simulationsergebnisse ist ein Erfahrungswissen von Vorteil, das auf dem Systemverständnis aufbauen kann, das mit Hilfe bedientheoretischer Betrachtungen erworben werden kann. Durch die analytische Betrachtung bedientheoretischer Modelle können kausale Beziehungen zwischen Veränderungen des Modells und beobachteten oder berechneten Leistungskennwerten hergestellt werden. Diese aus der Bedientheorie bekannten kausalen Beziehungen können genutzt werden, um zumindest einen Teil der Ergebnisse komplexerer Simulationsexperimente zu erklären und somit in der Folge die Analyse der Simulationsergebnisse erleichtern.

Eine Kombination bedientheoretischer Berechnungsverfahren mit Simulationsexperimenten kann genutzt werden, um z.B. in komplexen Optimierungsmodellen mit Hilfe bedientheoretischer Kennwertberechnungsverfahren und einem vereinfachten Systemmodell das Suchverfahren zu durchlaufen. Das aus dem Suchverfahren resultierende Endergebnis wird anschließend mit Hilfe eines Simulationsexperimentes und einem detaillierteren Modell auf Gültigkeit hin verifiziert.

1.4 Struktur der weiteren Vorgehensweise

Mit Hilfe eines bedientheoretischen Modells können verschiedene Fragestellungen aus dem Bereich der Materialflußplanung bearbeitet werden. Deshalb orientiert sich die Darstellung in den folgenden Kapiteln am Methodischen, um die für Materialflußplanungszwecke verwendbaren bedientheoretischen Bausteine und Berechnungsverfahren einzuführen.

In Kapitel 2 werden die Grundlagen für das Verständnis einzelner Bediensysteme gelegt. Beginnend mit der Einführung des Bediensystems als Grundbaustein der Bedientheorie wird mit einfachen Modellen der operationellen Analyse, die wenige, beobachtbare Eingangsgrößen verwenden, gezeigt, welche Ergebnisse mit diesen Informationen erreichbar sind. Der nächste weiterführende Schritt leitet

hin zur Verwendung stochastischer Prozesse zur Beschreibung der Ein- und Ausgangsgrößen und die Analyse von Bediensystemen mit Hilfe stochastischer Prozesse. Dabei werden sukzessive einschränkende Bedingungen bezüglich der Verteilungsannahmen der zugrunde liegenden stochastischen Prozesse gelockert, bis keine exakten Berechnungsverfahren mehr angegeben werden können. Für Bediensysteme mit generellen Verteilungen werden Näherungsverfahren dargestellt. Eingebettet in Kapitel 2 sind Beispiele aus dem Materialflußbereich, für die sich jeweils mit Hilfe der hergeleiteten Ergebnisse Leistungskennwerte ermitteln lassen.

Das Kapitel 3 beschäftigt sich mit vernetzten Bediensystemen. Die Bediensystemnetzwerke werden in geschlossene und offene Bediensystemnetzwerke unterschieden. Da sich die Berechnungsverfahren für offene Bediensystemnetzwerke unmittelbar auf den Ergebnissen für einzelne Bediensysteme abstützen, werden zunächst offene und danach geschlossene Bediensystemnetzwerke behandelt.

Wie bereits in Kapitel 2 wird gezeigt, für welche Fälle exakte Ergebnisse erreicht werden können. Auch hier wird deutlich, daß diese Bedingungen in der Praxis häufig nicht erfüllt sind, so daß zur Beantwortung vieler praxisrelevanter Fragestellungen die mathematischen Randbedingungen gelockert werden müssen. Aus diesem Grund kommen zur Bestimmung der Kennwerte Näherungsverfahren zum Einsatz. Für zwei Netzwerktypen wird jeweils ein Approximationsverfahren dargestellt.

Den Abschluß des Kapitels bilden Anwendungsbeispiele aus dem Bereich der Materialflußplanung, die am Institut für Fördertechnik und Logistiksysteme (IFL) entwickelt und in die Praxis übertragen wurden.

2 Einzelemente von Materialflußsystemen

Die Analyse vernetzter Materialflußelemente setzt voraus, daß das Verhalten der Einzelemente in Abhängigkeit von den jeweiligen Umgebungsbedingungen bekannt ist. Deshalb werden in diesem Kapitel die Grundlagen der bedientheoretischen Modellierung und Berechnung von Materialflußelementen als Einzelbediensysteme dargestellt.

Materialflußelemente verfügen im allgemeinen über mindestens einen Ein- und einen Ausgang. Durch das Element wird im einfachsten Fall ein Eingangsstrom von Fördereinheiten nach Erbringen einer „Dienstleistung“, meist mit Hilfe einer maschinellen Einrichtung, in einen Abgangsstrom umgewandelt. Konkurrieren mehrere Fördereinheiten um diese „Dienstleistung“, so kann es vorkommen, daß die Fördereinheiten auf den Beginn der Bedienung warten müssen. Diese Wartezeit soll im Sinne eines Materialflusses mit geringen Durchlaufzeiten so kurz wie möglich gehalten werden. Im Interesse einer wirtschaftlichen Planung besteht das Ziel darin, mit so wenig Ressourcen wie nötig eine Förderaufgabe (z.B. Heben auf ein anderes Niveau, Transport eines Behälters über ein FTS, Auslagern eines Behälters in einem automatisierten Hochregallager) innerhalb vorgegebener Betriebsgrenzen (z.B. mittlere Wartezeit) zu lösen. Von welchen Parametern die Wartezeit, und andere davon abgeleitete Größen, abhängen, wird in diesem Kapitel dargestellt werden.

Um von einer konkreten technischen Ausführung des Materialflußelementes unabhängig zu sein, gilt es, die grundlegenden Abläufe im Zusammenhang mit einem Materialflußelement geeignet zu modellieren. Sollen die Auswirkungen von Ressourcenkonkurrenz quantitativ bewertet werden, so bieten sich hierzu Modelle der Bedientheorie (auch häufig angelehnt an den englischen Begriff engl. *queuing theory* als Warteschlangentheorie bezeichnet) an. Mit Hilfe dieser Modelle werden allerdings nur diejenigen Eigenschaften abgebildet, die für die Entstehung von Wartezeiten und -schlangen wichtig sind.

2.1 Grundbestandteile eines bedientheoretischen Modells

Betrachtungsgegenstand der Bedientheorie sind die Bediensysteme, die aus einer oder mehreren Bedienstationen und einem Warteraum bestehen (siehe Abbildung 2.1). Besitzt das Bediensystem genau eine Bedienstation, so spricht man auch von einem Einkanalssystem, sind es mehrere Bedienstationen, so wird es auch als Mehrkanalsystem bezeichnet.

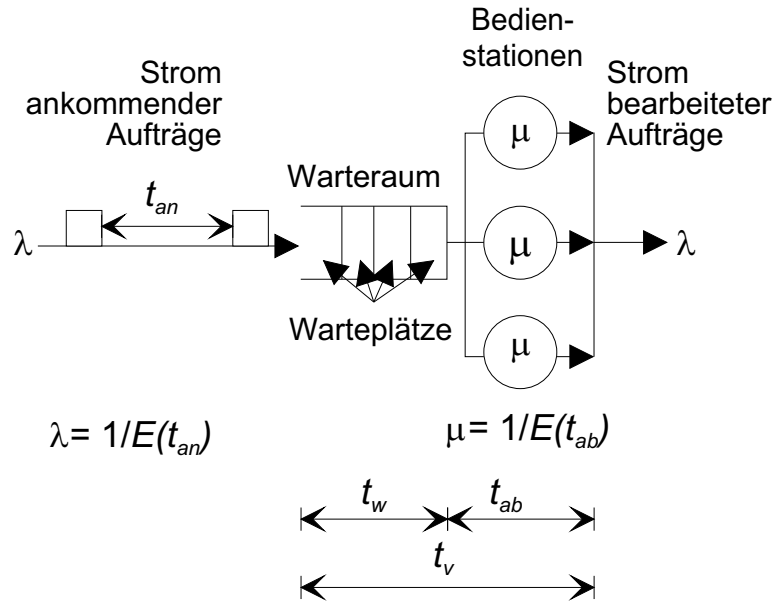


Abbildung 2.1: Bestandteile eines bedientheoretischen Modells

Die Ankünfte der Kunden erfolgen nach einem Zufallsprozeß, der durch die statistische Verteilung der Zeit zwischen zwei aufeinanderfolgenden Ankünften, der sogenannten *Zwischenankunftszeit* t_{an} charakterisiert ist. Mit Hilfe des Erwartungswerts der Zwischenankunftszeit $E(t_{an})$ kann auch die *Ankunftsrate*, also die mittlere Zahl der Ankünfte pro Zeiteinheit als $\lambda = 1/E(t_{an})$ ausgedrückt werden. Ist mindestens eine Bedienstation frei, so wird der eingetroffene Kunde sofort bedient. Die Dauer der Bedienung sei ebenfalls zufällig und durch t_{ab} beschrieben. Beträgt der Erwartungswert der *Bedienzeit* $E(t_{ab})$, so können im Mittel pro Zeiteinheit $\mu = 1/E(t_{ab})$ Kunden bedient werden. Diese Rate ist die sogenannte *Bedienrate* und gibt an, wieviele Kunden bei gegebener statistischer Verteilung der Bedienzeiten und voller Auslastung des Bediensystems an einer Bedienstation maximal pro Zeiteinheit bedient werden können.

Ist keine Bedienstation frei, so bleibt der Kunde zunächst im *Warteraum*, sofern ein solcher vorhanden ist und noch über freie Kapazität verfügt. Aus dem

Warteraum werden die Kunden in der sogenannten *Bediendisziplin* abgerufen. Die Disziplin beschreibt die Auswahlkriterien, nach denen die Kunden aus der Warteschlange ausgewählt werden.

Zur Klassifizierung verschiedener Bediensystemmodelle wurde von Kendall eine Notation eingeführt, die im Kern den Ankunftsprozeß, den Bedienprozeß, die Zahl paralleler Bedienstationen sowie die Bediendisziplin beinhaltet. Mit Hilfe dieser Klassifikation können die Modelle in einem Schema der Form $A|B|m - XXXX$ beschrieben werden. Hierbei beschreibt:

A den Ankunftsprozeß, wobei nachfolgend vor allem M als Symbol für den sogenannten Markov-Prozeß und G für durch generelle Verteilungen charakterisierte Prozesse verwendet wird

B den Bedienprozeß, für den dieselben Prozesse von Interesse sind

m die Anzahl paralleler Bedienstationen, sie kann auch unendlich sein

XXXX die Bediendisziplin, nach der Kunden aus der Warteschlange der nächsten freiwerdenden Bedienstation zugeordnet werden. Es werden verwendet:

FCFS Bei First-Come-First-Served oder synonym First-In-First-Out (FI-FO), werden die Kunden nach der Reihenfolge ihres Eintreffens der freien Bedienstation zugewiesen.

LCFS Last-Come-First-Served (LCFS) oder auch Last-In-First-Out weist den zuletzt eingetroffenen Kunden der freien Bedienstation zu.

RR Nach Round-Robin (RR) werden eintreffende Kunden reihum den Bedienstationen zur Bearbeitung zugewiesen, unabhängig davon, ob die Bedienstation frei ist. Werden die eintreffenden Kunden bei einem Bediensystem mit zwei Bedienstationen durchgezählt, so werden alle eintreffenden Kunden mit ungerader Nummer der einen Bedienstation und alle mit gerader Nummer der anderen Bedienstation zugeordnet und dort jeweils bedient.

PS Processor-Sharing teilt die Bedienstationen allen anwesenden Kunden in gleichen Teilen zu.

Außerdem können für Kundenklassen Prioritäten vergeben werden, wobei in diesem Fall wiederum zwischen unterbrechenden und nichtunterbrechenden Prioritäten unterschieden werden kann. Dabei stehen:

PRE für Preemptive, also unterbrechende und

NONPRE für nichtunterbrechende Prioritäten.

Eine detailliertere Beschreibung ist in Abschnitt 2.10 zu finden.

Für ein Bediensystem sollen Leistungskennwerte wie der Auslastungsgrad ρ , die Wartezeit im Warteraum t_w und die gesamte Verweilzeit im Bediensystem t_v berechnet werden. Weiterhin von Interesse sind die Zahl der im Mittel im Warteraum befindlichen Kunden N_w und die insgesamt im Mittel im Bediensystem befindliche Zahl von Kunden N . Obwohl hier auf die explizite Nennung eines Erwartungswertes (z.B. $E(t_w)$) verzichtet wurde, soll darauf hingewiesen werden, daß auch diese Werte den Charakter eines Erwartungswertes besitzen. Sollen sie als einzelne Realisation betrachtet werden, so wird darauf ausdrücklich hingewiesen werden.

Am Beispiel einer einfachen Hubeinrichtung sollen die Grundlagen der Modellbildung dargestellt werden.

2.2 Deterministisches Modell eines Materialflußelementes

Die in Abbildung 2.2 dargestellte Hubeinrichtung verbindet einen Arbeitsplatz, an dem Teile in einen Transportbehälter (Fördereinheit) gelegt werden, mit einer Förderstrecke, die unter der Hallendecke geführt ist. Der Arbeitsplatz stellt eine *Quelle* von Fördereinheiten dar, die von der Hubeinrichtung, die in diesem Fall die *Bedienstation* darstellt, bedient werden müssen.

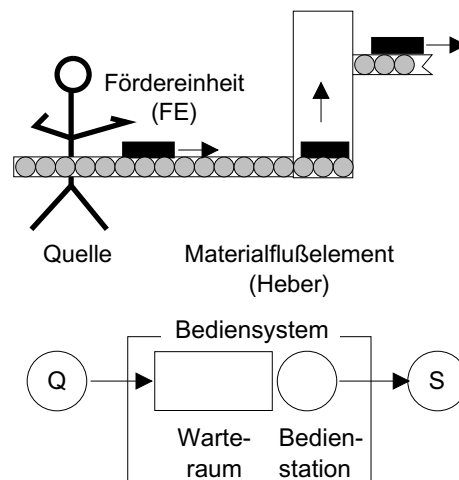


Abbildung 2.2: Einfaches Materialflußelement

Falls sich Fördereinheiten vor der Hubeinrichtung aufstauen, so geschieht dies auf dem Rollenförderer vor der Hubeinrichtung, der durch den *Warteraum* abgebildet wird. Das *Bediensystem* umfaßt den Warteraum und die Bedienstation. Die Fördereinheiten verlassen die Hubeinrichtung auf der Rollenbahn, die als Senke mit unbegrenzter Aufnahmefähigkeit betrachtet werden soll.

Es wird deutlich, daß bei der Modellierung Abstraktionen vorgenommen werden müssen. Auch wenn die Fördereinheit das System Hubeinrichtung bereits verlassen hat, ist die Bedienstation noch aktiv, denn das Lastaufnahmemittel muß zurück in die Ausgangsstellung abgesenkt werden. Aus diesem Grund wird die Anforderung eines Ressourcenbedarfs durch das Eintreffen eines *Kunden* modelliert. Die Bedienzeit eines Kunden soll der gesamten einer Fördereinheit zuzuordnenden Spielzeit der Fördereinrichtung entsprechen (siehe Abbildung 2.3)

Unter der Annahme, daß das Hubwerk für ein Spiel, das heißt Heben mit Fördereinheit 15 Sekunden in Anspruch nimmt und das anschließende Senken ebenfalls 15 Sekunden dauert, beträgt die Spielzeit 30 Sekunden. Dauert das Beladen eines Behälters immer genau 60 Sekunden, so wird die Hubeinrichtung wie folgt als Bediensystem modelliert: Die Zwischenankunftszeit beträgt stets 60 Sekunden, die Bedienzeit (entspricht der Spielzeit) nimmt stets 30 Sekunden in Anspruch.

Der zeitliche Verlauf einiger Systemkenngrößen läßt sich nun leicht ermitteln, und wie in Abbildung 2.3 zu sehen ist, grafisch darstellen.

Betrachten wir $N(t)$, die Zahl der Kunden, die sich zum Zeitpunkt t im Bediensystem aufhalten, so entwickelt sich diese wie folgt: Mit Eintreffen der ersten Fördereinheit zum Zeitpunkt 1 Minute nach Beginn der Arbeit befindet sich die Lastaufnahme in der unteren Position, die Fördereinheit kann also sofort einfahren, $N(t)$ nimmt den Wert „1“ an. Obwohl die Fördereinheit das System bereits vorher verlassen hat, endet die Aktivität der Hubeinrichtung erst zum Zeitpunkt 1,5 Minuten. Nach einer weiteren Minute trifft die nächste Fördereinheit ein, und der zuvor beschriebene Ablauf wiederholt sich identisch.

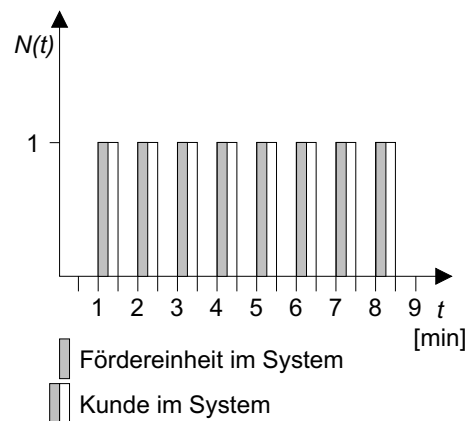


Abbildung 2.3: Belegung der Hubeinrichtung über der Zeit

Eine zentrale Kenngröße bei der Auslegung von Materialflußsystemen ist die *Auslastung* die beschreibt, welcher Teil der zur Verfügung gestellten Kapazität für die Durchführung der Förderaufgabe, also die Bedienung von Kunden genutzt wird. Die Aktivzeit sei die Zeit, die von der Hubeinrichtung pro Arbeitsspiel für die

Durchführung der Förderaufgabe benötigt wird. In diesem Falle besteht sie aus der Zeit für das Einfahren der Fördereinheit, das Anheben der Hubeinrichtung, das Ausfahren der Fördereinheit und die Rückkehr der leeren Hubeinrichtung zum Ausgangspunkt. Die Auslastung ρ , also das Verhältnis von Aktivzeit der Hubeinrichtung zur insgesamt zur Verfügung stehenden Zeit (Gesamtzeit), läßt sich aus den gegebenen Größen pro Arbeitsspiel bestimmen als:

$$\begin{aligned}\rho &= \frac{\text{Aktivzeit}}{\text{Gesamtzeit}} \\ &= \frac{30[s]}{60[s]} \\ &= \frac{1}{2}\end{aligned}\tag{2.1}$$

Da auch bei steigender Auslastung die Bedienung jedes Kunden beendet ist, bevor der nächste eintrifft kommt es zu keinen Wartezeiten vor der Hubeinrichtung.

Diese Eigenschaft bleibt erhalten, solange die Zwischenankunftszeit konstant und größer als die ebenfalls konstante Bedienzeit ist.

Konstante Zwischenankunfts- oder Bedienzeiten sind jedoch in realen Materialflusssystemen sehr selten anzutreffen. Sollen unregelmäßige Einflüsse, wie z.B. menschliche Arbeit, in unregelmäßigen Abständen eintreffende Kunden oder das Ausfallverhalten von Maschinen berücksichtigt werden, müssen umfassendere Ansätze gewählt werden, die auf einer stochastischen Beschreibung des Ressourcenbedarfs beruhen.

Aus diesem Grund beschäftigt sich dieses Kapitel im Folgenden mit der stochastischen Modellierung von Materialflußelementen. Hierbei wird auf Methoden der operationellen Analyse und auf grundlegende bedientheoretische Elemente zurückgegriffen.

2.3 Operationelle Analyse

Die von Denning und Buzen ([15] und [16]) entwickelte operationelle Analyse ermöglicht die Berechnung von Leistungskennwerten aus Größen, die im realen Betrieb oder während eines Simulationslaufes einfach zu beobachten sind. Die Beobachtung erfolge über einen endlichen Zeitraum der Länge T . In diesem Zeitraum werden an dem Bediensystem die folgenden Größen gemessen:

A : Anzahl angekommene Kunden während T (A vom Engl. *arrival*)

- C : Anzahl abgefertigte Kunden während T (C vom Engl. *completion*)
 B : Summe der Aktivzeit während T (B vom Engl. *busy-time*)
 W : Summe der Verweilzeiten aller Aufträge während T (W vom Engl. *waiting-time*)

Aus diesen Größen lassen sich bereits einige Kenngrößen berechnen, die die Leistungsfähigkeit des Bediensystems beschreiben. So erhält man:

- $\lambda = A/T$: Zugangsrate
 $X = C/T$: Abgangsrate
 $S = B/C$: mittlere Bedienzeit eines Kunden
 $U = B/T$: Auslastung der Bedienstation
 $\bar{k} = W/T$: mittlere Anzahl von Kunden im Bediensystem
 $R = W/C$: mittlere Verweilzeit eines Kunden

Erweitert man den Ausdruck für die Auslastung U des Bediensystems mit C/C so erhält man:

$$U = \frac{B}{T} = \frac{B}{T} \cdot \frac{C}{C} = \frac{C}{T} \cdot \frac{B}{C} = X \cdot S \quad (2.2)$$

Soll das System langfristig ein Gleichgewicht erreichen können, so muß gelten $A = C$. Daraus folgt dann für die Auslastung U :

$$U = \lambda \cdot S \quad (2.3)$$

Erweitert man den Ausdruck für die mittlere Anzahl von Kunden \bar{k} mit C/C , so ergibt sich:

$$\bar{k} = \frac{W}{T} = \frac{W}{T} \cdot \frac{C}{C} = \frac{C}{T} \cdot \frac{W}{C} \quad (2.4)$$

Gilt auch hier langfristig $C = A$, so resultiert daraus:

$$\bar{k} = \frac{C}{T} \cdot \frac{W}{C} = \frac{A}{T} \cdot \frac{W}{C} = \lambda \cdot R \quad (2.5)$$

Lediglich mit der Annahme, daß sich das betrachtete System im Gleichgewicht befindet, kann gezeigt werden, daß ein direkter, linearer Zusammenhang zwischen der Durchlaufzeit durch ein System (bspw. einer Fertigung) und dem Bestand in diesem System besteht. Soll also bei konstanter Ankunfts- bzw. Abgangsrate der Bestand reduziert werden, so ist dies nur möglich, wenn gleichzeitig auch die Durchlaufzeit reduziert wird.

Basierend auf dieser Überlegung kann (2.5) zur Überprüfung von Simulationsergebnissen verwendet werden. Um die notwendige Simulationslaufzeit zu reduzieren, wird bisweilen ein Simulationslauf mit bereits gefülltem System gestartet, was die Einschwingphase reduzieren soll. Vergleicht man die Ergebnisse mehrerer auf diese Weise durchgeführter Simulationsläufe, so ist die Einhaltung der durch Ausdruck (2.5) gegebenen Relationen eine notwendige (wenn auch nicht hinreichende) Voraussetzung für die Gültigkeit der auf diese Weise erzielten Ergebnisse.

Der Ausdruck (2.5) ist bekannt als Littles Gesetz in der Form der operationellen Analyse. Es wurde von J.D. Little 1961 bewiesen und von Wiendahl [65] als „Erstes Produktionslogistisches Gesetz“ in die Literatur eingeführt.

Für den Einsatz der operationellen Analyse in der Planung ist zu klären, ob mit ihrer Hilfe auch Prognosen von Leistungsgrößen möglich sind. Hierzu soll der stochastische Prozeß $N(t)$ an einem Bediensystem genauer betrachtet werden. In Abbildung 2.4 ist der Verlauf von $N(t)$ beispielhaft dargestellt. Während des Beobachtungszeitraumes T wurden für $N(t)$ Werte zwischen n_{min} und n_{max} beobachtet.

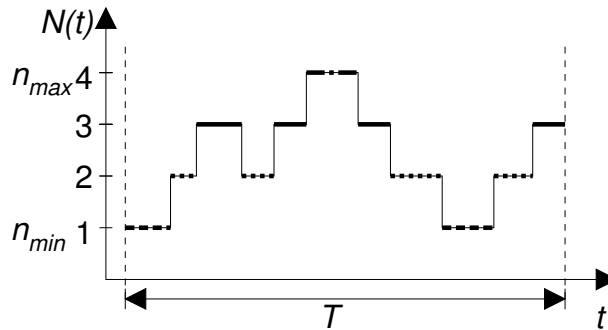


Abbildung 2.4: Verlauf von $N(t)$

Es sei $I(n)$ die Summe der Zeit, die der Prozeß $N(t)$ insgesamt im Zustand $N(t) = n$ verbracht hat, also genau n Kunden im System waren. Deshalb gilt $\sum_{n=n_{min}}^{n_{max}} I(n) = T$. Damit können die Zeitanteile $P(n)$ abgeleitet werden, in denen sich genau n Kunden im System befanden:

$$P(n) = \frac{I(n)}{T} \tag{2.6}$$

Weiterhin gilt:

$$\sum_{n=n_{min}}^{n_{max}} A(n) = A \tag{2.7}$$

$$\sum_{n=n_{min}}^{n_{max}} C(n) = C \tag{2.8}$$

Während der Zeit $I(n)$ wurden insgesamt $A(n)$ Zugänge und $C(n)$ Abgänge beobachtet. Es sind $A(n)$ die Zahl der Kunden, die an dem System eintrafen, als bereits genau n Kunden anwesend waren, und $C(n)$ die Zahl der Kunden, die das System verließen, als sich genau n Kunden im System befanden. Setzt man voraus, daß zu *einem* Zeitpunkt höchstens eine Ankunft oder eine Abfertigung stattfindet, sind die zustandsabhängigen Ankunfts- und Bedienraten $\lambda(n)$ bzw. $\mu(n)$ wie folgt zu ermitteln:

$$\lambda(n) = \frac{A(n)}{I(n)} \quad (2.9)$$

$$\mu(n) = \frac{C(n)}{I(n)} \quad (2.10)$$

Ist das System während des Beobachtungszeitraumes im Gleichgewicht, so sind am Ende des Beobachtungszeitraumes genausoviele Kunden im System, wie zu Beginn der Beobachtung. Es gilt $A = C$. $A(n)$ und $C(n)$ sind im allgemeinen nicht gleich groß, für das Beispiel aus Abbildung 2.4 ist $C(4) = 1$ und $A(4) = 0$. Die Zahl der Übergänge zwischen benachbarten Zuständen $A(n-1)$ und $C(n)$ muß stets identisch sein, um ein Gleichgewicht erreichen zu können:

$$A(n-1) = C(n) \quad \forall n = n_{min} + 1, \dots, n_{max} - 1 \quad (2.11)$$

Außerdem gilt im Bereich von $n = n_{min} + 1, \dots, n_{max} - 1$:

$$P(n-1) \cdot T \cdot \lambda(n-1) = A(n-1) \quad (2.12)$$

und:

$$\begin{aligned} P(n) \cdot T \cdot \mu(n) &= C(n) \\ \Rightarrow P(n) \cdot T \cdot \mu(n) &= P(n-1) \cdot T \cdot \lambda(n-1) \\ \Leftrightarrow P(n) &= P(n-1) \frac{\lambda(n-1)}{\mu(n)} \end{aligned} \quad (2.13)$$

Dies gilt für alle Zustände $n = n_{min} + 1 \dots n_{max} - 1$ und damit ist der Zeitanteil, mit dem sich das System im Zustand $N(t) = n$ befindet:

$$P(n) = P(n_{min}) \cdot \prod_{k=n_{min}+1}^n \frac{\lambda(k-1)}{\mu(k)} \quad \forall n = n_{min} + 1, \dots, n_{max} - 1 \quad (2.14)$$

Unter der Annahme, daß die Ankunfts- bzw. Bedienraten zustandsunabhängig (homogen) sind, also $\lambda(n) = \lambda$ und $\mu(n) = \mu$ gilt, vereinfacht sich vorstehendes Ergebnis für eine Beobachtungszeit $T \rightarrow \infty$ zu

$$P(n) = P(0) \left(\frac{\lambda}{\mu} \right)^n \quad (2.15)$$

da n_{min} für $T \rightarrow \infty$ bei Auslastungen von $\rho < 1$ zu 0 wird.

Dieses Ergebnis könnte auf den ersten Blick hin einen großen Gültigkeitsbereich besitzen, denn die Einschränkung, daß homogene Ankunfts- und Bedienraten vorliegen müssen, scheint plausibel zu sein, sofern nicht Steuerungsabläufe modelliert werden.

Betrachten wir jedoch das einleitende Beispiel (siehe Abbildung 2.3), so ist festzustellen, daß die Voraussetzung zustandsunabhängiger Ankunfts- und Bedienraten den Anwendungsbereich des Ergebnisses stark einschränkt.

Wir haben für das Beispiel bereits $\rho = 1/2$ berechnet. Die Wahrscheinlichkeit $P(0)$ dafür, daß das System leer ist, läßt sich als Komplement der Wahrscheinlichkeit, daß das System aktiv ist, ausdrücken. Wir erhalten:

$$P(0) = 1 - \rho = 1/2 \quad (2.16)$$

Damit ist die Wahrscheinlichkeit $P(1)$, daß sich genau ein Kunde im System befindet, als

$$P(1) = (1 - \rho)\rho = 1/2 \cdot 1/2 = 1/4 \quad (2.17)$$

auszudrücken. Weiterhin ist $P(2)$:

$$P(2) = (1 - \rho)\rho^2 = 1/2 \cdot (1/2)^2 = 1/8 > 0 \quad (2.18)$$

Durch Anwendung dieses Modells werden Wahrscheinlichkeiten, die größer als Null sind, für Zustände mit 2 und mehr Kunden im System berechnet. Solche Zustände können unter den getroffenen Zeitannahmen jedoch in Realität nicht auftreten. Unsere Modellannahmen (in diesem Fall die Zustandsunabhängigkeit der Ankunfts- und Bedienraten) waren offensichtlich für Aussagen, die über eine Prognose der Auslastung hinausgehen, ungeeignet. Für die Prognose von Leistungskennwerten werden offensichtlich genauere Beschreibungen der Ankunfts- und Bedienprozesse benötigt, als in dem vorstehenden Modell verwendet wurden. Deshalb wenden wir uns speziellen stochastischen Modellen zu, die mit den Methoden der Bedientheorie bearbeitet werden können.

2.4 Exkurs: Stochastische Prozesse

Die Ergebnisse, die in Abschnitt 2.3 mit Hilfe der operationellen Analyse erzielt werden konnten, waren nicht für die Prognose von Leistungskennwerten geeignet. Ein Grund hierfür liegt in der Tatsache, daß die Ankunfts- und Bedienprozesse jeweils nur mit einem Parameter, der Ankunftsrate λ beziehungsweise der Bedienrate μ beschrieben wurden.

Die Beschreibung der Ankunfts- und Bedienprozesse mit Mitteln der Statistik erlaubt es, weitere Informationen in den Modellierungs- und Berechnungsvorgang einfließen zu lassen. Obwohl die zugrundeliegenden Vorgänge häufig nicht wirklich zufällig sind, kann die Unterstellung des Zufalls eine ökonomisch vernünftige Entscheidung sein. So ist es theoretisch möglich, aus den physikalischen Parametern eines Würfelwurfs (z.B. Rotationsgeschwindigkeit des Würfels, Geschwindigkeit des Würfels, Auftreffwinkel, Elastizität des Untergrundes etc.) den Ausgang des Wurfs zu bestimmen. Es genügen jedoch bereits kleine Änderungen in diesen Parametern, um den Ausgang des Experiments zu verändern. Für viele Zwecke, insbesondere wenn nicht das einzelne Experiment sondern eine Folge von Experimenten von Interesse ist, kann es deshalb ökonomischer sein, den gesamten Vorgang als Zufallsexperiment zu betrachten (siehe auch Bamberg und Baur [59], S. 89ff.). Solche Überlegungen können auch für Materialflusssysteme angestellt werden. Darüber hinaus führt die Berücksichtigung von Störungen ohnehin dazu, daß das Gesamtsystem ein stochastisches System wird.

Ausgangspunkt der nachfolgenden Betrachtung ist ein Zufallsexperiment. Ein einzelnes Zufallsexperiment hat als Ergebnis ein Elementarergebnis ω . In der Menge der Elementarereignisse $\Omega = \{\omega | \omega \text{ ist Elementarereignis}\}$ können die möglichen Ergebnisse des Zufallsexperimentes zusammengefaßt werden. Ist das Zufallsexperiment beispielsweise das Würfeln mit einem Würfel, so kann die Augenzahl als Elementarereignis aufgefaßt werden. Die Menge der Elementarereignisse ist in diesem Falle $\Omega = \{1, 2, 3, 4, 5, 6\}$.

Üblicherweise interessiert, ob bei einem Zufallsexperiment ein Elementarereignis aus einer Menge A von Elementarereignissen eingetreten ist. Elementarereignisse ω können zu verschiedenen Mengen A von Elementarereignissen zusammengefaßt werden. Das Eintreffen eines Ereignisses aus einer dieser Mengen soll als Ereignis bezeichnet werden. Am Beispiel des Würfelwurfs könnte ein Ereignis als „Augenzahl ist kleiner als 4“ beschrieben werden. Die Menge aller aus den Elementarereignissen kombinierbaren Ereignisse ist deshalb die Potenzmenge \mathcal{A} , die alle möglichen Teilmengen von Ω enthält.

Mit der Wahrscheinlichkeit soll beschrieben werden, welche Chance besteht, daß ein bestimmtes Ereignis eintritt. Der Begriff Wahrscheinlichkeit ist schwierig zu fassen. Aus diesem Grund werden Wahrscheinlichkeitsmaße meist axiomatisch definiert, wobei das Ergebnis dieser Definitionen mit der „intuitiven“ Definition übereinstimmt.

Die Wahrscheinlichkeit P weist eine nichtnegative Zahl einer Menge von Elementarereignissen A zu und gilt als Wahrscheinlichkeitsmaß, wenn für alle möglichen A und eine endliche Menge Ω nachstehende drei Bedingungen erfüllt sind:

- $P(A) \geq 0$
- $P(\Omega) = 1$

- Wenn die Schnittmenge $A \cap B$ zweier Mengen A und B von Elementarereignissen leer ist, dann gilt für die Wahrscheinlichkeit des Eintritts eines Elementarereignisses aus der Vereinigungsmenge: $P(A \cup B) = P(A) + P(B)$

Ein Tripel (Ω, \mathcal{A}, P) wird Wahrscheinlichkeitsraum genannt. Zur Modellierung eines Problems mögen verschiedene Tripel geeignet sein, die die Wirklichkeit mehr oder minder gut abbilden. Es ist jedoch darauf zu achten, daß wenn das Tripel (Ω, \mathcal{A}, P) einmal festgelegt wurde, dieses für die weiteren Betrachtungen unverändert bleiben muß.

Ausgangspunkt sogenannter *stochastischer* Modelle sind die *Zufallsvariablen* oder *stochastischen* Variablen. Der Name Zufallsvariable ist etwas irreführend, da eine Zufallsvariable eine Funktion ist, die das Ergebnis eines Zufallsexperimentes festhält.

Im Gegensatz zum Wahrscheinlichkeitsmaß, das sich auf eine Menge bezieht, ordnet die Funktion der Zufallsvariablen jedem Element ω aus Ω eine reelle Zahl zu. Ein einfaches Beispiel ist die Indikatorfunktion I_A mit der das Eintreten eines Elementarereignisses aus einer Menge A einfacher darzustellen ist.

$$I_A(\omega) = \begin{cases} 1 & \text{wenn } \omega \in A \\ 0 & \text{wenn } \omega \notin A \end{cases}$$

Nun kann die Wahrscheinlichkeit, daß das Ereignis A eintritt als $P(I_A = 1) = P(A)$ und $P(I_A = 0) = 1 - P(A)$ geschrieben werden.

Im allgemeinen wird an die Zufallsvariable der Anspruch gestellt, daß sie jedem Elementarereignis aus dem Wahrscheinlichkeitsraum (Ω, \mathcal{A}, P) eine reelle Zahl zuordnet. Als Funktion ausgedrückt ist die Zufallsvariable X für reellwertige x : $\{\omega | X(\omega) \leq x\}$ und ist in \mathcal{A} enthalten.

Die zugehörige Verteilungsfunktion beschreibt die den Ereignissen zugeordnete Wahrscheinlichkeit und ist aufbauend auf $X(\omega)$ als $F(x) = P(\{\omega | X(\omega) \leq x\})$ abzuleiten. Durch das Weglassen des expliziten Bezugs zum Elementarereignis ω wird die Verteilungsfunktion üblicherweise kurz $F(x) = P(X \leq x)$ geschrieben.

Wir verwenden Zufallsvariablen nachfolgend vorwiegend zur Beschreibung von Zeiten. So können die Zwischenankunftszeit und die Bedienzeit des nächsten Kunden mit Hilfe von Zufallsvariablen beschrieben werden.

Zur Abbildung von Materialflußsystemen ist ein einziger Ausgang eines Zufallsexperimentes nicht ausreichend. Notwendig ist die Betrachtung einer Menge zeitlich aufeinanderfolgender Ereignisse, die aus dem gleichen oder zusammenhängenden Zufallsexperiment resultieren.

Verändert sich das Zufallsexperiment über der Zeit nicht, so wird es als *stationär* bezeichnet und kann deshalb durch einen *stationären stochastischen Prozeß* beschrieben werden. Sei X eine Zufallsvariable, die über der Zeit verschiedene

Realisationen $x(t_1), x(t_2), \dots$ hervorbringt, so ist $X(t)$ der zu X gehörende stochastische Prozeß.

Zur Modellierung von Materialflußsystemen können sowohl *zeitkontinuierliche* als auch *zeitdiskrete* stochastische Prozesse benutzt werden. Bei zeitkontinuierlichen Prozessen kann die Zeit t jeden Wert aus $T = [0, \infty)$ annehmen, im zeitdiskreten Fall ist T eine Menge natürlicher Zahlen $t = \{0, 1, \dots\}$.

Obwohl im Weiteren nicht verwendet, soll kurz erwähnt werden, daß auch instationäre stochastische Prozesse möglich sind, bei denen $X(t)$ zeitabhängig ist.

Sagt man, daß sich der stochastische Prozeß zur Zeit t im Zustand x befindet, so ist damit ausgedrückt, daß das Ereignis $X(t) = x$ eingetreten ist. Zuständen von Prozessen wird im Folgenden eine große Bedeutung zukommen.

2.5 Das Bediensystem $M|M|1$ -FCFS

Die Analyse von Bediensystemen des Typs $M|M|1$ -FCFS macht in einem starken Maß von den Eigenschaften der Ankunfts- und Bedienprozesse Gebrauch. Die beiden Prozessen zugehörigen Verteilungen sind Exponentialverteilungen, deren Dichteverteilung durch

$$f(t) = \lambda e^{-\lambda t} \quad (2.19)$$

mit entsprechender Verteilungsfunktion

$$F(x) = \int_0^x \lambda e^{-\lambda t} dt = 1 - e^{-\lambda x} \quad \forall x \geq 0 \quad (2.20)$$

gegeben ist. Die besondere Form der Exponentialverteilung ermöglicht einfache Analysen, da die Wahrscheinlichkeit eines zukünftigen Ereignisses (z.B. einer Ankunft) unabhängig von der Zeit ist, die seit dem letzten Ereignis verstrichen ist. Diese Eigenschaft wird häufig als *Gedächtnislosigkeit* bezeichnet.

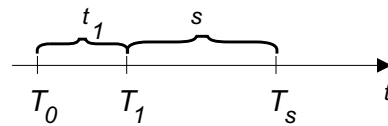


Abbildung 2.5: Position von T_0 , T_1 und T_s

Die Gedächtnislosigkeit kann einfach gezeigt werden. In Abbildung 2.5 markiere T_0 den Zeitpunkt des letzten Ereignisses. Inzwischen sei der Zeitpunkt T_1 erreicht, also die Zeitspanne t_1 verstrichen. Die Gedächtnislosigkeit ist identisch mit folgender Behauptung: Die Wahrscheinlichkeit, daß kein Ereignis bis zum Zeitpunkt T_s stattfindet unter der Bedingung, daß bereits die Zeitspanne t_1 verstrichen ist, ist genau so groß wie die Wahrscheinlichkeit, daß kein Ereignis in der Zeitspanne $t_1 + s$ stattfindet.

$$P(t > s + t_1 | t > t_1) = P(t > s) \quad (2.21)$$

Durch Umformulierung der bedingten Wahrscheinlichkeit erhält man

$$\begin{aligned} P(t > s + t_1 | t > t_1) &= \frac{P(t > s + t_1 \wedge t > t_1)}{P(t > t_1)} \\ &= \frac{P(t > s + t_1)}{P(t > t_1)} \\ &= \frac{1 - F(s + t_1)}{1 - F(t_1)} \end{aligned} \quad (2.22)$$

Setzt man Ausdruck 2.20 ein, so ergibt sich

$$\begin{aligned} P(t > s + t_1 | t > t_1) &= \frac{1 - (1 - e^{-\lambda(s+t_1)})}{1 - (1 - e^{-\lambda t_1})} \\ &= e^{-\lambda s} = P(t > s) \quad \text{q.e.d.} \end{aligned} \quad (2.23)$$

Mit der stetigen Exponentialverteilung korrespondiert die Poissonverteilung, die der Anzahl von Ereignissen in einem Zeitintervall entspricht, wenn die Zeit zwischen den Ereignissen exponentialverteilt ist.

Es sei n die Zahl der Ankünfte während eines Zeitintervalles δ_t . Die Wahrscheinlichkeit dafür, daß genau n Kunden in einem Intervall der Länge δ_t eintreffen ist beschrieben durch

$$P_n(\delta_t) = \frac{(\lambda \delta_t)^n}{n!} e^{-\lambda \delta_t} \quad (2.24)$$

Sind Zwischenankunfts- und Bedienzeiten beide exponentialverteilt, so kann der Zustand eines solchen Bediensystems vollständig mit Hilfe der Zahl der Kunden ausgedrückt werden, die sich im Bediensystem befinden, denn der stochastische Prozeß der Zahl $N(t)$ besitzt die sogenannte Markov-Eigenschaft, die wie folgt definiert ist:

Definition 2.1 (Markov-Eigenschaft) *Die Wahrscheinlichkeit eines zukünftigen Zustandes hängt nur vom gegenwärtigen Zustand ab, die Vergangenheit bildet sich vollständig im gegenwärtigen Zustand ab.*

Das $M|M|1$ -Bediensystem besitzt diese Eigenschaft, es wurde bereits gezeigt, daß für das Auftreten eines Ereignisses die Zeit seit dem Eintritt des letzten Ereignisses ohne Belang ist. Dies gilt offensichtlich für den Ankunftsprozeß, aber auch für die Bedienung eines Kunden.

Aufgrund der Markov-Eigenschaft kann der Zustand eines $M|M|1$ -Bediensystemes vollständig alleine durch die Zahl der Kunden k beschrieben werden, die sich in dem Bediensystem aufhalten.

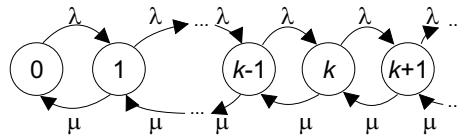


Abbildung 2.6: Markovkette für die Zustände eines $M|M|1$ -Bediensystems

Zur Bestimmung der Zustandswahrscheinlichkeiten wird die sogenannte *Markovkette* der möglichen Zustände und die Übergangsraten zwischen den Zuständen verwendet (siehe Abbildung 2.6).

Hiermit ist eine einfache Bestimmung der Zustandswahrscheinlichkeiten $P(N(t) = k) = p(k)$ möglich. Befindet sich das Bediensystem im Gleichgewicht ($\rho = \lambda/\mu < 1$), dann muß die Wahrscheinlichkeit in einen Zustand k zu gelangen, genau so groß sein, wie die Wahrscheinlichkeit diesen Zustand wieder zu verlassen. Dies bezeichnet man auch als *statistisches Gleichgewicht*. Für den Zustand „0“ gilt demnach:

$$\lambda p(0) = \mu p(1) \tag{2.25}$$

Für alle anderen Zustände mit zwei benachbarten Zuständen ist analog

$$(\lambda + \mu)p(k) = \mu p(k + 1) + \lambda p(k - 1) \tag{2.26}$$

Mit $\rho = \lambda/\mu$ wird aus (2.25):

$$p(1) = \rho p(0) \tag{2.27}$$

Für den Zustand $k = 1$ ergibt (2.26) ausgeschrieben

$$\begin{aligned} (\lambda + \mu)p(1) &= \mu p(2) + \lambda p(0) \\ \Leftrightarrow \lambda \rho p(0) + \mu \rho p(0) &= \mu p(2) + \lambda p(0) \\ \Leftrightarrow \lambda \rho p(0) + \lambda p(0) &= \mu p(2) + \lambda p(0) \\ \Leftrightarrow \rho^2 p(0) &= p(2) \end{aligned} \tag{2.28}$$

Durch rekursive Nutzung vorstehender Vorgehensweise bekommen wir:

$$p(k) = \rho p(k - 1) = \rho^k p(0) \tag{2.29}$$

Die Summe aller Zustandswahrscheinlichkeiten muß 1 ergeben. Dies kann als Normierungsbedingung genutzt werden, um die noch unbekannte Zustandswahrscheinlichkeit $p(0)$ zu bestimmen.

$$\begin{aligned} \sum_{k=0}^{\infty} p(k) &= 1 \\ \Leftrightarrow p(0) \sum_{k=0}^{\infty} \rho^k &= 1 \end{aligned} \tag{2.30}$$

Ersetzt man $\sum_{k=0}^{\infty} \rho^k$ durch den äquivalenten Ausdruck $1/(1 - \rho)$ dann ist

$$p(0) = 1 - \rho \tag{2.31}$$

und damit

$$p(k) = (1 - \rho)\rho^k = p(0)\rho^k \tag{2.32}$$

Dieses Ergebnis ist identisch mit demjenigen, daß in der operationellen Analyse unter der Annahme zustandsunabhängiger Ankunftsrate erzielt wurde. Diese beiden Annahmen sind also äquivalent, was sich bereits in der Erörterung der Gedächtnislosigkeit abzeichnete.

Bildet man die mit den Zuständen k gewichtete Reihensumme, um den Erwartungswert der Zahl von Kunden im System zu bestimmen, so resultiert:

$$N = \sum_{k=0}^{\infty} kp(k) = \sum_{k=0}^{\infty} k(1 - \rho)\rho^k = \frac{\rho}{1 - \rho} \tag{2.33}$$

2.6 Das Bediensystem $M|M|m$ -FCFS

Besitzt das Bediensystem mehr als eine Bedienstation, so kann prinzipiell die gleiche Vorgehensweise wie bei einem $M|M|1$ -Bediensystem angewendet werden, wenn auch die Durchführung aufwendiger ist.

Für den Auslastungsgrad ρ werden in der Literatur zwei verschiedene Formeln angegeben. Nachfolgend sei für m parallele Bedienstationen der Auslastungsgrad ρ wie folgt festgelegt:

$$\rho = \frac{\lambda}{m\mu} \tag{2.34}$$

Damit gibt ρ die durchschnittliche Auslastung einer Bedienstation und $m\rho$ die Zahl der im Mittel aktiven Bedienstationen an.

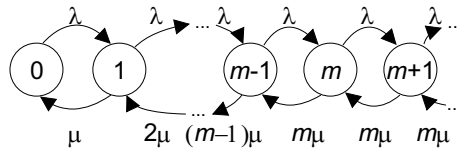


Abbildung 2.7: Markovkette für die Zustände eines $M|M|m$ -Bediensystems

Beim $M|M|m$ -Bediensystem kann die Übergangsrate von einem Zustand zu seinen Nachbarzuständen mit der Zahl der im System befindlichen Kunden variieren. In Abbildung 2.7 ist die Übergangsrate zu Zuständen mit größerer Kundenzahl konstant λ , während die Übergangsraten zu Zuständen mit geringerer

Kundenzahl, die das Abfertigen der Kunden widerspiegeln, vom aktuellen Zustand abhängen kann. Da für Zustände $k < m$ nicht alle Bedienstationen benutzt werden können, ist die Übergangsrate in diesem Bereich durch $k\mu$ gegeben, da bei $k < m$ Kunden im Bediensystem genau k Bedienstationen aktiv sind. Für Zustände $k \geq m$ ist die Übergangsrate dann konstant gleich $m\mu$.

Als Lösung der Gleichgewichtsgleichungen ergibt sich der Ausdruck:

$$p(k) = \begin{cases} \frac{(m\rho)^k}{k!} p(0) & k = 0, 1, \dots, m \\ \frac{m^m \rho^k}{m!} p(0) & k > m \end{cases} \quad (2.35)$$

Aus der Normierungsbedingung $\sum_{k=0}^{\infty} p(k) = 1$ kann anschließend der Wert von $p(0)$ bestimmt werden.

$$p(0) = \frac{1}{\frac{(m\rho)^m}{(1-\rho)m!} + \sum_{i=0}^{m-1} \frac{(m\rho)^i}{i!}} \quad (2.36)$$

Durch Addition der gewichteten Zustandswahrscheinlichkeiten kann nun die mittlere Zahl von Kunden im Bediensystem N bestimmt werden. Nach z.B. [39], S. 105 resultiert:

$$N = m\rho + p(m) \frac{\rho}{(1-\rho)^2} \quad (2.37)$$

Die insgesamt im Mittel in der Warteschlange befindliche Zahl von Kunden N_w beträgt damit

$$N_w = N - m\rho = p(m) \frac{\rho}{(1-\rho)^2} \quad (2.38)$$

2.7 Bediensysteme mit mehrphasigen Verteilungen

Während exponentialverteilte Zwischenankunftszeiten in der Praxis durchaus angetroffen werden können, ist die Annahme exponentialverteilter Bedienzeiten in der Praxis selten zu verifizieren. Dies macht es notwendig, Verteilungen zu finden, die die wesentlichen Eigenschaften der in der Praxis vorgefundenen Verteilungen abbilden können und gleichzeitig die mathematischen Eigenschaften der Exponentialverteilung weitgehend beibehalten. Diese Forderungen können teilweise mit Hilfe sogenannter *mehrphasiger* Verteilungen erfüllt werden. Die Abbildung von Bedienprozessen durch solche Verteilungen wird in den nachfolgenden Abschnitten dargestellt.

2.7.1 Mehrphasige Verteilungen

Aufbauend auf Exponentialverteilungen können andere Verteilungen dargestellt werden, indem der Bedienprozeß in mehrere Phasen zerlegt wird. Diese Phasen werden von einem Kunden vollständig oder, bestimmt durch eine Wahrscheinlichkeitsverteilung, teilweise durchlaufen. Es befindet sich höchstens ein Kunde in Bedienung, also in einer der vorhandenen Phasen.

Die Zeit, die ein Kunde in einer Phase verbringt, ist stets unabhängig exponentialverteilt, so daß bei einer Verwendung in einem Bediensystem nach wie vor Markovketten betrachtet werden können.

Die Erlangverteilung

Im einfachsten Fall wird die gesamte Bedienzeit t_{ab} als Summe von n identisch exponential verteilten Phasen dargestellt. Der Erwartungswert für die Bedienzeit in einer der Phasen muß

$$E(t_{ph}) = \frac{E(t_{ab})}{n} = \frac{1}{\mu n} \quad (2.39)$$

betragen, damit die vorgegebene mittlere Gesamtbedienzeit als

$$E(t_{ab}) = n \cdot E(t_{ph}) = \frac{1}{\mu} \quad (2.40)$$

erreicht wird.

Eine derart definierte Verteilung ist als n -Erlangverteilung bekannt. Ihr Modell ist in Abbildung 2.8 dargestellt.

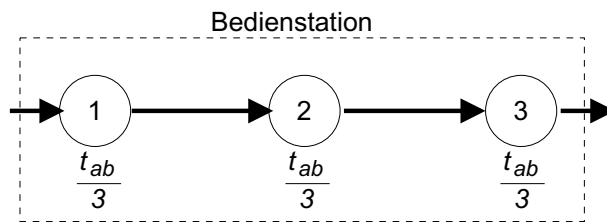


Abbildung 2.8: Erlangverteilung mit $n = 3$ Phasen

Da die Bedienzeiten in den einzelnen Phasen unabhängig voneinander sind, kann die Varianz der Summe der mittleren Bedienzeiten als Summe der Varianzen der einzelnen Phasen bestimmt werden. Die Varianz der exponentialverteilten Zeit in einer Phase beträgt $1/(n\mu)^2$. Damit ist die Varianz der Erlangverteilung mit nachstehender Formel beschrieben:

$$Var(t_{ab}) = n \cdot \frac{1}{(n\mu)^2} = \frac{1}{n\mu^2} \quad (2.41)$$

Zusammen mit Formel (2.40) ergibt sich als Maß für das Verhältnis von Varianz zum quadrierten Erwartungswert die Variabilität c_{ab}^2 des Bedienprozesses

$$c_{ab}^2 = \frac{\text{Var}(t_{ab})}{E(t_{ab})^2} = \frac{1}{n} \quad (2.42)$$

Wird statt der Varianz die Standardabweichung als Maß für die Streuung der Bedienzeiten verwendet, so kann ein Variationskoeffizient c_{ab} von

$$c_{ab} = \frac{\sqrt{\text{Var}(t_{ab})}}{E(t_{ab})} = \frac{1}{\sqrt{n}} \quad (2.43)$$

bestimmt werden.

Die Ermittlung der Dichtefunktion ist über die Transformation der Bedienzeiten in den einzelnen Phasen möglich. Die Bedienzeit ist als Summe von n identisch exponentialverteilten Zufallsvariablen darzustellen. Deshalb kann die resultierende Dichtefunktion durch Faltung der Exponentialverteilungen der einzelnen Phasen erzeugt werden. Zur Herleitung der resultierenden Dichtefunktion ist eine Laplace-Transformation der Dichtefunktion der Zeit in einer der Bedienphasen vorzunehmen und wir erhalten:

$$F_{ph}^*(z) = \frac{n\mu}{z + n\mu} \quad (2.44)$$

Durch n -fache Multiplikation der Transformatierten erhält man die Laplace-Transformierte der gesuchten Dichtefunktion,

$$F^*(z) = F_{ph}^*(z)^n = \left(\frac{n\mu}{z + n\mu} \right)^n \quad (2.45)$$

und nach Rücktransformation resultiert die Dichtefunktion der Erlangverteilung:

$$f(t) = \frac{n\mu(n\mu t)^{n-1}}{(n-1)!} e^{-n\mu t} \quad (2.46)$$

Eine ausführliche Darstellung dieser Vorgehensweise findet sich auf den Seiten 119 ff. bei Kleinrock ([40]).

Die Dichtefunktion der Erlangverteilung ist in Abbildung 2.9 für einige Werte von n dargestellt. Aus dieser Abbildung und aus Gleichung (2.46) ist zu erkennen, daß mit $n = 1$ die Exponentialverteilung als Spezialfall der Erlangverteilung dargestellt werden kann. Mit großen Werten von n können taktförmige Bedienzeiten beliebig gut approximiert werden. Aufgrund der Ganzzahligkeit von n ist jedoch die Approximation von Verteilungen, deren Variationskoeffizient zwischen $1/\sqrt{2}$ und 1 liegt, nicht gut möglich. Weiterhin ist eine Darstellung von Verteilungen mit Variationskoeffizienten > 1 überhaupt nicht möglich. Dies führt zu der im übernächsten Abschnitt beschriebenen Hyperexponentialverteilung.

Sinngemäß gelten die obigen Ausführungen auch für Zwischenankunftszeitverteilungen, wobei hier t_{an} und λ an die Stelle von t_{ab} und μ treten.

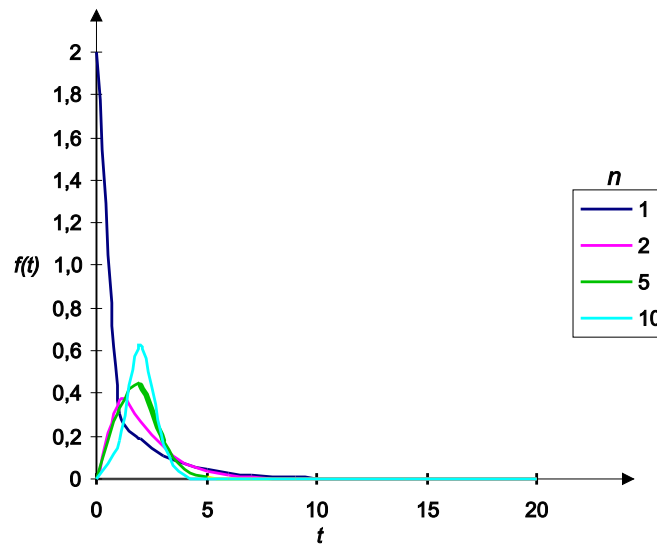


Abbildung 2.9: Dichtefunktionen von Erlangverteilungen mit $n = 1, 2, 5$ und 10 Phasen

Die Hypoexponentialverteilung

Die Hypoexponentialverteilung ist eine Verallgemeinerung der Erlangverteilung, bei der die Bedienzeiten in den Phasen nicht mehr die gleichen Parameter besitzen müssen.

Die Dichtefunktion ist wie bei der Erlangverteilung als Summe von unabhängigen Zufallsvariablen darzustellen. Damit ist es möglich, über den Weg einer Laplacetransformation der Dichtefunktionen der Phasen die Dichtefunktion der Hypoexponentialverteilung zu bestimmen.

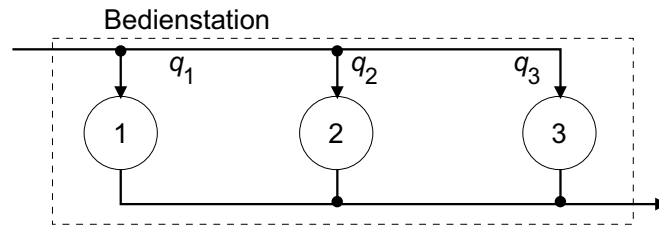
Der Wertebereich der Variabilität einer Hypoexponentialverteilung liegt zwischen 0 und 1 , kann jedoch im Gegensatz zur Erlangverteilung beliebig in diesem Intervall variiert werden.

Die Hyperexponentialverteilung

Mit Hilfe von Hyperexponentialverteilungen können Verteilungen mit Variabilitäten $c^2 > 1$ auf der Basis von Exponentialverteilungen approximiert werden.

Wird als Bedienzeit unter n verschiedenen Exponentialverteilungen jeweils eine mit gegebener Wahrscheinlichkeit q_j ausgewählt, so entsteht eine Hyperexponentialverteilung (siehe Abbildung 2.10).

Der Erwartungswert $E(t_{ab})$ der Funktion ist durch Gewichtung der jeweiligen Mittelwerte $t_{ph,i} = 1/\mu_i$ mit der zugehörigen Auswahlwahrscheinlichkeit q_i der Phase zu ermitteln.


 Abbildung 2.10: Modell einer Hyperexponentialverteilung mit $n = 3$ Phasen

$$E(t_{ab}) = \sum_{i=1}^n q_i t_{ph,i} \quad (2.47)$$

Die Dichtefunktion entsteht aus der Überlagerung der Dichtefunktionen der Phasen, ebenfalls mit der zugehörigen Wahrscheinlichkeit gewichtet.

$$f(t_{ab}) = \sum_{i=1}^n q_i \mu_i \cdot e^{-\mu_i t_{ab}} \quad (2.48)$$

Mit Hilfe des Mittelwerts und der Dichtefunktion wird die Varianz zu

$$Var(t_{ab}) = 2 \sum_{i=1}^n \frac{q_i}{\mu_i^2} - \frac{1}{\mu^2} \quad (2.49)$$

bestimmt. Der aus Mittelwert und Varianz resultierende Variationskoeffizient

$$c_{ab} = \sqrt{2\mu^2 \sum_{i=1}^n \frac{q_i}{\mu_i^2} - 1} \quad (2.50)$$

ist stets größer als 1.

Vergleicht man das Bild einer Hyperexponentialverteilung mit dem einer Exponentialverteilung (siehe Abbildung 2.11), so stellt man fest, daß bei der Hyperexponentialverteilung kleine Werte wahrscheinlicher sind als bei der Exponentialverteilung und daß gleichzeitig die Wahrscheinlichkeit für große Werte kleiner als bei einer Exponentialverteilung ist.

Die Standardabweichung der Hyperexponentialverteilung ist stets größer als der Mittelwert, woraus der bereits hergeleitete Variationskoeffizient größer als 1 resultiert. Aus diesem Grunde wird die Hyperexponentialverteilung häufig zur Approximation von Verteilungen mit Variationskoeffizienten, die größer als 1 sind, eingesetzt (z.B. durch Kühn [43]).

Dabei muß jedoch beachtet werden, daß die Hyperexponentialverteilung von der Gestalt her der Exponentialverteilung ähnlich ist (siehe auch Abbildung 2.11). Weicht die zu modellierende Verteilung sehr von diesem Verlauf ab, so sind die erzielten Ergebnisse mit Vorsicht zu bewerten.

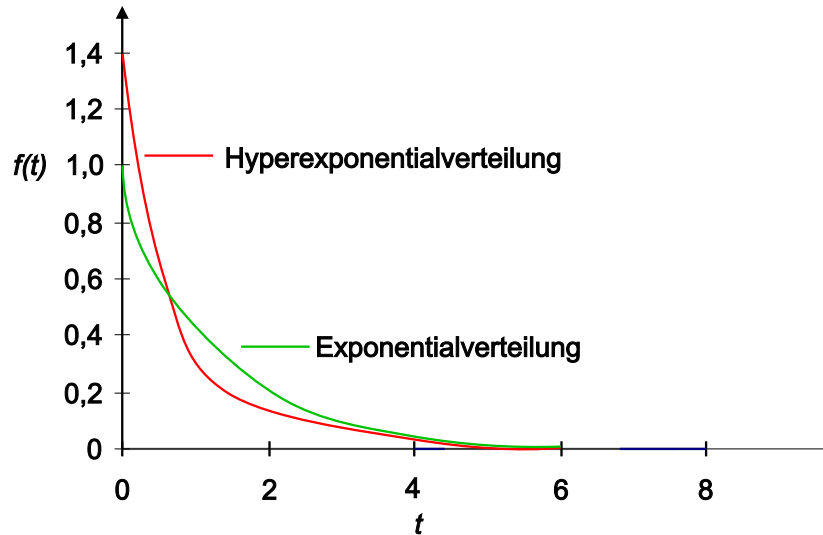


Abbildung 2.11: Vergleich einer Exponentialverteilung ($f(0) = 1$) mit einer Hyperexponentialverteilung gleichen Mittelwerts (erkennbar an $f(0) = 1,4$) mit zwei Phasen und den Parametern $q_1 = 5/6$, $q_2 = 1/6$, $\mu_1 = 10/6$, $\mu_2 = 2/6$

Die Cox-Verteilung

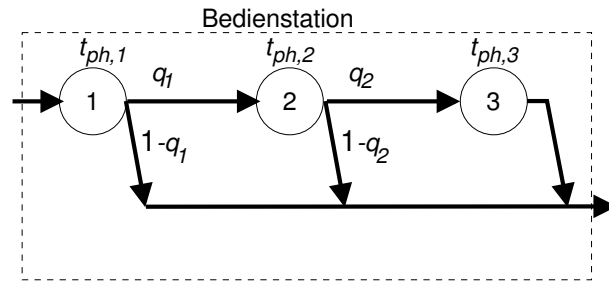
Durch Hintereinanderschaltung von n exponential verteilten Phasen mit verschiedenen Parametern μ_i , $i = 1, \dots, n$ entsteht die Cox-Verteilung, sofern mit Wahrscheinlichkeit $1 - q_i$ die Bedienung nach der i -ten Phase beendet wird (siehe Abbildung 2.12). Die Anzahl der Phasen wird auch zur Beschreibung der Cox-Verteilung verwendet, so bedeutet beispielsweise Cox-2-Verteilung (kurz C_2 -Verteilung), daß es sich um eine Cox-Verteilung mit zwei Phasen handelt.

Der Cox-Verteilung liegt die Vorstellung zugrunde, daß eintreffende Kunden eine Phase durchlaufen, um dann mit einer bekannten Wahrscheinlichkeit q_1 zumindest für eine weitere Phase im System zu verweilen oder mit der Wahrscheinlichkeit $1 - q_1$ das System zu verlassen. Diese Entscheidungen wird nach jeder Phase $i = 2, \dots, n$, mit den Wahrscheinlichkeiten q_i wiederholt, bis der Kunde das System verläßt.

Der resultierende Mittelwert der Bedienzeitverteilung kann bestimmt werden, indem für jede Phase i die Phasenbedienzeit $t_{ph,i} = 1/\mu_i$ mit der Wahrscheinlichkeit ihres Auftretens gewichtet wird.

$$E(t_{ab}) = \frac{1}{\mu_1} + \sum_{i=2}^n \left(\frac{1}{\mu_i} \cdot \prod_{j=2}^i q_{j-1} \right) \quad (2.51)$$

Aufbauend auf dem zweiten Moment, welches mit Hilfe der erwarteten, quadrierten Gesamtbedienzeit in allen Phasen bestimmt wird


 Abbildung 2.12: Modell einer Cox-Verteilung mit $n = 3$ Phasen

$$E(t_{ab}^2) = \sum_{i=2}^n \left(\prod_{j=2}^i q_{j-1} \cdot \left[\frac{1}{\mu_1^2} + \sum_{j=2}^i \frac{1}{\mu_j} \right]^2 \right) \quad (2.52)$$

kann anschließend die Varianz der Verteilung aus

$$\text{Var}(t_{ab}) = E(t_{ab}^2) - E(t_{ab})^2 \quad (2.53)$$

berechnet werden. Es kann gezeigt werden (siehe Chandy und Sauer [54] bzw. King [39]), daß mit Hilfe der Cox-Verteilung alle Verteilungen die eine rationale Laplace-Transformierte besitzen, beliebig genau approximiert werden können. Insbesondere Erlang- und Hyperexponentialverteilungen sind einfach darzustellen.

Die Schwierigkeit des praktischen Einsatzes der Cox-Verteilung liegt in der Schätzung der Parameter. Berücksichtigt man die Ähnlichkeit zur Erlang- und Hyperexponentialverteilung, so kann, einem Vorschlag von Chandy und Sauer ([54], S. 48 ff.) folgend, die Approximation wie folgt durchgeführt werden.

Für $c_{ab}^2 \leq 1$ wird als Basis eine Erlangverteilung unterstellt, bei der nach der ersten Phase mit Wahrscheinlichkeit $1 - q_1$ die Bedienung bereits beendet wird. Die Anzahl der Phasen ergibt sich dann wie bei der Erlangverteilung durch Aufrundung auf die nächstgrößere Zahl für die

$$n = \left\lceil \frac{1}{c_{ab}^2} \right\rceil \quad (2.54)$$

gilt. Mit Hilfe einer Erlangverteilung können nur Variabilitäten dargestellt werden, die als rationale Zahl $1/n$ auszudrücken sind. Für Variabilitäten in der Größenordnung von $1/10 \leq c_{ab}^2 \leq 1/2$ kann dies eine erhebliche Ungenauigkeit bedeuten. Dies kann mit Hilfe der Wahrscheinlichkeit q_1 korrigiert werden. Mit Hilfe eines Minimierungsansatzes bestimmt sie sich zu

$$q_1 = 1 - \frac{2nc_{ab}^2 + n - 2 - \sqrt{n^2 + 4 - 4nc_{ab}^2}}{2(c_{ab}^2 + 1)(n - 1)} \quad (2.55)$$

und die Bedienraten pro Phase seien

$$\mu_i = \frac{n - (1 - q_1) \cdot (n - 1)}{E(t_{ab})} \quad \forall i = 1, \dots, n \quad (2.56)$$

Die Benutzung mehrphasiger Cox-Verteilungen als Ausgangspunkt für die nachfolgende bedientheoretische Berechnung von Kennzahlen führt zu aufwendigen Berechnungsvorschriften. Die Nutzung von Cox-2-Verteilungen dagegen ist wesentlich einfacher. Auch mit Cox-2-Verteilungen kann bei geeigneter Wahl der Verteilungsparameter q_1 , μ_1 und μ_2 ein weiterer Bereich von Variabilitäten c_{ab}^2 abgedeckt werden.

Zur Schätzung der Parameter stehen die Gleichungen (2.51) bis (2.53) für $n = 2$ zur Verfügung. Sie stellen die Identität von Mittelwert und Varianz der zu approximierenden Verteilung und der sie approximierenden Cox-2-Verteilung sicher. Da insgesamt drei Parameter zu bestimmen sind, wird als zusätzliche Bedingung eingefügt, daß die Erwartungswerte der Verweilzeiten in den Phasen gleich groß sein sollen.

$$1/\mu_1 \stackrel{!}{=} q_1/\mu_2 \quad (2.57)$$

Mit Hilfe des Erwartungswertes kann ein Ausdruck für μ_1 gewonnen werden.

$$\begin{aligned} E(t_{ab}) = 1/\mu &\stackrel{!}{=} 1/\mu_1 + q_1/\mu_2 \\ \Rightarrow \mu_1 &= \frac{\mu\mu_2}{\mu_2 - q_1\mu} \end{aligned} \quad (2.58)$$

Durch Kombination von (2.57) und (2.58) ist es möglich, μ_1 zu eliminieren, und es ergibt sich nachstehender Ausdruck für q_1 :

$$q_1 = \frac{\mu_2}{2\mu} \quad (2.59)$$

Statt die Varianzen zu vergleichen, wird nachfolgend als gleichwertiges Ziel die Übereinstimmung der zweiten Momente verfolgt. Hieraus resultiert ein nur noch von den bekannten Parametern μ und c abhängiger Ausdruck.

$$\begin{aligned} E(t_{ab}^2) = \frac{1}{\mu^2}(1 + c_{ab}^2) &\stackrel{!}{=} \frac{1 - q_1}{\mu_1^2} + q_1 \cdot \left(\frac{1}{\mu_1} + \frac{1}{\mu_2} \right)^2 \\ \Rightarrow \mu_2 &= \frac{2\mu}{1 + 4c_{ab}^2} \end{aligned} \quad (2.60)$$

Setzt man dieses Ergebnis in (2.59) ein, so ergibt sich für q_1 :

$$q_1 = \frac{1}{1 + 4c_{ab}^2} \quad (2.61)$$

und für μ_1

$$\mu_1 = 2\mu \quad (2.62)$$

Die hier abgeleiteten Formeln zur Bestimmung geeigneter Parameter q_1 , μ_1 und μ_2 entsprechen nicht denjenigen, die von Marie in [50] vorgeschlagen werden ($\mu_1 = 2\mu$, $\mu_2 = \mu/c_{ab}^2$ und $q_1 = 1/(2c_{ab}^2)$). Wie leicht überprüft werden kann, können diese Ausdrücke zu teilweise erheblichen Abweichungen von den zu approximierenden Werten führen .

2.7.2 Das Bediensystem $\lambda(k)|C_n|1$

Naheliegender wäre als Erweiterung des Bediensystems $M|M|1$ die Betrachtung von Bediensystemtypen mit phasenverteilten Bedienzeiten wie zum Beispiel $M|E_n|1$. Wie in Abschnitt 2.9 gezeigt wird, können jedoch genaue Ergebnisse für mittlere Wartezeit und mittlere Zahl der Kunden in der Warteschlange unter weniger einschränkenden Verteilungsannahmen gefunden werden.

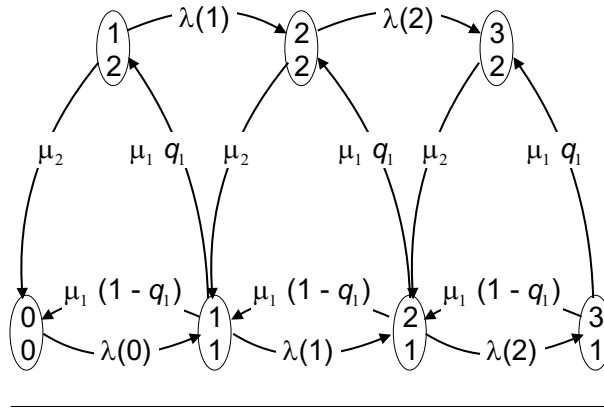
Die Bedeutung der Phasenverteilungen liegt in der Möglichkeit, *zustandsabhängige* Ankunfts- und Bedienraten modellieren zu können. Hierauf wird im Kapitel 3.5 zurückgegriffen werden, um Kennwerte für geschlossene Bediensystemnetzwerke zu berechnen.

Am Beispiel des $\lambda(k)|C_2|1$ -Bediensystems kann das Vorgehen bei der Analyse komplexerer Bediensysteme, die mit Hilfe von Phasenverteilungen beschrieben werden, gut gezeigt werden. Das Bediensystem ist charakterisiert durch einen zustandsabhängigen Ankunftsprozeß, bei dem die Ankunftsrate λ eine Funktion der bereits im Bediensystem befindlichen Zahl von Kunden k ist. Der Bedienprozeß sei durch eine Cox-Verteilung dargestellt, die jedoch die Besonderheit hat, daß nur $q_1 \neq 1$ sein kann. Mit einer solchen Cox-Verteilung können Verteilungen mit Variationskoeffizienten größer und kleiner als 1 angenähert dargestellt werden.

Der Zustandsraum eines solchen Bediensystems mit zwei Bedienphasen und zustandsabhängigen Ankunftsraten ist in Abbildung 2.13 dargestellt. Der Zustand des Bediensystems sei in diesem Fall durch einen zweidimensionalen Vektor $(k, l)^T$ dargestellt, dessen erste Komponente die Anzahl von Kunden im Bediensystem angibt. Die zweite zeigt, in welcher Phase der Bedienung sich der aktuell in der Bedienstation aufhaltende Kunde befindet.

Mit Hilfe des Zustandsdiagramms können die Gleichgewichtsgleichungen für dieses System aufgestellt und gelöst werden, so daß die Zustandswahrscheinlichkeiten $p((k, l)^T)$ berechnet werden können.

Die Gleichgewichtsgleichungen (auch als stationäre Chapman-Kolmogorov Gleichungen bekannt) lauten für $k > 1$:



Anzahl Kunden im Bediensystem (1)
In Bedienung befindlicher Kunde ist in Phase (2)

Abbildung 2.13: Zustandsdiagramm eines Bediensystems $\lambda(k)|C_2|1$ mit zwei Phasen und Ankunftsraten $\lambda(k) = 0 \quad \forall \quad k \geq 3$

$$\lambda(k-1)p(k-1,1) + \sum_{j=1}^n (1-q_j)\mu_j p(k+1,j) = [\lambda(k) + \mu_1]p(k,1) \quad (2.63)$$

und

$$\lambda(k-1)p(k-1,j) + q_{j-1}\mu_{j-1}p(k,j-1) = [\lambda(k) + \mu_j]p(k,j) \quad \forall j = 2, \dots, n \quad (2.64)$$

Sie werden ergänzt für die Randzustände $k \leq 1$:

$$\lambda(0)p(0) = \sum_{j=1}^n (1-q_j)\mu_j p(1,j) \quad (2.65)$$

$$\text{und} \quad q_{j-1}\mu_{j-1}p(1,j-1) = [\lambda(1) + \mu_j]p(1,j) \quad \forall j = 2, \dots, n \quad (2.66)$$

Die Auswertung dieser Ausdrücke kann jedoch sehr aufwendig werden, sofern die Anzahl der Phasen groß ist. Eine große Zahl von Phasen tritt bei der Modellierung getakteter oder taktähnlicher Bedienzeiten auf, da wie bei der Erlangverteilung die Abbildung kleiner Variationskoeffizienten eine große Zahl von Phasen notwendig macht.

Marie hat jedoch gezeigt (siehe [51] und [50]), daß es möglich ist, die Zustandswahrscheinlichkeiten für eine konstante Anzahl von Kunden zu kumulieren

$$p(k) = \sum_{l=1}^n p(k,l) \quad (2.67)$$

und zustandsabhängige Bedienraten $\nu(k)$, die nur noch von der Zahl der Kunden im Bediensystem abhängig sind, als

$$\nu(k) = \frac{\sum_{l=1}^n (1 - q_l) \mu_l p(k, l)}{p(k)} \quad (2.68)$$

zu definieren, um anschließend die Gleichgewichtsgleichungen für die kumulierten Wahrscheinlichkeiten zu lösen.

$$p(k) \cdot \nu(n) = \lambda(k - 1) \cdot p(k - 1) \quad \forall k = 1, 2, \dots \quad (2.69)$$

Die obenstehende Gleichung entspricht den Gleichgewichtsgleichungen der sogenannten *Geburts-* und *Sterbeprozesse* und ist strukturell identisch mit Ausdruck (2.13). Bei einer endlichen Zahl k^{limit} möglicher Zustände, wie z.B. im Falle von $\lambda(k) = 0 \quad \forall k > k^{limit}$, ist die Lösung deshalb gleichartig und lautet:

$$p(k) = p(0) \prod_{j=1}^{k^{limit}} \frac{\lambda(j - 1)}{\nu(j)} \quad \forall k = 0, 1, \dots, k^{limit} \quad (2.70)$$

$$p(0) = \frac{1}{1 + \sum_{j=1}^{k^{limit}} \prod_{k=1}^j \frac{\lambda(k-1)}{\nu(k)}} \quad (2.71)$$

Die Anwendung der Phasenverteilungen zur Abbildung des Bedienprozesses setzt voraus, daß die $\nu(k)$ bestimmt werden können. Hierfür hat Marie in [50] ausgeführt, wie für Cox-2 Verteilungen, Erlangverteilungen und Cox-n Verteilungen die Raten ν zu bestimmen sind. Die grundsätzliche Vorgehensweise soll am Beispiel der Cox-2-Verteilung dargestellt werden.

Zur Vereinfachung der Notation wird ein Wahrscheinlichkeitskoeffizient $c(n, j) = p(n, j)/p(0)$ eingeführt, der die relative Wahrscheinlichkeit eines Phasenzustandes im Verhältnis zur Wahrscheinlichkeit des leeren Systems angibt.

Die Gleichung (2.64) kann mit Hilfe der Wahrscheinlichkeitskoeffizienten zu nachstehendem Ausdruck umgeformt

$$c(k, 1) = \frac{\lambda(k) + \mu_2}{q_1 \mu_1} \cdot c(k, 2) - \frac{\lambda(k - 1)}{q_1 \mu_1} \cdot c(k - 1, 2) \quad \forall k > 1 \quad (2.72)$$

und mit Hilfe der Wahrscheinlichkeitskoeffizienten kann (2.68) zu

$$\nu(k) = \frac{\mu_1 q_1 c(k, 1) + \mu_2 c(k, 2)}{c(k, 1) + c(k, 2)} \quad (2.73)$$

$$\Rightarrow \frac{c(k, 1)}{c(k, 2)} = \frac{\mu_2 - \nu(k)}{\nu(k) - \mu_1 q_1} \quad (2.74)$$

vereinfacht werden.

Für den Fall einer Cox-2 Verteilung sind die Gleichungen (2.69) und (2.72) schließlich zu

$$\nu(k) = \frac{\lambda(k)\mu_1(1 - q_1) + \mu_1\mu_2}{(\lambda(k) + \mu_1 + \mu_2) \cdot \nu(k - 1)} \quad \forall k > 1 \quad (2.75)$$

zusammenzufassen. Für den besonderen Fall der Bedienraten $\nu(1)$ ist aus der Gleichung (2.66) in gleicher Weise

$$c(1, 1) = \frac{[\lambda(1) + \mu_2]}{q_1\mu_1} c(1, 2) \quad (2.76)$$

abzuleiten. Dies Ergebnis führt anschließend zu dem gesuchten Ausdruck für $\nu(1)$:

$$\nu(1) = \frac{\lambda(1)\mu_1(1 - q_1) + \mu_1\mu_2}{\lambda(1) + \mu_2 + q_1\mu_1} \quad (2.77)$$

2.7.3 Modellierung des Ausfalls- und Reperaturverhaltens eines FTS mit Hilfe des Bediensystems $\lambda(k)|C_n|1$

Das $\lambda(k)|C_n|1$ -Bediensystem hat neben seiner Verwendung bei der approximativen Berechnung geschlossener Bediensystemnetzwerke vor allem eine Bedeutung bei der Modellierung von Reparaturproblemen. Dies soll an einem Beispiel erläutert werden.

Eine FTS-Anlage verfüge über $K = 10$ Fahrzeuge, bei denen angenommen wird, daß die Zeit, in der ein Fahrzeug ungestört im Einsatz ist exponential verteilt mit dem Erwartungswert $E(t_f) = 120h$ ist. Tritt eine Störung auf, so sei angenommen, daß die Zeitdauer vom Eingang der Meldung bis zur Wiederinbetriebnahme des Fahrzeuges im Mittel 2 Stunden dauert. Die Variabilität c_{ab}^2 der Reparaturdauer betrage 2, d.h die Varianz der Reparaturdauer sei $8h^2$. Für die Reparaturen ist während der Betriebszeit jeweils eine Person zuständig, sind also zwei Fahrzeuge ausgefallen, so muß zunächst das erste ausgefallene Fahrzeug repariert werden, bevor mit dem zweiten Fahrzeug begonnen wird.

Der Betriebsdurchsatz der Anlage kann noch erreicht werden, wenn 8 Fahrzeuge betriebsbereit sind. Es stellt sich nun die Frage, mit welcher Wahrscheinlichkeit mindestens 8 Fahrzeuge betriebsbereit sind.

Diese Frage kann beantwortet werden, indem die defekten Fahrzeuge als Kunden modelliert werden, die die Bedienstation „Reparatur“ benötigen. Die defekten, aber noch nicht in Reparatur befindlichen Fahrzeuge stellen die Warteschlange mit der Länge N_w dar. Zusammen mit dem gerade eben reparierten Fahrzeug sind k Fahrzeuge im Bediensystem „Reparatur“.

Die Ankunftsrate der Fahrzeuge an dem Bediensystem ist von der Zahl der ungestörten Fahrzeuge $K - k$ abhängig und hat pro intaktem Fahrzeug die Rate $\lambda = 1/120h$.

Die Ankunftsrate ist in dem betrachteten Beispiel zustandsabhängig, da nur die intakten Fahrzeuge ausfallen können. Die zustandsabhängige Ankunftsrate reparaturbedürftiger Fahrzeuge beträgt damit:

$$\lambda(k) = (K - k) \frac{1}{E(t_f)} = (10 - k) \frac{1}{120h} \quad (2.78)$$

Die Abbildung des Reparatur- (bzw. Bedien-)prozesses ist mit Hilfe der obigen Angaben nur näherungsweise möglich, da die Verteilungsform nicht bekannt ist. Bei einer Variabilität von $c_{ab}^2 = 2$ bietet sich eine Modellierung der Reparaturzeitverteilung als Cox-2-Verteilung an. Der Mittelwert der Bedienzeit beträgt 2 Stunden, was einer Bedienrate von $\mu = 1/2$ entspricht.

Vor dem Beginn der Berechnung sind die Parameter der Cox-Verteilung zu schätzen. Mit Hilfe der Formeln (2.60) bis (2.62) werden folgende Werte ermittelt (siehe auch Tabelle 2.1) :

$$\begin{aligned} \mu_2 &= \frac{2 \cdot 1/2}{3 + 4 \cdot 2} = \frac{1}{11} \\ q_1 &= \frac{1}{3 + 4 \cdot 2} = \frac{1}{11} \\ \mu_1 &= 2 \cdot 1/2 = 1 \end{aligned}$$

Die Ankunftsraten $\lambda(k)$ defekter Fahrzeuge sind mit (2.78) zu bestimmen. Mit Hilfe von (2.75) und (2.77) werden anschließend die von k abhängigen Bedienraten berechnet. Im nächsten Schritt sind die Terme von (2.70) und (2.71) auszuwerten.

k	Werte für $c_{ab}^2 = 2$					Werte für $c_{ab}^2 = 0,5$
	$\lambda(k)$	$\nu(k)$	$\frac{\lambda(k-1)}{\nu(k)}$	$\prod \frac{\lambda(k-1)}{\nu(j)}$	$p(k)$	$p(k)$
0	0,0833	-	-	-	0,8411	0,8419
1	0,0750	0,5981	0,1393	0,1393	0,1172	0,1357
2	0,0667	0,2418	0,3101	0,0432	0,0363	0,0195
3	0,0583	0,5762	0,1157	0,0050	0,0042	0,0025
4	0,0500	0,2325	0,2509	0,0013	0,0011	0,0003
5	0,0417	0,5527	0,0905	0,0001	0,0001	0,0000
6	0,0333	0,2225	0,1873	0,0000	0,0000	0,0000
7	0,0250	0,5275	0,0632	0,0000	0,0000	0,0000
8	0,0167	0,2117	0,1181	0,0000	0,0000	0,0000
9	0,0083	0,5001	0,0333	0,0000	0,0000	0,0000
10	0,0000	0,1999	0,0417	0,0000	0,0000	0,0000

Tabelle 2.1: Berechnung der Wahrscheinlichkeit $p(k)$ in Reparatur befindlicher Fahrzeuge

Die Berechnung zeigt, daß mit einer Wahrscheinlichkeit von 5,3% drei und mehr Fahrzeuge defekt (in Reparatur oder auf Reparatur wartend) sind, so daß der

Durchsatz in dieser Zeit nicht erbracht werden kann. Gelingt es jedoch, die Variabilität (bzw. die Varianz) der Reparaturzeit entscheidend zu verringern (auf $c^2 = 0,5$) dann verringert sich die Wahrscheinlichkeit, daß sich mehr als $k \geq 3$ Fahrzeuge außer Betrieb befinden, auf 2,8%, obwohl die durchschnittliche Ausfallrate und Reparaturdauer gleich geblieben sind.

2.8 Little's Gesetz

Als Voraussetzung für eine weitere Verallgemeinerung der Modelle soll Little's Gesetz, das bereits in Abschnitt 2.3 für die operationelle Analyse gezeigt wurde, für stochastische Bediensysteme hergeleitet werden.

Die Beobachtung, daß die Verweilzeit der Kunden im System und die Zahl der Kunden im System zueinander proportional sind, wurde schon vor längerer Zeit gemacht. Den Zusammenhang zu beweisen und den Proportionalitätsfaktor zu bestimmen, gelang J.D. Little bereits 1961 ([47]). Eine einfachere Herleitung in Anlehnung an King [39] soll die Zusammenhänge hier erläutern.

Analog zu den Beobachtungen der operationellen Analyse sei $A(t)$ die Zahl der Ankünfte bis zum Zeitpunkt t und $C(t)$ die Zahl der Abgänge aus einem Bediensystem bis zu eben diesem Zeitpunkt. Die Zahl $V(t)$ der Kunden, die sich zum Zeitpunkt t im Bediensystem befinden, läßt sich als Differenz

$$V(t) = A(t) - C(t) \tag{2.79}$$

bestimmen.

Kumuliert man die insgesamt von allen Kunden bis zum Zeitpunkt t im System verbrachte Zeit, so ergibt sich:

$$T_v(t) = \int_0^t V(s) ds \tag{2.80}$$

Mittelt man durch Division mit der Anzahl Kunden, die diese Verweilzeit aufgebracht haben, so resultiert die mittlere Verweilzeit $t_v(t)$ als:

$$t_v(t) = \frac{T_v(t)}{A(t)} \tag{2.81}$$

Die mittlere Zahl $N(t)$ von Kunden im System bis zum Zeitpunkt t läßt sich ebenfalls mit Hilfe von $V(t)$ berechnen:

$$N(t) = \frac{1}{t} \int_0^t V(s) ds \tag{2.82}$$

Mit einer bis t gemittelten Ankunftsrate $\lambda(t) = A(t)/t$ folgt:

$$\begin{aligned}\lambda(t)t_v(t) &= \frac{A(t)}{t} \frac{\int_0^t V(s)ds}{A(t)} \\ &= \frac{1}{t} \int_0^t V(s)ds = N(t)\end{aligned}\quad (2.83)$$

Falls sich das System immer wieder leert, was für $\rho < 1$ der Fall ist, kann für obigen Ausdruck für $t \rightarrow \infty$ ein Grenzwert

$$\lambda t_v = N \quad (2.84)$$

bestimmt werden.

Die gleiche Betrachtung kann für alle Abschnitte des Bediensystems angestellt werden. So ergibt sich für die Wartezeit t_w und die Zahl der Elemente in der Warteschlange N_w entsprechend

$$\lambda t_w = N_w \quad (2.85)$$

Stellt man die Betrachtung für die Zahl der Elemente in der Bedienstation an, so resultiert

$$\lambda t_{ab} = \lambda \frac{1}{\mu} = \rho \quad (2.86)$$

was in Übereinstimmung mit der bereits getroffenen Definition für die Auslastung des Bediensystems gilt.

Für Einkanal-Bediensysteme entsteht aus 2.85 und 2.86 eine im Folgenden noch häufig gebrauchte Beziehung für die Zusammensetzung der Anzahl Kunden im System.

$$t_v = t_w + t_{ab} \quad (2.87)$$

$$\Rightarrow \lambda t_v = \lambda t_w + \lambda t_{ab} \quad (2.88)$$

$$\Rightarrow N = N_w + \rho \quad (2.89)$$

Zur Herleitung von Littles Gesetz wurden nur wenige Annahmen über die Existenz des Grenzwertes benötigt. In der Monographie von King [39] wird gezeigt, daß Littles Gesetz einen sehr weiten Geltungsbereich hat. Die durchaus konstruierbaren Ausnahmen sind jedoch aufgrund der endlichen Ausprägungen der Zeiten im Materialflußbereich dort nicht relevant. Weiterhin läßt sich belegen, daß die oben getroffenen Aussagen für die Verweilzeit auch auf offene Bediensystemnetzwerke ausgedehnt werden können, denn auf spezielle Eigenschaften eines einzelnen Bediensystems mußte nicht zurückgegriffen werden.

2.9 Das Bediensystem $M|G|1$ -FCFS

Die in Abschnitt 2.5 getroffene Annahme, daß die Bedienzeiten einer Exponentialverteilung gehorchen, ist in der Praxis nur selten verifizierbar. Zu ersten praktisch verwertbaren Abschätzungen für Wartezeit und Warteschlangenlänge kann man durch geeignete Anwendung des $M|G|1$ -Bediensystemes kommen. Hier liegt ein Markov-Prozeß bezüglich der Ankünfte vor. Die Bedienzeitverteilung wird durch eine generelle, nur durch Mittelwert und Varianz festgelegte Verteilung beschrieben. Damit können beispielsweise Materialflußelemente wie Verzweigungen abgebildet werden, die für die verschiedenen Richtungen nur wenige, annähernd diskrete Bedienzeiten erzeugen (siehe Abbildung 2.14). Die im Mittel vor einem solchen Element auftretenden Warteschlangen können mit Hilfe der Pollazek-Khintchine Formel bestimmt werden, die in den nächsten Abschnitten hergeleitet wird.

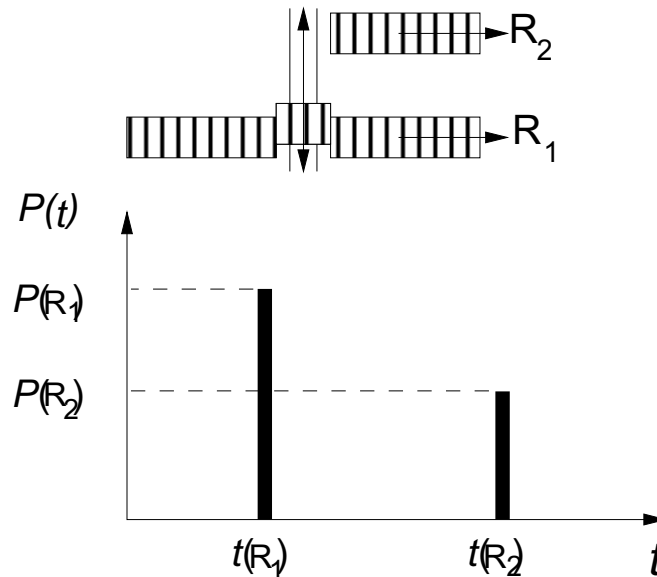


Abbildung 2.14: Teilstetiges Verzweigungselement für zwei Richtungen und seine Bedienzeitverteilung

Zur Herleitung der Wartezeit an einem $M|G|1$ -Bediensystem betrachten wir einen Kunden, der das System in seinem Durchschnittszustand antrifft. Dies ist zulässig, da der Ankunftsprozeß ein Markov-Prozeß ist und deshalb die Kunden in ihrer Gesamtheit das System in einem „Durchschnittszustand“ erleben (PASTA-Theorie, bewiesen von Wolff [66]).

Ein eintreffender Kunde finde bei seiner Ankunft $n \geq 0$ Kunden wartend und einen in Bedienung vor. Bis seine Bedienung beginnt muß also die Restbedienzeit t_r des gerade in Bedienung befindlichen Kunden und die gesamte Abfertigungszeit $t_{ab,1} + t_{ab,2} + \dots + t_{ab,n}$ der vor ihm eingetroffenen Kunden verstrichen sein. Die

Restbedienzeit tritt allerdings nur mit der Wahrscheinlichkeit $1 - p(0) = \rho$ auf, also mit der Wahrscheinlichkeit, daß die Bedienstation auch belegt ist.

Die Wartezeit t_w des ankommenden Kunden läßt sich damit als

$$t_w = (1 - p(0))t_r + \sum_{i=1}^n t_{ab,i} \quad (2.90)$$

berechnen.

Nun erfolgt der Übergang zu Mittelwerten. Zur besseren Unterscheidung werden für diesen Abschnitt die mittlere Wartezeit mit \bar{t}_w und die mittlere Restbedienzeit mit \bar{t}_r bezeichnet. Nach diesem Übergang ist der Zusammenhang aus Ausdruck 2.90 mit Hilfe der mittleren Restbedienzeit \bar{t}_r als

$$\bar{t}_w = (1 - p(0))\bar{t}_r + N_w E(t_{ab}) \quad (2.91)$$

zu beschreiben. Durch Anwendung von Littles Gesetz 2.84 und Einsetzen von $\rho = 1 - p(0)$ resultiert:

$$\begin{aligned} \bar{t}_w &= \rho \bar{t}_r + \lambda \bar{t}_w E(t_{ab}) \\ &= \rho \bar{t}_r + \bar{t}_w \rho \\ \Leftrightarrow \bar{t}_w &= \frac{\rho \bar{t}_r}{1 - \rho} \end{aligned} \quad (2.92)$$

Für ein $M|M|1$ -Bediensystem wäre diese Gleichung leicht zu lösen, denn der Erwartungswert der Restbedienzeit t_r beträgt dort aufgrund der Markov-Eigenschaft immer $1/\mu$. Im Falle allgemein verteilter Bedienzeiten ist eine nur wenig aufwendigere Überlegung notwendig.

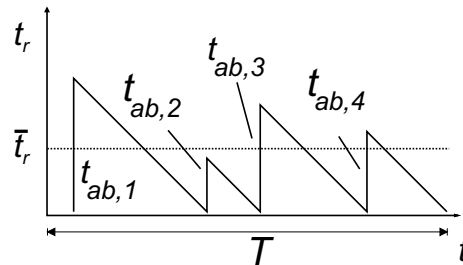


Abbildung 2.15: Verlauf der Restbedienzeit

Bei der Betrachtung der jeweiligen Restbedienzeiten t_r über der Zeit (siehe Abbildung 2.15) können die Brachperioden (Bedienstation leer) vernachlässigt werden, da hierfür die Restbedienzeit bereits mit ρ gewichtet wurde. Die mittlere Restbedienzeit findet man als Quotienten aus der Fläche F unter der Kurve und dem Betrachtungszeitraum T . Aufgrund des Sägezahnverlaufes ist die Fläche F für m bediente Kunden als

$$F = \sum_{i=1}^m \frac{t_{ab,i}^2}{2} \quad (2.93)$$

zu berechnen. Während der Bedienung der m Kunden ist die Bedienstation

$$T = \sum_{i=1}^m t_{ab,i} \quad (2.94)$$

Zeiteinheiten aktiv gewesen. Somit ergibt sich als mittlere Restbedienzeit eines zu einem beliebigen Zeitpunkt eintreffenden Kunden:

$$\bar{t}_r = \frac{\sum_{i=1}^m t_{ab,i}^2}{2 \sum_{i=1}^m t_{ab,i}} \quad (2.95)$$

Nach Grenzbetrachtung für einen unendlich langen Beobachtungszeitraum erhält man:

$$t_r = \frac{E(t_{ab}^2)}{2 \cdot E(t_{ab})} \quad (2.96)$$

Durch Einsetzen in Gleichung 2.92 resultiert:

$$t_w = \frac{\rho E(t_{ab}^2)}{2E(t_{ab})(1 - \rho)} \quad (2.97)$$

$$(2.98)$$

mit der Variabilität:

$$c_{ab}^2 = \frac{Var(t_{ab})}{E^2(t_{ab})} = \frac{E(t_{ab}^2)}{E^2(t_{ab})} - 1$$

$$t_w = \frac{\rho E(t_{ab})(1 + c_{ab}^2)}{2(1 - \rho)} \quad (2.99)$$

Mit Anwendung von Littles Gesetz können ebenfalls N_w

$$N_w = \frac{\rho^2(1 + c_{ab}^2)}{2(1 - \rho)} \quad (2.100)$$

und N

$$N = \rho + \frac{\rho^2(1 + c_{ab}^2)}{2(1 - \rho)} \quad (2.101)$$

bestimmt werden. Diese Formel ist als die Pollazek-Khintchine Formel für die mittlere Zahl der Kunden im Bediensystem bekannt.

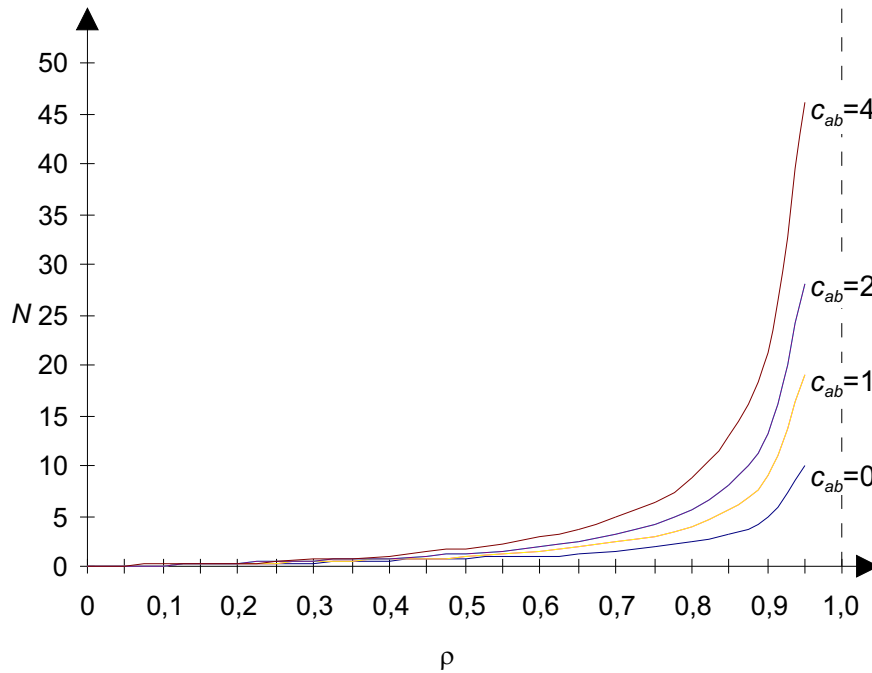


Abbildung 2.16: Verlauf der mittleren Zahl von Kunden im Bediensystem für ein $M|G|1$ -Bediensystem mit Variabilitäten $c_{ab}^2 \in \{0, 1, 2, 4\}$

Die Variabilität zur Beschreibung der Streuung einer stochastischen Größe hat den Vorzug, durch den Vergleich mit dem Mittelwert eine gegen Transformationen (z.B. der Zeiteinheit) unempfindliche Größe zu sein. Weiterhin ist durch den Vergleich mit den bekannten Variabilitäten $c_{ab}^2 = 0$ eines deterministischen Prozesses und $c_{ab}^2 = 1$ einer exponentialverteilten Größe eine überschlägige Einordnung eines stochastischen Prozesses möglich.

In Abbildung 2.16 ist zu sehen, welche wichtige Rolle der Variabilität bei der Bestimmung der Anzahl wartender Kunden zukommt. Der enge Abstand der Kurven darf nicht darüber hinweg täuschen, daß sich z.B. bei konstanter Auslastung $\rho = 0,8$ die mittlere Zahl der Kunden im System von $N = 4$ auf $N = 8,8$ mehr als verdoppelt, wenn sich die Variabilität von $c_{ab}^2 = 1$ auf $c_{ab}^2 = 4$ erhöht.

2.9.1 Modellierung von Verzweigungselementen mit Hilfe von $M|G|1$ -Bediensystemen

Mit Hilfe von $M|G|1$ -Bediensystemen können bereits realistische Kennwerte für einzelne Materialflüsselemente berechnet werden. Wie bereits in der Einleitung dargestellt, kann mit Hilfe eines $M|G|1$ -Bediensystem die mittlere Wartezeit an einem teilstetigen Verzweigungselement (siehe Abbildung 2.14) bestimmt werden. Die Spielzeiten, also die Zeit die für das Passieren der Verzweigung und das bei

der unstetigen Richtung notwendige Rückfahren des Verschiebewagens benötigt wird, stellen in diesem Fall die Bedienzeiten dar.

In einem Beispiel, zu finden bei Arnold ([3], S. 21 ff.), beträgt der Grenzdurchsatz in die unstetige Richtung 2 300 Fördeinheiten pro Stunde (FE/h), die Bedienzeit für die Fördereinheiten, die in Richtung 2 verzweigt werden, dauert demzufolge 12 Sekunden. Für die stetige Richtung 1 betragen die entsprechenden Werte 600 FE/h und 6 Sekunden.

Es sollen jeweils im Mittel 240 FE/h in Richtung 1 und 160 FE/h in Richtung 2 verzweigt werden. Unter der Annahme exponentialverteilter Zwischenankunftszeiten sind nun Kennwerte für die mittlere wartende Anzahl Fördereinheiten und die mittlere Wartezeit zu bestimmen.

Im bedientheoretischen Modell beträgt die Ankunftsrate:

$$\lambda = 160 \text{ FE/h} + 240 \text{ FE/h} = 400 \text{ FE/h.}$$

Mit den Anteilen $3/5$ für Richtung 1 und $2/5$ für Richtung 2 ergibt sich eine mittlere Bedienzeit von $3/5 \cdot 6 \text{ Sekunden} + 2/5 \cdot 12 \text{ Sekunden} = 8,4 \text{ Sekunden}$, was einer Bedienrate $\mu = 428,57 \text{ FE/h}$ gleich kommt. Dies entspricht einer Auslastung von $\rho = 0,93$.

Das zweite Moment kann aus $3/5 \cdot 6^2 \text{ Sekunden}^2 + 2/5 \cdot 12^2 \text{ Sekunden}^2$ zu $79,2 \text{ Sekunden}^2$ bestimmt werden. Damit ergibt sich eine Variabilität der Bedienzeit $c_{ab}^2 = 0,1224$.

Durch Einsetzen dieser Werte in (2.100) kann die mittlere Anzahl wartender Fördereinheiten als $N_w = 7,33 \text{ FE}$ und mit Hilfe von Littles Gesetz die mittlere Wartezeit als 66 Sekunden berechnet werden.

Obwohl diese Werte (je nach Auslegung der Fördereinrichtungen) noch beherrschbar zu sein scheinen, genügt eine Erhöhung der Ankunftsrate für Richtung 2 um 10 % auf 176 FE/h, um wahrscheinlich einen kritischen Zustand herbeizuführen, denn die Werte für $N_w = 40,91 \text{ FE}$ und $t_w = 354 \text{ Sekunden}$ steigen trotz einer Erhöhung der gesamten Ankunftsrate um lediglich 5,5 % auf ungefähr fünffache Werte.

Der gleiche nachteilige Effekt tritt auch dann ein, wenn durch eine technische Veränderung des Verzweigungselementes die Variabilität der Bedienzeiten ansteigt, selbst wenn die Auslastung konstant bleibt.

2.9.2 Losgrößenoptimierung mit Hilfe eines $M|G|1$ -Bediensystems

Das nachfolgende Anwendungsbeispiel zeigt, welche Einflüsse operative Entscheidungen (hier die Losgrößenplanung) auf die Aufgaben und Randbedingungen der Materialflußplanung (hier die Flächen- und Pufferplanung) haben und wie die

Losgrößenplanung Durchlaufzeiten und Bestände beeinflusst.

Kern bedientheoretischer Modelle ist die Abbildung von Ressourcenkonkurrenz und die daraus resultierenden Folgen für Verweil- und Wartezeiten. Dies macht es möglich, bedientheoretische Modelle zur Analyse der Auswirkungen von Losgrößenentscheidungen auf den Bestand und damit den Pufferbedarf einer Werkstattfertigung zu benutzen.

Exemplarisch sei eine einzelne Bearbeitungsmaschine betrachtet, die einen Bearbeitungsschritt durchführt, der typischerweise etwa in der Mitte eines Arbeitsplanes auftritt. Auf dieser Maschine sollen P verschiedene Produkte bearbeitet werden, deren Bedarfe $d_h, \forall h = 1, \dots, P$ im Dispositionszeitraum T_d bekannt sind. Für die Produkte sind die jeweiligen Stückbearbeitungszeiten $t_{b,h}$ und die (reihenfolgenunabhängigen) Rüstzeiten $t_{r,h}$ bekannt. Die Rüstzeiten fallen genau einmal zu Beginn der Fertigung eines neuen Loses des Produktes h an. Die Fertigung eines Produktes erfolgt in Losen stets gleicher Größe l_h .

Im Rahmen der Losgrößenplanung, die eine Teilaufgabe der Produktionsplanung darstellt (siehe Günther und Tempelmeier [35], S. 174ff.), ist darüber zu entscheiden, welche Stückzahl l_h bei einer Auflage eines Produktes auf dieser Maschine gefertigt werden soll.

Die nachfolgende Modellierung der Fertigungsmaschine als Bediensystem unterstellt, daß ein Los durch einen Kunden des Bediensystems repräsentiert wird. Die Parameter des bedientheoretischen Modells müssen aus der Losgrößenentscheidung und den zugehörigen Mengengerüsten abgeleitet werden. Hierzu wird zuerst die Losbearbeitungszeit ermittelt.

Bei nicht unterbrochener Fertigung resultiert aus der Stückzahlentscheidung eine Losbearbeitungszeit

$$t_{l,h} = t_{r,h} + l_h \cdot t_{b,h} \quad (2.102)$$

sowie eine produktspezifische Ankunftsrate.

$$\lambda_h = d_h / l_h \quad (2.103)$$

Die Ankunftsrate λ an der Bearbeitungsmaschine ergibt sich als Summe der produktspezifischen Ankunftsraten

$$\lambda = \sum_{h=1}^P \lambda_h \quad (2.104)$$

Es sei im weiteren unterstellt, daß die Zwischenankunftszeiten exponentialverteilt sind. In Abschnitt 3.3.2 wird gezeigt, daß diese Annahme als Arbeitsgrundlage akzeptabel ist.

Nachdem der Ankunftsprozeß der Fertigungslose beschrieben ist, sind nun die Parameter des stochastisch modellierten Bedienprozesses zu ermitteln. Der Anteil p_h der Lose für Produkt h am Gesamtstrom ist durch

$$p_h = \frac{\lambda_h}{\lambda} \quad (2.105)$$

gegeben und kann zur Bestimmung der mittleren Losbearbeitungszeit dienen. Diese mittlere Losbearbeitungszeit entspricht dem Erwartungswert der Bedienzeit.

$$E(t_{ab}) = \sum_{h=1}^P p_h \cdot t_{l,h} \quad (2.106)$$

Gleichermaßen ist das zweite Moment der Bedienzeit als

$$E(t_{ab}^2) = \sum_{h=1}^P p_h \cdot t_{l,h}^2 \quad (2.107)$$

zu berechnen.

Die Auslastung des Bediensystems ist durch

$$\rho = \frac{\lambda}{\mu} = \lambda \cdot E(t_{ab}) \quad (2.108)$$

zu ermitteln.

Damit sind alle Informationen bereitgestellt, die zur Ermittlung der mittleren Wartezeit an dem Bediensystem mit Hilfe von Ausdruck 2.97 benötigt werden.

Mit diesem Modell lassen sich auch die Auswirkungen einer Losgrößenveränderung auf den Fertigungsbestand darstellen.

Es sei $\vec{l} = (l_1, l_2, \dots, l_P)^T$ der transponierte Vektor der Losgrößen an der betrachteten Bearbeitungsmaschine. Es sei angenommen, daß die Losgrößen l_1, \dots, l_{P-1} fest seien und die Losgröße von Produkt P verändert werde.

Zur Vereinfachung der Darstellung seien $E(\bar{t}_{ab})$, $E(\bar{t}_{ab}^2)$, $\bar{\lambda}$ und $\bar{\rho}$ die auf die Produkte mit festgehaltenen Losgrößen bezogenen Größen.

Diese Größen sind wie folgt definiert:

$$\bar{p}_h = \frac{\lambda_h}{\sum_{i=1}^{P-1} \lambda_i}$$

$$E(\bar{t}_{ab}) = \sum_{h=1}^{P-1} \bar{p}_h \cdot t_{l,h} \quad (2.109)$$

$$E(\bar{t}_{ab}^2) = \sum_{h=1}^{P-1} \bar{p}_h \cdot t_{l,h}^2 \quad (2.110)$$

$$(2.111)$$

Die durchschnittliche Losgröße dieser Produkte sei \bar{l} .

Die Ankunftsrate λ_P von Losen des Produktes P in Abhängigkeit von der Losgröße ist durch

$$\lambda_P = \frac{d_P}{l_P} \quad (2.112)$$

beschrieben. Damit ist auch die Ankunftsrate insgesamt von l_p abhängig.

$$\lambda = \bar{\lambda} + \frac{d_P}{l_P} \quad (2.113)$$

Der Anteil p_P der Lose von Produkt P in Abhängigkeit von l_p ist durch

$$p_P = \frac{d_P/l_P}{\bar{\lambda} + d_P/l_P} \quad (2.114)$$

darzustellen.

Zur Analyse der Einflüsse werden 2.106, 2.107 und 2.108 erweitert:

$$E(t_{ab}) = p_P(t_{r,P} + l_P t_{b,P}) + (1 - p_P)E(\bar{t}_{ab}) \quad (2.115)$$

$$E(t_{ab}^2) = p_P(t_{r,P} + l_P t_{b,P})^2 + (1 - p_P)E(\bar{t}_{ab})^2 \quad (2.116)$$

$$\rho = \bar{\lambda}E(\bar{t}_{ab}) + \lambda_P(t_{r,P} + l_P t_{b,h}) \quad (2.117)$$

Damit ist die Wartezeit in Abhängigkeit von l_P auch ausdrückbar als:

$$t_w = \frac{\left(\bar{\lambda}E(\bar{t}_{ab}) + \frac{d_P(t_r+t_b l_P)}{l_P}\right)}{2 \left(E(\bar{t}_{ab}) \left[1 - \frac{d_P}{\bar{\lambda}+d_P/l_P}\right] + \frac{d_P(t_r+t_b l_P)}{l_P(\bar{\lambda}+d_P/l_P)}\right)} \cdot \frac{\left(E(\bar{t}_{ab}^2) \left[1 - \frac{d_P}{\bar{\lambda}+d_P/l_P}\right] + \frac{d_P(t_r+t_b l_P)^2}{l_P(\bar{\lambda}+d_P/l_P)}\right)}{\left(1 - \bar{\lambda}E(\bar{t}_{ab}) - \frac{d_P(t_r+t_b l_P)}{l_P}\right)} \quad (2.118)$$

Mit Hilfe von Littles Gesetz kann aus der Wartezeit auf die mittlere Anzahl wartender Aufträge geschlossen werden:

$$\begin{aligned} N_w &= t_w \cdot \lambda \\ &= t_w \cdot (\bar{\lambda} + d_P/l_P) \end{aligned} \quad (2.119)$$

Der Gesamtbestand von Aufträgen an der Maschine n resultiert schließlich aus:

$$N = N_w + \rho \quad (2.120)$$

Mit Hilfe dieses Ergebnisses kann das klassische optimale Losgrößenmodell nach *Harris* und *Wilson* bzw. *Andler* (siehe z.B. Hillier und Lieberman [36], S. 627 ff.), welches Rüstkosten K_r , mit Hilfe des Rüstkostensatzes k_r durch

$$\begin{aligned} K_r &= \sum_{h=1}^P k_r \frac{d_h}{l_h} \\ &= k_r(\bar{\lambda} + d_P/l_P) \end{aligned} \quad (2.121)$$

abbildet und Lagerhaltungskosten K_l des Fertigproduktes mit dem Lagerkostensatz k_l

$$\begin{aligned} K_l &= \sum_{h=1}^P k_l \frac{l_h}{2} \\ &= (P-1) \cdot \frac{\bar{l}}{2} \cdot k_l + k_l \frac{l_P}{2} \end{aligned} \quad (2.122)$$

bewertet, erweitert werden. Greiling zeigt in [30] (siehe auch Dörrsam, Furmans und Greiling [19]), daß eine sinnvolle Erweiterung der Kostenfunktion durch Berücksichtigung der Fertigungsbestände möglich ist. Hierzu wird mit k_f ein Kostensatz pro Einheit eines der Produkte $1, \dots, P$ eingeführt. In ihm sind die fertigungsbestandsabhängigen Kosten, wie z.B. Kapitalbindung, Flächenkosten, Verschrottungsrisiko etc. abgebildet. Der Kostensatz könnte produktspezifisch vereinbart werden, zur Vereinfachung der nachfolgenden Darstellung wird jedoch ein über alle Produkte einheitliche Kostensatz unterstellt.

Die Kosten der Fertigungsbestände sind damit durch

$$K_f = k_f \cdot N \cdot ((1 - p_P)\bar{l} + p_P l_P) \quad (2.123)$$

bestimmt.

Die Ermittlung der Kosten ist nur im Bereich zulässiger Lösungen sinnvoll, also dann wenn $\rho < 1$ gilt. Bedingung für die Existenz zulässiger Lösungen ist, daß die Auslastung zumindest bei sehr großen Losgrößen unter den Wert 1 fällt, was gleichbedeutend damit ist, daß ein Grenzwert $\lim_{l_P \rightarrow \infty} \rho < 1$ existiert. Außerdem ist in einem Los mindestens eine ganze Einheit eines Produktes herzustellen. Falls eine gültige Lösung existiert, ist die Bestimmung der unteren Grenze des Definitionsbereiches folglich mit

$$l_{P,min} = \max\left\{\frac{d_P \cdot t_r}{1 - \bar{\lambda}E(t_{ab}) - d_P \cdot t_b}; 1\right\} \quad (2.124)$$

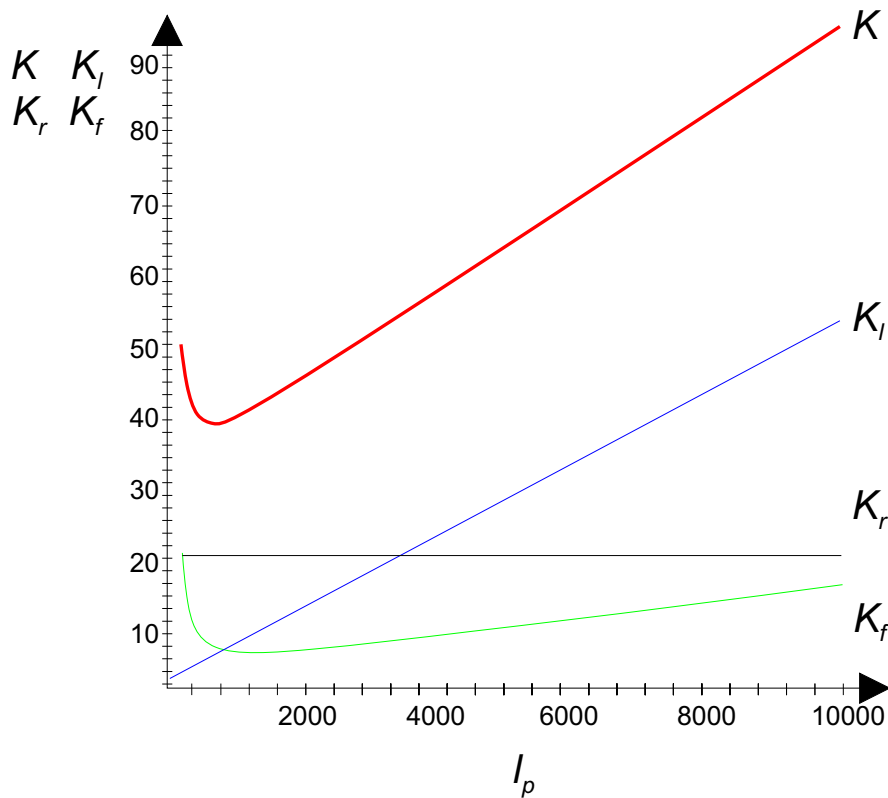
vorzunehmen.

Der Verlauf der Gesamtkostenfunktion in Abhängigkeit von der Losgröße l_P des zu betrachtenden Produktes hat im allgemeinen einen konvexen Verlauf. Der Darstellung in Abbildung 2.17 liegen die in Tabelle 2.2 aufgeführten Parameter zugrunde.

Für das klassische Losgrößenproblem ist bekannt, daß die Summe aus $K_l + K_r$ konvex ist und ein eindeutiges Minimum besitzt. Für K_f kann gezeigt werden,

$E(t_{ab})$	0,20	ZE
$E(t_{ab}^2)$	0,44	ZE ²
$\bar{\lambda}$	4,50	Lose / ZE
d_P	0,10	Stück
t_r	100,00	ZE
t_b	0,10	ZE
\bar{l}	1.000,00	Stück / Los
k_r	5,00	GE / Rüstvorgang
k_l	0,01	GE / (Stück · ZE)
k_f	0,00005	GE / (Stück · ZE)

Tabelle 2.2: Parameter des Losgrößenmodells

Abbildung 2.17: Verlauf der Kostenfunktionen K , K_r , K_l und K_f für zwei Produkte mit den Parametern aus Tabelle 2.2

daß der Grenzwert der 1. Ableitung für $l_P \rightarrow l_{P,min}$ negativ und der Grenzwert der 1. Ableitung für $l_P \rightarrow \infty$ positiv ist. Daraus folgt, daß mindestens ein echtes Minimum existiert. Es ist noch nicht gelungen, die Existenz eines eindeutigen Minimums nachzuweisen. Zahlreiche Optimierungsversuche mit variierenden Startwerten haben jedoch immer zum gleichen Ergebnis geführt. Es ist kein Fall

bekannt, bei dem zwei verschiedene Minima gefunden wurden (siehe hierzu Greiling [30], Anhang A).

In den meisten praktischen Anwendungsfällen sind für alle, oder zumindest für sehr viele Produkte die Losgrößen frei bestimmbar. Aus diesem Grund ergibt sich ein mehrdimensionales, nichtlineares Optimierungsproblem. Der Lösungsraum mit seiner Zielfunktion kann für Probleme mit zwei Produkten noch graphisch dargestellt werden. Aufgrund der anzunehmenden Konvexität des Lösungsraumes bietet sich eine Suche nach dem Gradientenverfahren (siehe z.B. Hillier und Lieberman [36], S. 448ff) an. Dabei wird ausgehend von einem Startvektor $(l_1, l_2)^T$ jeweils ein Schritt in Richtung des Gradienten gemacht. Nach diesem Schritt wird der neue Zielfunktionswert ermittelt. Ist dieser im Vergleich zum vorherigen Wert verbessert, wird der Gradient erneut bestimmt und ein weiterer Schritt in dessen Richtung vorgenommen. Das Verfahren terminiert, wenn die Verbesserung des Zielfunktionswertes unterhalb eines vorher bestimmten Schwellwertes liegt.

Für die Flächen- und Layoutplanung werden neben den Flächenbedarfen und gegebenenfalls Abmessungen der Betriebsmittel einschließlich der Wartungs- und Zugangsflächen sowie den Flächen für Wege auch die Pufferflächen in der Fertigung für halbfertige Produkte benötigt.

Mit Hilfe der bedientheoretisch ermittelten mittleren Zahl von Losen in der Warteschlange können diese Flächenbedarfe zumindest als Mittelwerte näherungsweise bestimmt werden (siehe auch Furmans [26]). Hierfür ist für jedes Produkt mit Hilfe der ermittelten oder vorgegebenen Losgröße l_h die notwendige Behältergröße und -anzahl zu bestimmen. Darauf aufbauend kann der erwartete Flächenbedarf a_h pro Los des Produktes h bestimmt werden.

Die Zusammensetzung der Warteschlange ist proportional zu den Anteilen p_h der Produkte am Gesamtfluß λ durch das Bediensystem. Der mittlere Flächenbedarf A_h für alle Lose des Produktes h in der Warteschlange beträgt somit

$$A_h = N_w \cdot a_h p_h \quad (2.125)$$

und der mittlere Gesamtflächenbedarf für den Bestand an unbearbeiteten Aufträgen ist mit

$$A^{\text{gesamt}} = \sum_{i=1}^P A_h \quad (2.126)$$

zu bestimmen.

Ein Ansatz für eine genauere Bestimmung des Flächenbedarfs unter Berücksichtigung verschiedener Pufferkonfigurationen und der Wahrscheinlichkeitsverteilung für die Zahl der Lose in der Warteschlange wird mit Hilfe von zeitdiskreten Bediensystemen durch Furmans und Zillus in [27] dargestellt.

2.10 Das $M|G|1$ -Bediensystem mit Prioritäten

Es gibt zahlreiche Versuche, durch die Einführung von Prioritätsregeln die Reaktionszeit oder Durchlaufzeit eines Materialflusssystems zu verkürzen oder andere Kenngrößen zu verbessern. Mit Hilfe einfacher, auf dem $M|G|1$ -Bediensystem basierenden Überlegungen, läßt sich zeigen, daß die Auswirkungen von Prioritätsregeln einigen Grundregeln genügen müssen.

Die nachfolgenden Betrachtungen beschränken sich auf Prioritätsregeln, die *arbeitserhaltend* sind, d.h. auch bei Unterbrechung der Arbeit geht die bereits geleistete Arbeit nicht verloren. Die Eigenschaft arbeitserhaltend sei wie folgt definiert:

Definition 2.2 (Arbeitserhaltung) *Eine Prioritätsregel ist arbeitserhaltend, wenn ein Kunde mit Bedienzeitbedarf t_{ab} zum Zeitpunkt t im Umfang $\alpha \cdot t_{ab}$ bedient wurde und dessen Restbedienzeit t_r nach Unterbrechung noch $t_r = t_{ab} \cdot (1 - \alpha)$ beträgt.*

Zuerst wird die noch zu leistende Arbeit im System betrachtet. Es zeigt sich, daß diese von arbeitserhaltenden Reihenfolgeregeln nicht zu beeinflussen ist. Die Summe der Arbeitszeit, die im Bediensystem zu einem Zeitpunkt t noch zu leisten ist, bildet einen stochastischen Prozeß, der mit $\{V(t)\}$ bezeichnet sei¹. $V(t)$ besteht aus der Restbedienzeit t_r des gerade in Bedienung befindlichen Kunden und der Summe der Bedienzeiten $t_{ab,i}$ der wartenden Kunden. Ein beispielhafter Verlauf von $V(t)$ ist in Abbildung 2.18 gezeigt.

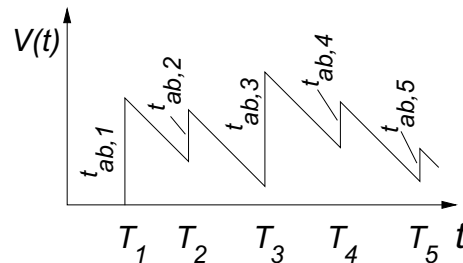


Abbildung 2.18: Verlauf der noch zu leistenden Arbeitszeit $V(t)$ im Bediensystem. Die Ankunftszeitpunkte der Kunden sind mit T_1 bis T_5 gekennzeichnet

Satz 2.1 *Für den Erwartungswert $E(V)$ der noch zu leistenden Arbeitszeit $V(t)$ in einem Bediensystem mit einer Bedienstation gilt:*

¹Der Variablenbezeichner $V(t)$ rührt von der englischen Bezeichnung *virtual work* her.

$$E(V) = \lim_{t \rightarrow \infty} \frac{\int_0^t V(u) du}{t} \quad (2.127)$$

sofern der Grenzwert existiert. $E(V)$ ist gleich für alle Reihenfolgen.

Dies kann durch folgende Überlegungen begründet werden (Beweis siehe King [39]): Die Abarbeitungsrate (Steigung) ist konstant und unabhängig von der Ankunfts- und Bearbeitungsreihenfolge. Da die Zugangszeitpunkte unabhängig von der Abarbeitungsreihenfolge sind, ist auch der Gesamtverlauf unabhängig von der Bedienungsreihenfolge.

Es ist noch anzumerken, daß $E(V)$ zwar unabhängig von der Reihenfolge ist, die Anzahl der Kunden in der Warteschlange allerdings sehr wohl durch die Reihenfolge beeinflussbar ist, wie in Abschnitt 2.10.2 deutlich werden wird.

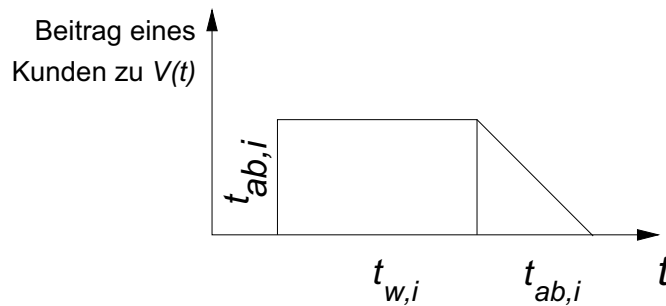


Abbildung 2.19: Verlauf des Beitrages eines Kunden zu $V(t)$

Wie ist $E(V)$ zu bestimmen? In Abbildung 2.19 ist zu sehen, welchen Beitrag ein Kunde zu $E(V)$ vom System fordert. Der Beitrag besteht aus dem gesamten Arbeitsinhalt für die Dauer der Wartezeit und der über der Zeit abnehmenden Restbedienzeit für die Dauer der Abarbeitung.

$$V(t) = \int_0^t V(u) du = \sum_{i=1}^{A(t)} (t_{w,i} \cdot t_{ab,i} + t_{ab,i}^2/2) + \epsilon(t) \quad (2.128)$$

Der Term $\epsilon(t)$ bildet als Fehlerterm die Restbedienzeit des noch in Bedienung befindlichen Kunden ab. Dividiert man 2.128 durch t um zu der Betrachtung von Raten überzugehen und läßt $t \rightarrow \infty$ streben, dann strebt gleichzeitig auch die Zahl der bedienten Kunden n gegen ∞ . Es resultiert nachstehender Ausdruck:

$$E(V) = \lambda \left(\underbrace{E(t_w t_{ab})}_{\lim_{n \rightarrow \infty} \sum_{i=1}^n \frac{t_{ab,i} t_{w,i}}{n}} + \underbrace{E(t_{ab}^2)/2}_{\lim_{n \rightarrow \infty} \sum_{i=1}^n \frac{t_{ab,i}^2}{2n}} \right) \quad (2.129)$$

Der Term $\epsilon(t)/t$ verschwindet, da der Wertebereich von $\epsilon(t)$ begrenzt ist.

Setzt man $E(V) = t_w$, so kann aus 2.129 auch die Pollazek-Khinchine Formel (2.101) hergeleitet werden.

Welche Auswirkungen kann die Anwendung von Prioritätsregeln haben? Hierzu werden die nachfolgenden Betrachtungen auf zwei Typen von Prioritätsregeln beschränkt, auf

unterbrechende², d.h. den gerade in Bedienung befindlichen Kunden verdrängende, und

nicht unterbrechende³, d.h. erst nach Beendigung der Bedienung des aktuellen Kunden eingreifende

Regeln. Außerdem ist (s.o.) vorausgesetzt, daß die Unterbrechungen arbeitserhaltend seien.

Die Kunden werden in durchnummerierte Klassen $j = 1, 2, \dots$ eingeteilt, wobei die Klasse 1 die höchste Priorität besitzt. Es sei p_j die Wahrscheinlichkeit, daß ein eintreffender Kunde der Klasse j angehört und die erwartete Bedienzeit $E(t_{ab,j})$ fordert. Der Erwartungswert der Bedienzeit $E(t_{ab,ges})$ über alle Klassen kann durch

$$E(t_{ab,ges}) = \sum_j p_j E(t_{ab,j}) \quad (2.130)$$

berechnet werden. Die partielle Auslastung des Bediensystems durch Kunden der Klasse j ist als

$$\rho_j = \lambda p_j E(t_{ab,j}) \quad (2.131)$$

definiert. Hieraus resultiert eine Gesamtauslastung von:

$$\rho = \lambda E(t_{ab,ges}) = \sum_j \rho_j \quad (2.132)$$

Die mittlere Anzahl von wartenden Kunden einer Klasse j , die sich im System befindet, sei mit $N_{w,j}$ bezeichnet und $N_w = \sum_j N_{w,j}$ sei die Gesamtzahl wartender Kunden. $E(t_{w,j})$ ist die mittlere Wartezeit der Kunden der Klasse j und $E(t_w) = \sum_j p_j t_{w,j}$ die mittlere Wartezeit aller Kunden im System.

Kleinrocks Erhaltungssatz

Für Bediensysteme mit Prioritäten, bei denen kein systematischer Zusammenhang zwischen Bedienzeit und Wartezeit besteht, also $E(t_w \cdot t_{ab}) = E(t_w)E(t_{ab})$ gilt, kann *Kleinrocks Erhaltungssatz* bewiesen werden.

²Im Englischen als preemptive, kurz PRE bekannt und als solche in Übersichten aufgeführt

³Im Englischen als non-preemptive, kurz NONPRE, aufgeführt

Durch Erweiterung von $N_{w,j} \cdot E(t_{ab,j})$ mit Hilfe von Littles Gesetz in der Form $N_w = \lambda E(t_w)$ erhält man

$$N_{w,j} \cdot E(t_{ab,j}) = \overbrace{\lambda \cdot p_j \cdot E(t_{w,j})}^{N_{w,j}} E(t_{ab,j}) = \rho_j \cdot E(t_{w,j}) \quad (2.133)$$

als Ausdruck für den durchschnittlichen Arbeitsvorrat für Kunden der Klasse j . Der Erwartungswert $E(V)$, wie in Gleichung 2.129 beschrieben, kann wie dort aus zwei Komponenten, jeweils eine für den in der Warteschlange befindlichen Teil $E(V_{Warten})$ und den in Bedienung befindlichen Teil $E(V_{Bedienung})$ zusammengesetzt werden.

$$\begin{aligned} E(V) &= \underbrace{\sum_j \rho_j \cdot t_{w,j}}_{E(V_{Warten})} + \overbrace{\lambda \frac{E(t_{ab}^2)}{2}}^{E(V_{Bedienung})} \\ &= \sum_j N_{w,j} \cdot E(t_{ab,j}) + \lambda \cdot \frac{E(t_{ab}^2)}{2} \end{aligned} \quad (2.134)$$

Da $E(V)$, wie bereits in Satz 2.1 gezeigt, unabhängig von der Reihenfolge der Bearbeitung ist, gilt

Satz 2.2 (Kleinrocks Erhaltungssatz)

$$E(V_{Warten})N_{w,j}E(t_{ab,j}) = \sum_j \rho_j t_{w,j} = \sum_j = \rho t_{w,FIFO} \quad (2.135)$$

falls die Erwartungswerte von t_w und t_{ab} voneinander unabhängig sind.

Aus Kleinrocks Erhaltungssatz läßt sich ableiten, daß durch Priorisierungen, die nicht auf der Dauer der Bedienzeit beruhen, keine Verkürzung der mittleren Wartezeit möglich ist. Die Wartezeit wird lediglich zwischen den Klassen verschoben. Der Erwartungswert für die mittlere Wartezeit bei Bediendisziplin FIFO kann deshalb beispielsweise auf der Basis eines $M|G|1$ -Bediensystems ermittelt werden. Das Theorem gilt jedoch auch in allgemeineren Fällen, die über $G|G|1$ -Bediensysteme hinausgehen (siehe Kleinrock [41]).

In den nachfolgenden beiden Abschnitten wird dargestellt werden, wie sich die Wartezeiten der einzelnen Klassen bei verschiedenen Priorisierungen verändern.

2.10.1 $M|G|1$ -Bediensystem mit nicht unterbrechenden Prioritäten

In diesem Abschnitt wird die Wartezeit für verschiedene Kundenklassen sukzessive ermittelt. Es wird angenommen, daß die Kunden der Klasse 1 die höchste Priorität haben. Die Betrachtung beschränkt sich zuerst auf zwei Klassen, da sich die Ergebnisse für mehrere Kundenklassen durch Subsummieren aller Klassen mit höherer Priorität in eine einzige Klasse ermitteln lassen.

Bei nicht unterbrechenden Prioritäten (auch verschiebbare Prioritäten genannt) wird, falls gerade ein Kunde einer niedrigeren Prioritätsklasse in der Bedienstation ist, der soeben eintreffende höher priorisierte Kunde erst dann bedient, wenn die Bedienung des in der Bedienstation befindlichen Kunden abgeschlossen ist.

Zur Vereinfachung der Notation werden nachfolgend die Erwartungswerte (z.B. $E(t_w)$) verkürzt notiert (in diesem Fall durch t_w). Ausnahme sind die zweiten Momente der Bedienzeit (z.B. $E(t_{ab,ges}^2)$)

Die Wartezeit $t_{w,1}$ von Kunden der Klasse 1, besteht aus der Bedienzeit aller bereits eingetretener Kunden der Klasse 1 sowie aus der Restbedienzeit des gerade in Bedienung befindlichen Kunden.

$$\begin{aligned}
 t_{w,1} &= \sum \left(\begin{array}{l} \text{Bedienzeit aller bereits} \\ \text{eingetretener Kunden} \\ \text{Klasse 1} \end{array} \right) + \text{Restbedienzeit des gerade} \\
 &\quad \text{bedienten Kunden} \\
 t_{w,1} &= \rho_1 t_{w,1} + \frac{\lambda E(t_{ab,ges}^2)}{2} \qquad (2.136)
 \end{aligned}$$

Wird obiger Ausdruck nach $t_{w,1}$ aufgelöst, so ergibt sich:

$$t_{w,1} = \frac{\lambda E(t_{ab}^2)}{2(1 - \rho_1)} = \frac{\rho E(t_{ab}^2)}{2E(t_{ab})(1 - \rho_1)} \qquad (2.137)$$

Dieser Ausdruck unterscheidet sich von Formel (2.99) dadurch, daß ein Kunde der Klasse 1 lediglich auf Kunden seiner eigenen Klasse warten muß, deshalb tritt in Ausdruck (2.137) im Nenner nur $1 - \rho_1$ auf. Der Ausdruck für die Restbedienzeit ist jedoch identisch, da für die Dauer der Belegung der Bedienstation die Bedienzeit der Gesamtheit der Kunden maßgeblich ist, denn die Bedienung des aktuell bedienten Kunden wird in jedem Fall abgeschlossen bevor ein Kunde der Klasse 1 bedient wird.

Für einen soeben eingetroffenen Kunden der Klasse 2 besteht die Wartezeit aus den Summanden

- Restbearbeitungszeit des gerade bedienten Kunden
- Bearbeitungszeit aller bereits eingetroffenen Kunden der Klasse 1

- Bearbeitungszeit aller bereits wartenden Kunden der Klasse 2
- Bearbeitungszeit aller während der Wartezeit eintreffenden Kunden der Klasse 1

Durch die während der Wartezeit eintreffenden Kunden der Klasse 1 verlängert sich die Wartezeit $t_{w,2}$ der Kunden der Klasse 2 zusätzlich, denn während $t_{w,2}$ treffen $\lambda_1 t_{w,2}$ zusätzliche Kunden der Klasse 1 ein, um deren gesamte Bearbeitungszeit $\lambda_1 t_{w,2} E(t_{ab,1})$ sich die Wartezeit der Kunden der Klasse 2 verlängert. Die gesamte Wartezeit ist also aus

$$t_{w,2} = \lambda E(t_{ab,ges}^2)/2 + \rho_1 t_{w,1} + \rho_2 t_{w,2} + \lambda_1 t_{w,2} E(t_{ab,1}) \quad (2.138)$$

zu bestimmen. Wird diese Gleichung nach $t_{w,2}$ aufgelöst, so resultiert:

$$t_{w,2}(1 - \rho_1 - \rho_2) = \rho_1 t_{w,1} + \lambda E(t_{ab}^2)/2$$

Durch Einsetzen von 2.137 folgt:

$$\begin{aligned} t_{w,2} &= \frac{\rho_1 \lambda E(t_{ab}^2)/2}{(1 - \rho_1)(1 - \rho_1 - \rho_2)} + \frac{\lambda E(t_{ab}^2)/2}{1 - \rho_1 - \rho_2} \\ \Leftrightarrow t_{w,2} &= \frac{\lambda E(t_{ab}^2)}{2(1 - \rho_1)(1 - \rho_1 - \rho_2)} \end{aligned} \quad (2.139)$$

Zum besseren Verständnis kann die rechte Seite von Ausdruck (2.139) in zwei Teile zerlegt werden:

$$t_{w,2} = \underbrace{\frac{\lambda E(t_{ab}^2)}{2(1 - \rho_1 - \rho_2)}}_A \cdot \underbrace{\frac{1}{1 - \rho_1}}_B \quad (2.140)$$

Die Wartezeit aufgrund der Gesamtauslastung wird durch den Faktor A repräsentiert und entspricht mit $\rho = \rho_1 + \rho_2$ der erwarteten Wartezeit für ein $M|G|1$ -Bediensystem mit der Auslastung ρ . Die durch die Priorisierung von Klasse 1 entstehende Wartezeit wird durch den Faktor $B = 1/(1 - \rho_1)$ erfasst, er wird deshalb auch als *Expansionsterm* bezeichnet.

In der allgemeinen Form für beliebig viele Klassen lautet die Formel für die Berechnung der mittleren Wartezeit für die k -te Klasse (siehe Wolf [66], S. 441):

$$t_{w,k} = \frac{\lambda E(t_{ab}^2)}{2(1 - \sum_{i < k} \rho_i)(1 - \sum_{i \leq k} \rho_i)} \quad (2.141)$$

Grundsätzlich ist die Wartezeit der priorisierten Kunden geringer als die der nicht priorisierten Kunden. Jedoch wird dieser Effekt immer geringer, wenn mehr Kunden in die priorisierte Klasse aufgenommen werden. In Abbildung 2.20 ist zu sehen, wie sich die Wartezeit der Kunden mit der höchsten Priorität vergrößert, wenn sich der Anteil der priorisierten Kunden bei ansonsten gleichen Parameterwerten des Bediensystems erhöht.

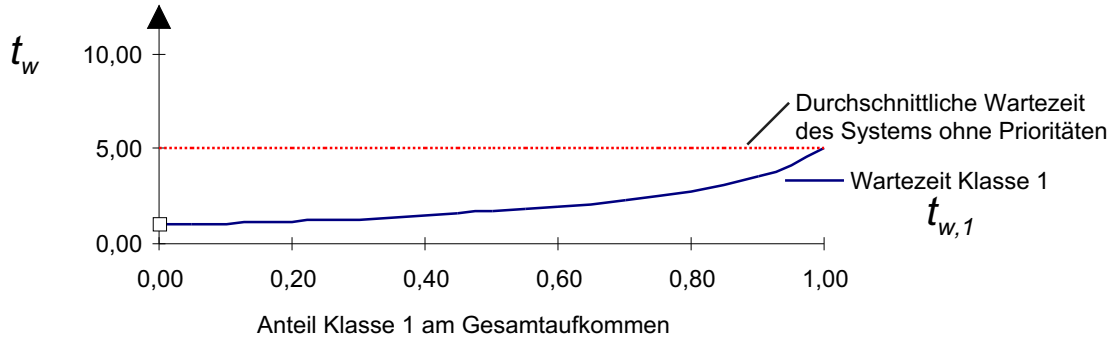


Abbildung 2.20: Entwicklung der Wartezeiten für die priorisierte Klasse eines $M|G|1$ -Bediensystem mit zwei Klassen bei Erhöhung des Anteils der priorisierten Klasse am Gesamtstrom.

Parameterwerte: $\lambda = 2$; $\rho_{ges} = 0,8 = \text{konst.}$; $E(t_{ab}^2) = 1$

Zum Vergleich Wartezeit des äquivalenten Bediensystems mit gleicher Auslastung und einer Klasse

Gleichzeitig kommt es bei steigender Zahl der priorisierten Kunden zu einer überproportionalen Steigerung der Wartezeit von Kunden der Klasse 2 (siehe Abbildung 2.21), der dadurch zu erklären ist, daß ein immer kleiner werdender Teil der Kunden die Wartezeitverkürzung von immer mehr priorisierten Kunden kompensiert, da die mittlere Gesamtwartezeit konstant bleibt.

In der Praxis kann dieser Effekt z.B. bei der Prioritätssteuerung in PPS-Systemen beobachtet werden. Wenn bei steigendem Auslastungsgrad der Fertigung beobachtet wird, daß die Durchlaufzeiten steigen, wird unter Umständen versucht, durch Priorisierung wichtiger Aufträge eine Verkürzung der Durchlaufzeiten für diese Aufträge zu erreichen.

Dadurch werden die Durchlaufzeiten anderer, nicht priorisierter Aufträge noch länger, wodurch die Notwendigkeit entsteht, noch mehr Aufträge zu priorisieren. Dies kann beliebig fortgesetzt werden, da immer mehr nicht priorisierte Aufträge (die hier den Kunden entsprechen) durch zu lange Durchlaufzeiten kritisch werden. Gleichzeitig wächst auch die Durchlaufzeit der priorisierten Aufträge, so daß die Notwendigkeit zu entstehen scheint, noch mehr Priorisierungsklassen zu schaffen.

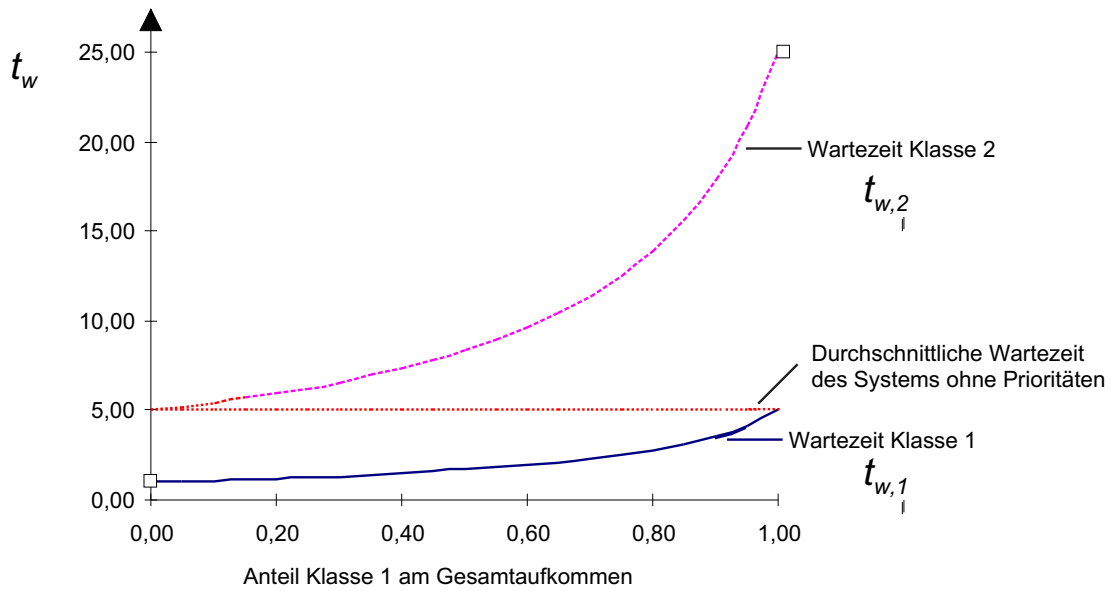


Abbildung 2.21: Entwicklung der Wartezeiten für die priorisierte und die nicht priorisierte Klasse eines $M|G|1$ -Bediensystem mit zwei Klassen bei Verschiebung des Anteils der priorisierten Klasse am Gesamtstrom.

Parameterwerte: $\lambda = 2$; $\rho_{ges} = 0,8 = \text{konst.}$; $E(t_{ab}^2) = 1$

Zum Vergleich Wartezeit des äquivalenten Bediensystems mit gleicher Auslastung und einer Klasse

2.10.2 Modellierung von Zusammenführungselementen mit Hilfe von $M|G|1$ -Bediensystemen mit nicht unterbrechenden Prioritäten

Mit Hilfe eines $M|G|1$ -Bediensystems können neben Verzweigungselementen (siehe Abschnitt 2.9.1) unter anderem auch stetige oder teilstetige Zusammenführungen wie sie in Abbildung 2.22 gezeigt sind, modelliert werden.

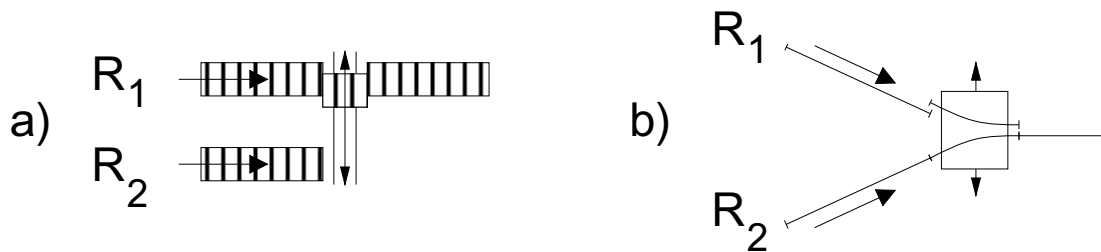


Abbildung 2.22: Teilstetiges (a) und stetiges (b) Zusammenführungselement für zwei Richtungen, nach [3]

Hierbei bildet die Bedienzeit den Zeitbedarf ab, der aus der Durchfahrt einer Fördereinheit durch das Zusammenführungselement resultiert. Sie besteht aus der Zeit, die eine Fördereinheit zum Durchfahren des Zusammenführungselementes benötigt und zusätzlich im un stetigen Fall der Zeit, die das Zusammenführungselement für die Rückkehr in die Ausgangsstellung benötigt. Beim teilstetigen Element ist dies die Zeit, um den Verschiebewagen aus der Ausfahrtposition für Richtung 2 wieder in die Ausgangsposition zurückzubewegen, so daß die nachfolgende Fördereinheit in das Verzweigungselement einfahren kann.

Sind die gleichen technischen Voraussetzungen wie in Abschnitt 2.9.1 gegeben, so treten bei Abfertigung in der Folge des Eintreffens (sogenannte stochastische Abfertigung) die gleichen Wartezeiten auf wie bei einem äquivalenten Verzweigungselement. Allerdings verteilen sich die wartenden Fördereinheiten nun auf zwei vor dem Zusammenführungselement liegende Staustrecken.

Soll dagegen eine der beiden Richtungen nach der Regel „Eingeschränkte Vorfahrt“ (siehe Arnold [3], S. 32) priorisiert werden, so sind die Wartezeiten für beide Richtungen getrennt zu ermitteln. Da es nicht sinnvoll ist, eine bereits begonnene Bedienung einer Fördereinheit zu unterbrechen, kann bei ansonsten gleichen Voraussetzungen das $M|G|1$ -Bediensystemmodell mit nicht unterbrechenden Prioritäten zur Ermittlung der mittleren Wartezeiten verwendet werden.

Die bereits in Abschnitt 2.9.1 ermittelten Werte für λ , $E(t_{ab})$ und $E(t_{ab}^2)$ können weiterverwendet werden. Die Auslastung für die jeweiligen Richtungen betragen $\rho_1 = 0,4$ und $\rho_2 = 0,53$.

Zur Berechnung der Wartezeit werden die Ausdrücke 2.137 und 2.139 verwendet.

Priorität für die stetige Richtung: In diesem Fall hat die stetige Richtung 1 Priorität und ist identisch mit Klasse 1. Wir erhalten für diese Fördereinheiten den Wert $t_{w,1} = 7,3$ Sekunden. Für die Fördereinheiten der Klasse 2 ergibt sich $t_{w,2} = 110$ Sekunden. Eine mit ihrem Anteil am Gesamtstrom gewichtete Betrachtung der Wartezeiten ergibt eine mittlere Wartezeit $t_w = 3/5 \cdot 7,3 + 2/5 \cdot 110 = 48,4$ Sekunden. Diese Verringerung resultiert aus der systematischen Bevorzugung der Klasse mit den kürzeren Bedienzeiten. Aus diesem Grund gilt hier Kleinrocks Erhaltungssatz nicht.

Priorität für die un stetige Richtung: Für diese Konstellation, bei der die Richtung 2 der Klasse 1 entspricht, ergeben sich $t_{w,1} = 9,4$ Sekunden und $t_{w,2} = 141,4$ Sekunden. Die mittlere Wartezeit beträgt 88,6 Sekunden. Durch Priorisierung der Kunden mit der längeren Abfertigungszeit tritt also eine Verschlechterung der Verhältnisse durch Umkehrung der vorigen Effekte ein. Dies entspricht zusammen mit dem obigen Ergebnis auch den aus der Reihenfolgeplanung für Ein-Maschinen-Probleme bekannten Gesetzen.

Auch wenn es gelingt, die mittlere Wartezeit einer Kundenklasse zu reduzieren, so ist zu berücksichtigen, daß die Belastung des Gesamtsystem unverändert ist, und daß die mittlere Zeit bis zur Beendigung der Bedienung der Kunden der letztpriorisierten Klasse bei eingeschränkter Vorfahrt gleich groß bleibt.

2.10.3 $M|G|1$ -Bediensystem mit unterbrechenden Prioritäten

Eine verstärkte Form der Priorisierung sind die unterbrechenden Prioritäten. Im Falle unterbrechender Prioritäten werden Kunden geringerer Priorität, die sich in Bearbeitung befinden, sofort unterbrochen, und die Bedienung des Kunden der höheren Klasse beginnt.

Da Kunden der niederen Prioritätsstufen Kunden der höchsten Priorität nie behindern, wird die Warte- und Verweilzeit der Kunden der höchsten Prioritätsstufe nur durch Aufträge der gleichen Klasse beeinflusst. Die Wartezeit der Kunden der höchsten Prioritätsstufe kann deshalb mit Hilfe eines einfachen $M|G|1$ -Bediensystems ermittelt werden, bei dem nur die Ankunftsrate und die Bedienzeitparameter der höchstpriorisierten Klasse eingesetzt werden.

Die Wartezeit $t_{w,1}$ beträgt deshalb:

$$t_{w,1} = \frac{\lambda_1 E(t_{ab,1}^2)}{2(1 - \rho_1)} \quad (2.142)$$

Für Kunden der Klasse 2 bleibt die Wartezeit die gleiche wie im Falle der nicht unterbrechenden Prioritäten, denn bis zum Beginn der Bedienung eintreffende Kunden der Klasse 1 werden auf jeden Fall vor Beginn der Bedienung eines Kunden der Klasse 2 abgefertigt. Ein Unterschied ergibt sich jedoch für die Verweilzeit $t_{v,2}$, da die Verweilzeit durch mögliche Unterbrechungen während der Bedienung verlängert werden kann.

Unterstellt man, daß die Unterbrechung arbeitserhaltend ist, das heißt, daß die beanspruchte Ressource insgesamt nicht länger beansprucht wird, als es bei ununterbrochener Abfertigung der Fall gewesen wäre, so besteht die Verweilzeit eines Kunden der Klasse 2 aus

- der Wartezeit $t_{w,2}$ (zu bestimmen nach 2.139) und
- der Bedienzeit $t_{ab,2}$ sowie
- der Bedienzeit von Kunden der Klasse 1, die während der Bedienung, oder während dieselbe bereits durch Kunden der Klasse 1 unterbrochen wurde, eintreffen.

Sei $t_{k,2}$ die Zeit, die vom Beginn der Bedienung eines Kunden der Klasse 2 bis zur Beendigung der Bedienung vergeht, dann gilt:

$$\begin{aligned} t_{k,2} &= E(t_{ab,2}) + t_{k,2}\lambda_1 E(t_{ab,1}) \\ &= E(t_{ab,2}) + t_{k,2}\rho_1 \\ \Leftrightarrow t_{k,2} &= \frac{E(t_{ab,2})}{1 - \rho_1} \end{aligned} \quad (2.143)$$

Damit ist die Verweilzeit $t_{v,2}$ mit Hilfe des nachstehenden Ausdrucks zu berechnen:

$$t_{v,2} = t_{w,2} + t_{k,2} = \frac{\lambda E(t_{ab}^2)}{2(1 - \rho_1)(1 - \rho_1 - \rho_2)} + \frac{E(t_{ab,2})}{1 - \rho_1} \quad (2.144)$$

Die Zeit ab Beginn der Bedienung verlängert sich für Kunden der Klasse 2 im Vergleich zu nicht unterbrechenden Prioritäten um den Faktor $1/(1 - \rho_1)$.

Der Vergleich der Verweilzeiten (siehe Ausdruck (2.144)) bzw. Wartezeiten (siehe (2.141)) bei unterbrechenden und bei nicht unterbrechenden Prioritäten zeigt, daß die Warte- und Verweilzeiten für Klasse 1 nur dann verändert werden, wenn sich der Erwartungswert der Bedienzeit von Klasse 1 vom Mittelwert der Gesamtheit aller Kunden unterscheidet. Eine Verkürzung der mittleren Verweilzeit ist also nur möglich, wenn der Erwartungswert der Bedienzeit der Klasse 1 kleiner als der der Gesamtheit aller Kunden ist.

2.10.4 Modellierung von Störungen mit Hilfe unterbrechender Prioritäten

Mit Hilfe der im vorigen Abschnitt vorgestellten Modelle ist der Einfluß von Störungen auf die Leistungskennwerte eines Bediensystemes, welches störungsfrei als $M|G|1$ -Bediensystem abgebildet werden kann, einfach abzuschätzen.

Im Unterschied zu Abschnitt 2.7.3 in dem die Reparaturwerkstatt als Bediensystem abgebildet wurde soll nun eine Maschine betrachtet werden, an der stochastisch Störungen auftreten. Die Maschine wird mit Hilfe der Bedienstation modelliert, die zu bearbeitenden Aufträge und die Störungen werden mit verschiedenen Kundenklassen abgebildet.

Unter der Annahme, daß Störungen jederzeit eintreten können und dabei, falls vorhanden, die Bedienung des soeben in Bedienung befindlichen Kunden unverzüglich unterbrechen, können die Störungen quasi als Kunden oberster Priorität abgebildet werden. Störungen sollen als seltenes Ereignis betrachtet werden, so daß eine Modellierung des Ankunftsprozesses der Störungen als Poisson-Prozess mit exponentialverteilten Zwischenankunftszeiten möglich ist (siehe z.B. Arnold [3], S. 236 ff.). Der Erwartungswert der Zwischenankunftszeit entspricht in diesem Fall dem Erwartungswert der störungsfreien Einsatzzeit MTBF.

Für die Zeitdauer von der Entdeckung der Störung bis zur Behebung der Störung ist das Bediensystem durch die „Bedienung“ der Störung belegt. Die mittleren Ausfallzeitdauern MTTR werden deshalb durch den Erwartungswert der Bedienzeiten der höchstpriorisierten Klasse modelliert. Außerdem läßt das verwendete $M|G|1$ -Modell zu, die Variabilität der Ausfallzeiten miteinzubeziehen.

Die Fertigungsaufträge werden, wie bereits in Abschnitt 2.9.2 gezeigt, als Kunden der Klasse 2 abgebildet.

Werden zu dem in Abschnitt 2.9.1 eingeführten Beispiel Störungen in der Modellierung als höchstpriorisierte Klasse hinzugefügt, so ist die Vergrößerung des Auslastungsgrades einfach nachzuvollziehen.

Die MTBF sei 200 Stunden, die MTTR betrage 0,5 Stunden, damit erhöht sich die Auslastung des Bediensystems von 0,9333 auf 0,9358.

Eine Erhöhung der Auslastung im Modell ohne Störungen, z.B. durch eine Erhöhung der Ankünfte für die stetige Richtung um 1,5 FE/h, also von 240 FE/h auf 241,5 FE/h, hätte nur geringfügige Folgen, die Wartezeit erhöht sich von 66 Sekunden auf 68,7 Sekunden. Die mittlere Bedienzeit beträgt in diesem Fall 8,4 Sekunden, so daß sich eine Verweilzeit von $t_v = 77,1$ Sekunden ergibt.

Führen wir die Betrachtung mit Störungen weiter und nehmen eine deterministische Stördauer an, so ergibt sich durch Einsetzen in Formel 2.142 eine Wartezeit für Klasse 1 von 2,3 Sekunden, d.h. Störungen warten auf die Beseitigung von Störungen. Die Wartezeit für Kunden der Klasse 2, also für die eigentlichen Fördereinheiten kann wie zuvor mit Hilfe von Ausdruck 2.139 bestimmt werden. Dabei ist jedoch zu beachten, daß die Störungen bei der Ermittlung des zweiten Momentes mit zu berücksichtigen sind. Es resultiert ein höherer Wert von $E(t_{ab}^2) = 119,7$ Sekunden² und eine Variabilität von 0,68. Damit steigt die mittlere Wartezeit der Fördereinheiten $t_{w,2}$ auf 103,9 Sekunden.

Durch die Kombination von Priorität und erhöhter Variabilität, ausgedrückt im zweiten Moment, steigt die mittlere Wartezeit wesentlich stärker, als durch den Effekt der Auslastungserhöhung.

Die mittlere Verweilzeit der Fördereinheiten erhöht sich darüber hinaus in Folge der Bedienungsunterbrechungen während der Ausfallzeit. Mit Hilfe von 2.144 ist die mittlere Verweilzeit der Fördereinheiten als $t_{v,2} = 112,3$ Sekunden berechenbar. Der Expansionseffekt (siehe Seite 54) durch Unterbrechungen ist in diesem Fall durch den geringen Ausfallzeitanteil von 0,25% fast zu vernachlässigen.

Das Beispiel zeigt, daß auch geringe Ausfallzeitanteile durch die mit den Ausfällen üblicherweise verbundene Vergrößerung der Variabilität des Bedienprozesses zu erheblich höheren Warte- und Verweilzeiten führen können. So lange der Auslastungsgrad durch Störungen gering bleibt ist die Wirkung des Expansionsfaktors vergleichsweise gering.

Bei der Modellierung von Störungen durch das Modell $M|G|1$ mit unterbrechen-

den Prioritäten wird ein Warteprozess auch für Störungen mitberücksichtigt, d.h. auch eine bereits gestörte Einrichtung kann weiteren Störungen unterliegen. Dies kann bei entsprechend groß angelegtem Betrachtungszeitraum durchaus der Fall sein, bei der Anwendung des Modells sollte zur Kontrolle die entsprechende Wartezeit $t_{w,1}$ bestimmt und auf ihre Realitätsnähe hin kritisch beurteilt werden.

Sollte diese Abbildung aufgrund hoher Ausfallzeitanteile zu ungenau sein, so kann zum Beispiel bei King (siehe [39], S. 81ff.) eine exakte Formel für die Anzahl der Elemente in der Warteschlange gefunden werden.

2.11 Approximationen für $M|G|m$ -Bediensysteme

Insbesondere für die Abbildung von Werkstattfertigungen, in denen Gruppen gleichartiger Maschinen vorhanden sind, wäre es notwendig, Bediensysteme mit mehreren parallelen Bedienstationen abbilden zu können.

Es ist jedoch heute nicht möglich, mit den für das $M|G|1$ -Bediensystem angewendeten Methoden, Formeln für exakte Kennwerte des $M|G|m$ -Bediensystems zu entwickeln, da die Belegung der Bedienstationen abhängig von der Ankunftsreihenfolge ist.

Deshalb werden zur Abbildung von Bediensystemen mit mehreren Bedienstationen Näherungsverfahren eingesetzt. Eine einfache Approximation beruht auf einem Analogieschluß, bei dem zunächst die Unterschiede zwischen einem $M|M|1$ - und einem $M|M|m$ -Bediensystem analysiert werden.

Diese Unterschiede werden genutzt, um einen Analogieschluß von einem $M|G|1$ -System auf ein $M|G|m$ -System durchzuführen.

Aus Formel (2.33) und durch Anwendung von Littles Gesetz folgt für das $M|M|1$ -Bediensystem:

$$\begin{aligned} N_w &= \frac{\rho}{1-\rho} - \rho \\ &= \frac{\rho^2}{1-\rho} \end{aligned} \tag{2.145}$$

$$\begin{aligned} \Rightarrow t_w &= \frac{N_w}{\lambda} \\ &= \frac{\rho^2}{(1-\rho)\lambda} \\ &= \frac{\rho E(t_{ab})}{1-\rho} \end{aligned} \tag{2.146}$$

Dieses Ergebnis kann mit (2.99) verglichen

$$t_w^{M|G|1} = t_w^{M|M|1} \cdot \frac{1 + c_{ab}^2}{2} \quad (2.147)$$

und auf $M|G|m$ -Bediensysteme übertragen werden.

$$t_w^{M|G|m} \approx t_w^{M|M|m} \cdot \frac{1 + c_{ab}^2}{2} \quad (2.148)$$

Für die Ermittlung von $t_w^{M|M|m}$ kann Ausdruck (2.38) und Littles Gesetz eingesetzt werden.

2.12 Näherungen und Grenzen für $G|G|1$ -FCFS Bediensysteme

Die im vorigen Abschnitt vorgenommenen Betrachtungen profitieren von der Markov-Eigenschaft des Ankunftsprozesses. Da diese Eigenschaft nicht immer gegeben ist, werden die bisher betrachteten Modelle in diesem Abschnitt um Ankunftsprozesse erweitert, die ebenfalls durch Verteilungen beschrieben werden, von denen nur Mittelwert und Varianz, bzw. daraus abgeleitete Größen, bekannt sind.

Für $G|G|1$ -Bediensysteme sind nur mit größerem Aufwand exakte Werte für die t_w oder N_w zu berechnen (siehe z.B. Grassmann und Jain in [38] und [29]). Es ist jedoch möglich, Obergrenzen und Näherungen abzuleiten, die alleine auf der Basis der ersten beiden Momente eine einfache Abschätzung von Kennwerten erlauben. Mit Hilfe der in [66], S. 474ff dargestellten Vorgehensweise können handhabbare Formeln hergeleitet und einige Einsichten gewonnen werden.

Zur Untersuchung von Bediensystemen mit generell und unabhängig voneinander verteilten Ankunfts- und Bedienprozessen soll für jeden Kunden die von ihm benötigte Bedienzeit mit der Zeitlücke bis zum Eintreffen des nächsten Kunden verglichen werden. Daraus können die Auswirkungen auf den Arbeitsvorrat im System, d.h. die Menge noch zu leistender Bedienzeit im System, bestimmt werden.

Ist die Zeitlücke größer als die Bedienzeit des eingetroffenen Kunden, so verringert sich der Arbeitsvorrat im System, ist sie kleiner, so nimmt der Arbeitsvorrat im System zu. Es muß weiterhin beachtet werden, daß Brachzeiten entstehen, der Arbeitsvorrat in einigen Intervallen also 0 ist.

Es seien (siehe auch 2.23):

$K_0, K_1, K_2, \dots, K_n, \dots$	Kunden
$T_{an,0}, T_{an,1}, T_{an,2}, \dots, T_{an,n}, \dots$	die zugehörigen Zeitlücken bis zum Eintreffen des nächsten Kunden
$T_{ab,0}, T_{ab,1}, T_{ab,2}, \dots, T_{ab,n}, \dots$	die Bedienzeit der jeweiligen Kunden. Außerdem sollen $T_{an,n}$ und $T_{ab,n}$ unabhängig verteilte Zufallsvariablen sein.
$X_n = T_{ab,n} - T_{an,n}$	die Differenz zwischen hinzukommender und abfließender Arbeitszeit
$T_{w,1}, T_{w,2}, \dots, T_{w,n}, \dots,$	die Wartezeit der Kunden $K_1, K_2,$
$Y_n = \min\{(T_{w,n} + T_{ab,n} - T_{an,n}), 0\} $	die Brachzeit, die nach dem n -ten Kunden auftritt.

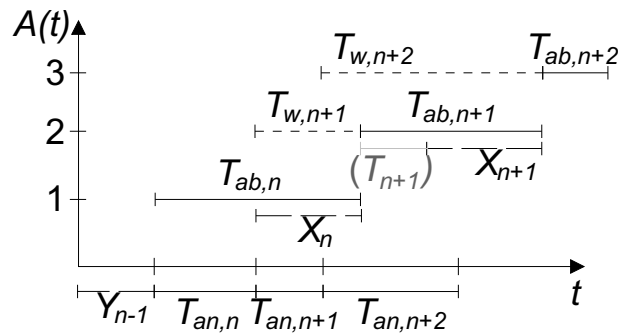


Abbildung 2.23: Darstellung von $T_{an,n}$, $T_{ab,n}$, X_n , T_w und Y_n für ein $G|G|1$ -Bediensystem

Die Brachzeit Y_n nach der Bedienung des n -ten Kunden ist größer als 0, falls die Verweilzeit des n -ten Kunden im Bediensystem kleiner ist als die Zeitlücke bis zur Ankunft des nächsten Kunden K_{n+1} . In einem solchen Fall ist

$$T_{w,n} + T_{ab,n} - T_{an,n} < 0$$

und der Betrag dieses Ausdruckes gibt die Dauer der Brachzeit an. Ist die Zeitlücke kleiner, so entsteht keine Brachzeit und Y_n erhält den Wert 0.

Ersetzt man $T_{ab,n} - T_{an,n}$ durch X_n , so resultiert als kürzerer Ausdruck:

$$Y_n = |\min\{(T_{w,n} + T_{ab,n} - T_{an,n}), 0\}| = |\min\{(T_{w,n} + X_n), 0\}| \quad (2.149)$$

Die Wartezeit des Kunden K_{n+1} kann aus der Wartezeit des Vorgängers $T_{w,n}$ und X_n abgeleitet werden. Im Vergleich zur Wartezeit des Vorgängers verlängert sie sich um die Differenz von zufließender und hinzukommender Arbeit, kann jedoch nicht kleiner als 0 werden.

$$T_{w,n+1} = \max\{(T_{w,n} + X_n), 0\} \quad (2.150)$$

Durch Kombination der Ausdrücke (2.149) und (2.150) folgt nachstehender Ausdruck.

$$T_{w,n+1} - Y_n = T_{w,n} + X_n \quad (2.151)$$

Es greift entweder Ausdruck (2.149) oder Ausdruck (2.150). Nur im Fall $Y_n = 0$ treffen beide zu, dann ist das Ergebnis jedoch identisch.

Es können nicht gleichzeitig Warte- und Brachzeiten auftreten. Deshalb gilt für das Produkt:

$$Y_n \cdot T_{w,n+1} = 0 \quad (2.152)$$

Man kann zeigen, daß die Verteilung von Y_n punktweise gegen Y_∞ konvergiert. Außerdem konvergieren die beiden Zufallsvariablen $T_{w,n}$ und $T_{w,n+1}$ beide gegen den gleichen Erwartungswert $E(T_w)$, da sie aus dem gleichen stochastischen Prozeß stammen.

Durch Bildung der Erwartungswerte in Gleichung (2.151) und unter Ausnutzung von $E(T_{w,n}) = E(T_{w,n+1})$ folgt:

$$E(Y_\infty) = -E(X) \quad (2.153)$$

Der Erwartungswert der Brachzeit kann außerdem auch mit Hilfe der Erwartungswerte für Zwischenankunfts- und Bedienzeit ausgedrückt werden.

$$E(Y_n) = E(T) - E(T_{ab}) = \frac{1}{\lambda} - \frac{1}{\mu} \quad (2.154)$$

Quadriert man Ausdruck 2.151 und geht zu Erwartungswerten über, so kann abermals die Identität von $E(T_{w,n})$ und $E(T_{w,n+1})$ ausgenutzt werden und es entsteht:

$$-2E(T_w \cdot Y_\infty) + E(Y_\infty^2) = 2t_w E(X) + E(X^2) \quad (2.155)$$

Mit Hilfe von 2.152 läßt sich der vorstehende Ausdruck in eine Formel für den gesuchten Erwartungswert der Wartezeit umformen.⁴

$$E(t_w) = \frac{-E(X^2)}{2E(X)} - \frac{E(Y_\infty^2)}{2E(Y_\infty)} \quad (2.156)$$

In obigem Ausdruck sind bekannt: $E(X)$ und $E(Y_\infty)$. $E(X^2)$ kann ebenfalls auf bekannte Parameter zurückgeführt werden,

$$E(X^2) = Var(T_{an}) + 1/\lambda^2 - \frac{2}{\mu\lambda} + Var(T_{ab}) + 1/\mu^2 \quad (2.157)$$

⁴Im Gegensatz zum ersten Eindruck ist der Ausdruck für t_w nicht negativ, da $E(X) < 0$ gilt.

so daß bis auf $E(Y_\infty^2)$ alle Parameter bekannt sind.

Der Erwartungswert der Wartezeit könnte also bestimmt werden, wäre das zweite Moment (oder die Varianz) der Brachzeit bekannt. Die Ermittlung von $E(Y_\infty^2)$ ist jedoch allein mit den ersten beiden Momenten von Bedien- und Zwischenankunftszeit nicht möglich⁵. Das zweite Moment von Y_n kann jedoch mit Hilfe des Mittelwertes $E(Y_\infty)$ und der Varianz $Var(Y_\infty)$ abgeschätzt werden.

$$\begin{aligned} E(Y_\infty^2) &= E^2(Y_\infty) + Var(Y_\infty) \\ \Rightarrow E(Y_\infty^2) &\geq E^2(Y_\infty) \end{aligned} \quad (2.158)$$

Damit kann eine Obergrenze für t_w angegeben werden.

$$\begin{aligned} E(t_w) &\leq \frac{\lambda(Var(T_{an}) + Var(T_{ab}))}{2(1 - \rho)} \\ \Leftrightarrow E(t_w) &\leq \frac{c_{an}^2 + \rho^2 c_{ab}^2}{2\lambda(1 - \rho)} \end{aligned} \quad (2.159)$$

Formel (2.159) ist als „Kingman’s upper bound“ bekannt. Beachtenswert ist, daß c_{an}^2 und c_{ab}^2 nicht in gleichem Maße die Wartezeit beeinflussen. Eine bessere obere Schranke wurde von Wolff in [66], S. 477 mit Hilfe eines Ergebnisses von Daley [13] hergeleitet und lautet:

$$E(t_w) \leq \frac{\lambda(Var(T_{an}) + Var(T_{ab}) - (1 - \rho)^2 Var(T_{an}))}{2(1 - \rho)} \quad (2.160)$$

Diese Schranke ist vor allem bei kleineren und mittleren ρ sowie größeren c_{an}^2 schärfer als „Kingman’s upper bound“.

Unter Beobachtung der Güte der hier vorgestellten oberen Schranken wurden zahlreiche Näherungsformeln entwickelt. Allerdings besitzt keine der Näherungsformeln eine Gütegarantie, d.h. es können beliebig große relative Fehler auftreten. Der Grund hierfür liegt in der Tatsache, daß die mittlere Wartezeit nicht nur von den ersten beiden Momenten, sondern auch von den anderen Momenten beeinflusst wird, die in diesen Formeln keine Berücksichtigung finden.

Marchal [48] gibt als Näherungsformel

$$E(t_w) \approx \frac{1 + c_{ab}^2}{1/\rho^2 + c_{ab}^2} \cdot \frac{\lambda(Var(T_{an}) + Var(T_{ab}))}{2(1 - \rho)} \quad (2.161)$$

an. In dieser Formel ist die Nähe zu Kingmans Obergrenze in Ausdruck (2.159) deutlich zu erkennen.

⁵Für ein $M|G|1$ -Bediensystem wäre dies möglich, da aufgrund der Markov-Eigenschaft des Ankunftsprozesses, die Zeit bis zur Ankunft des nächsten Kunden immer mit den Mittelwerten $1/\lambda$ und der Varianz $1/\lambda^2$ beschrieben werden kann.

Sehr häufig wird in Verfahren, die für $G|G|1$ -Bediensysteme erzielte Ergebnisse verwenden, z.B. für die Kennwertberechnung in Netzwerken, die Formel von Krämer und Langenbach-Belz [42] angewendet. Die größte Quelle von Ungenauigkeiten bei der approximativen Berechnung von Kennwerten für $G|G|1$ -Bediensysteme mit einfachen Ausdrücken, wie z.B. (2.161), liegt in der uniformen Behandlung von Ankunftsprozessen mit Variationskoeffizienten $c_{an}^2 < 1$ und Variationskoeffizienten $c_{an}^2 > 1$. Deshalb wird ein Korrekturterm B eingeführt, der für diese beiden Fälle unterschiedlich bestimmt wird. Es sei:

$$B = \begin{cases} e \left[-\frac{2(1-\rho)}{3\rho} \cdot \frac{(1-c_{an}^2)^2}{c_{an}^2 + c_{ab}^2} \right] & \text{für } c_{an}^2 \leq 1 \\ e \left[-(1-\rho) \frac{c_{an}^2 - 1}{c_{an}^2 + 4c_{ab}^2} \right] & \text{für } c_{an}^2 > 1 \end{cases} \quad (2.162)$$

Mit Hilfe des geeigneten B wird t_w durch

$$E(t_w) \approx \frac{\rho(c_{an}^2 + c_{ab}^2)}{2\mu(1-\rho)} \cdot B \quad (2.163)$$

angenähert. Die Näherungsformel wurde anhand von Simulationsläufen überprüft und wird z.B. von Whitt als Bestandteil des Berechnungspaketes *Queueing Network Analyzer* (siehe [62] und Abschnitt 3.3.2) verwendet, da sie als gute Näherung gilt. Dennoch muß angemerkt werden, daß auch Ausdruck (2.163) zu großen relativen Fehlern führen kann.

Die Verwendung von $G|G|1$ -Bediensystemen zur Modellierung von Komponenten von Materialflußsystemen setzt auch voraus, daß die Verteilung, oder zumindest wichtige Parameter der Abgangsverteilung, bestimmt werden können, da diese unabdingbare Voraussetzung zur Berechnung des Ankunftsstromes am nachfolgenden Bediensystem sind.

Die Zeit $T_{d,n}$ zwischen zwei aufeinanderfolgenden Kunden, die das Bediensystem verlassen, kann durch Y_n und $T_{ab,n}$ ausgedrückt werden.

$$T_{d,n} = Y_n + T_{ab,n} \quad (2.164)$$

Nach Übergang auf die Erwartungswerte folgt:

$$\begin{aligned} E(T_d) &= E(Y_\infty) + E(T_{ab,n}) \\ &= \underbrace{1/\lambda - 1/\mu}_{E(Y_\infty)} + \underbrace{1/\mu}_{E(T_{ab,n})} \\ \Rightarrow E(T_d) &= 1/\lambda \end{aligned} \quad (2.165)$$

Dieses Ergebnis ist nicht überraschend, da der Fluß durch das Bediensystem erhalten bleiben muß. Die Bestimmung des Variationskoeffizienten c_d^2 ist etwas aufwendiger.

$$\begin{aligned}
 c_d^2 &= \frac{\text{Var}(T_d)}{E^2(T_d)} = \lambda^2 [\text{Var}(Y_\infty) + \text{Var}(T_{ab})] \\
 &= \lambda^2 [E(Y_\infty^2) - E^2(Y_\infty) + \text{Var}(T_{ab})] \\
 &= \lambda^2 [E(Y_\infty^2) - (1/\lambda - 1/\mu)^2 + \text{Var}(T_{ab})] \\
 &= \lambda^2 E(Y_\infty^2) - (1 - \rho)^2 + \lambda^2 \text{Var}(T_{ab}) \\
 &= \lambda^2 E(Y_\infty^2) - (1 - \rho)^2 + \rho^2 c_{ab}^2
 \end{aligned} \tag{2.166}$$

Wie auch bei der Bestimmung der Wartezeit sind bis auf $E(Y_\infty^2)$ alle Bestandteile des vorstehenden Ergebnisses bekannt. Deshalb kann auch die Varianz oder das zweite Moment der Zwischenabgangszeit T_d nur näherungsweise bestimmt werden. Hierzu kann zum einen 2.156 mit einem näherungsweise berechneten t_w nach $E(Y_\infty)$ aufgelöst und dieses $E(Y_\infty)$ dann in 2.166 eingesetzt werden. Zum anderen können verfeinerte Approximationsverfahren eingesetzt werden, die speziell für den Einsatz bei der approximativen Berechnung von Netzwerken entwickelt wurden (siehe Abschnitt 3.3).

Mit (2.166) können jedoch die Parameter des Abgangsprozesses für $M|G|1$ -Bediensysteme bestimmt werden, da für diese ein exaktes Ergebnis für t_w berechnet werden kann.

Aus (2.156) folgt:

$$E(Y_\infty^2) = 2E(Y_\infty) \left[\frac{-E(X^2)}{2E(X)} - t_w \right] \tag{2.167}$$

Da die Zwischenankunftszeiten exponentialverteilt sind, kann $\text{Var}(T_{an}) + 1/\lambda^2$ zu $2/\lambda^2$ zusammengefaßt werden. Setzt man die bekannten Parameter in 2.166 ein und vereinfacht, so entsteht ein einfacher Ausdruck für den quadrierten Variationskoeffizienten der Zwischenabgangszeit:

$$c_d^2 = 1 + \rho^2(c_{ab}^2 - 1) \tag{2.168}$$

Für $M|M|1$ -Bediensysteme als Spezialfall der $M|G|1$ -Bediensysteme mit $c_{ab}^2 = 1$ vereinfacht sich (2.168) zu $c_d^2 = 1$. In Abschnitt 3.2 wird dieses Ergebnis vertieft.

Von den zahlreichen entwickelten Näherungsformeln für die mittlere Wartezeit t_w und die Variabilität c_d^2 des Abgangsprozesses werden in Tabelle 2.3 Formeln zusammengestellt, die zu brauchbaren Näherungsergebnissen führen können. Bei einem am IFL durchgeführten Vergleich von Näherungsformeln mit exakten Ergebnissen (bestimmt mit Hilfe zeitdiskreter Modelle, siehe Jain und Grassman [29]) konnten jedoch keine eindeutigen Präferenzen bezüglich der Genauigkeit der Formeln ermittelt werden.

Nr.	Quelle	Mittlere Wartezeit t_w	Variabilität des Abgangsprozesses c_d^2	Bemerkungen
1	[34]	$\rho \frac{c_{an}^2 + c_{ab}^2}{2(\lambda - \mu)}$	mit Hilfe von Näherung für t_w , (2.166) und (2.156)	
2	[9], S. 75	$\frac{\rho^2(1 + c_{ab}^2)}{1 + \rho^2 c_{ab}^2} \cdot \frac{c_{an}^2 + \rho^2 c_{ab}^2}{2(1 - \rho)} \cdot \frac{1}{\lambda}$	$(1 - \rho^2) \frac{\rho^2(1 + c_{ab}^2)}{1 + \rho^2 c_{ab}^2} + \rho^2 c_{ab}^2$	Ungenau für $c_{an}^2 > 2$
3	[9], S. 75	$\frac{\rho(1 + c_{ab}^2)}{2 - \rho + \rho c_{ab}^2} \cdot \frac{\rho(2 - \rho)c_{an}^2 + \rho^2 c_{ab}^2}{2\lambda(1 - \rho)}$	$\frac{1 - \rho^2 + \rho^2 c_{ab}^2 + (c_{an}^2 - 1) \cdot (1 - \rho^2)(2 - \rho) + \rho c_{ab}^2(1 - \rho)^2}{2 - \rho + \rho c_{ab}^2}$	Ungenau für $c_{an}^2 > 2$
4	[9], S. 75	$\frac{\rho c_{an}^2(1 - (1 - \rho)c_{an}^2) + \rho^2 c_{ab}^2}{2(1 - \rho) \cdot \lambda}$	$(1 - \rho)(1 + \rho c_{an}^2)c_{an}^2 + \rho^2 c_{ab}^2$	Ungenau für $c_{an}^2 > 1$
5	[48]	$\frac{1 + c_{ab}^2}{1/\rho^2 + c_{ab}^2} \cdot \frac{\lambda(\text{Var}(T_{an}) + \text{Var}(T_{ab}))}{2(1 - \rho)}$	mit Hilfe von Näherung für t_w , (2.166) und (2.156)	
6	[62]	$\frac{\rho(c_{an}^2 + c_{ab}^2)g}{2\mu(1 - \rho)}$ <p>mit</p> $g = e^{-\left[\frac{2(1-\rho)}{3\rho} \frac{(1-c_{an}^2)^2}{c_{an}^2 + c_{ab}^2}\right]}$ <p>falls $c_{an}^2 < 1$, sonst $g = 1$</p>	$\rho^2 c_{ab}^2 + (1 - \rho)c_{an}^2$	Vorwiegend zur Benutzung in Näherungsverfahren für Bediensystemnetzwerke

 Tabelle 2.3: Näherungsformeln für t_w und c_d^2 bei $G|G|1$ -Bediensystemen

2.13 Approximationen für $G|G|m$ -Bediensysteme

Da es nicht möglich ist, exakte Werte für den Mittelwert der Wartezeit oder der Warteschlangenlänge für $M|G|m$ -Bediensysteme zu berechnen, ist es auch nicht möglich, für $G|G|m$ -Bediensysteme exakten Formeln anzugeben.

Wie schon bei der Entwicklung von Näherungsformeln für $M|G|m$ -Bediensysteme, wird über einen Analogieschluß der Unterschied zwischen $M|M|1$ - und $M|M|m$ -Bediensysteme auf $G|G|1$ - und $G|G|m$ -Systeme übertragen, wobei im Unterschied zu Abschnitt 2.11 nicht auf exakte Ergebnisse, sondern auf Approximationen für die Einkanalssysteme zurückgegriffen werden muß.

$$E(t_w)^{G|G|m} \approx \frac{E(t_w)^{M|M|m}}{E(t_w)^{M|M|1}} E(t_w)^{G|G|1} = \frac{E(t_w)^{G|G|1}}{E(t_w)^{M|M|1}} E(t_w)^{M|M|m} \quad (2.169)$$

Aus Formel 2.38 wird mit Hilfe von Littles Gesetz:

$$E(t_w)^{M|M|m} = \frac{N_w}{\lambda} = p(m) E(t_{ab}) \frac{1}{(1 - \rho)^2} \quad (2.170)$$

Der Wert von $t_w^{G|G|1}$ kann mit verschiedenen Näherungen approximiert werden. Unter Verwendung von 2.163 resultiert ein Quotient $E(t_w)^{G|G|1} / E(t_w)^{M|M|1}$ von:

$$\frac{E(t_w)^{G|G|1}}{E(t_w)^{M|M|1}} \approx \frac{c_{an}^2 + c_{ab}^2}{2} \cdot B \quad (2.171)$$

Damit ist die aus der Formel von Krämer / Langenbach-Belz abgeleitete Näherung für die mittlere Wartezeit am Mehrkanalbediensystem als

$$E(t_w)^{G|G|m} \approx \frac{c_{an}^2 + c_{ab}^2}{2} \cdot B \cdot E(t_w)^{M|M|m} \quad (2.172)$$

auszudrücken, wobei B entsprechend Formel 2.162 zu bestimmen ist.

Whitt schlägt in [62] eine Formel zur Bestimmung der Variabilität der Zwischenabgangszeit vor, die auf Ergebnissen für $G|G|1$ -Bediensystemen beruht.

$$c_d^2 = 1 + (1 - \rho^2)(c_{an}^2 - 1) + \frac{\rho^2}{\sqrt{m}}(c_{ab}^2 - 1) \quad (2.173)$$

Der Ausdruck stimmt in den exakt überprüfbaren Fällen $M|M|m$ - und $M|G|\infty$ -Bediensystem mit den bekannten, genauen Ergebnissen überein.

3 Bediensystemnetzwerke

Bei der Leistungsbewertung von Materialflußsystemen ist es selten sinnvoll, nur Kennwerte für einzelne Elemente zu berechnen. Zur Erzielung sinnvoller Aussagen über die Leistungsfähigkeit des Materialflußsystems muß dem Verflechtungscharakter des Materialflußsystems Rechnung getragen werden.

Die Berechnungsverfahren für Bediensystemnetzwerke wurden in den letzten beiden Jahrzehnten erheblich verbessert, so daß nun bei der Modellierung und Kennwertberechnung für Materialflußsysteme auf wesentlich leistungsfähigere Methoden zurückgegriffen werden kann, als es in den Arbeiten von Gudehus ([34]) und Großschallau ([31]) möglich war.

In diesem Kapitel werden Modelle und Berechnungsverfahren für vernetzten Bediensystemnetzwerke vorgestellt, die grundsätzlich geeignet sind, Materialflußsysteme abzubilden. Auf der Basis der vorgestellten Modelle wurden in der Folge verfeinerte Modelle und Berechnungsverfahren entwickelt, die auf spezifische Problemstellungen ausgerichtet wurden.

Das Verständnis der hier vorgestellten Basisverfahren ist als Basis für diese weiterentwickelte Berechnungsmethoden notwendig, da diese im wesentlichen auf den hier dargestellten Grundlagen beruhen.

Die Berechnungsverfahren unterscheiden sich in ihrer Struktur grundlegend bezüglich des Modellmerkmals Abgeschlossenheit. Deshalb wird im Rahmen dieses Kapitels die weitere Vorgehensweise nach offenen und geschlossenen Netzwerken unterschieden.

Materialflußsysteme können als offene, geschlossene oder auch gemischte Bediensystemnetzwerke abgebildet werden. Dabei können für eine gegebene Problemstellung durchaus zwei verschiedene Modellierungsmöglichkeiten bestehen.

Offene Bediensystemnetzwerke modellieren Materialflußsysteme, bei denen die Kunden an einem oder mehreren Knoten in das Netzwerk eintreten und nach Abschluß aller Bedienschritte das Netzwerk wieder verlassen. Die Anzahl von Kunden im Gesamtsystem unterliegt deshalb einem stochastischen Prozeß und wird durch die Abläufe innerhalb des Netzwerkes bestimmt. Mit solchen Modellen können beispielsweise Werkstattfertigungen abgebildet werden, wobei Kunden die Fertigungsaufträge abbilden, während die

von ihnen beanspruchten Ressourcen (Bearbeitungsmaschinen) durch die Bediensysteme repräsentiert werden.

Geschlossene Bediensystemnetzwerke eignen sich zur Abbildung von Materialflusssystemen, in denen eine feste Zahl von Kunden im Netz umläuft. Die Kunden verlassen das Netz nie, die Gesamtzahl der Kunden ist konstant. Betrachtet man Fahrerlose Transportsysteme, so könnten die Fahrzeuge, die ja stets im System bleiben, durch die Kunden abgebildet werden, die Blockstrecken werden durch die Bediensysteme abgebildet.

Gemischte Bediensystemnetzwerke kombinieren die bereits angeführten Systeme, in dem sich für einige Kundenklassen das Netz als geschlossen darstellt, während die Kunden anderer Kundenklassen das Netz betreten und auch wieder verlassen können. Solche Modelle könnten dann genutzt werden, wenn auf einem Fördersystem sowohl von außen kommende Fördereinheiten gefördert werden und gleichzeitig Systemträger umlaufen. Dies wäre z.B. bei einer Palettenförderanlage der Fall, die für den Transport von palettierten Gütern und gleichzeitig für den Umlauf von belegten und leeren Systemträgern für nicht palettierte Ware genutzt wird.

Bediensystemnetzwerke können aufgrund weiterer Merkmale klassifiziert werden, hierzu gehören unter anderem

- die Zahl verschiedener Kundenklassen, deren Mitglieder jeweils spezielle Wege durch das Bediensystemnetzwerk nehmen,
- die Beschränktheit bzw. Unbeschränktheit der Warteräume und, hier für die weitere Gliederung dieses Kapitels verwendet,
- die Art der Zwischenankunfts- und Bedienzeitverteilungen.

Die Art der Verteilungen ist entscheidend dafür, ob die Berechnung von Kennzahlen exakt oder nur approximativ möglich ist. Dies wird bereits aufgrund der im vorangehenden Kapitel dargestellten Ergebnisse für Einzelbediensysteme deutlich. Diese Ergebnisse werden in den nachfolgenden Abschnitten weiterverwendet. Zunächst sind jedoch die Parameter für die Beschreibung von Bediensystemnetzwerken zu erweitern.

3.1 Beschreibung von Bediensystemnetzwerken

Da nun insgesamt M Bediensysteme miteinander vernetzt sind, werden die für isolierte Bediensysteme bereits eingeführten Beschreibungsparameter mit einem Index z.B. $i = 1, \dots, M$, wie zum Beispiel bei $t_{ab,i}$ versehen, die die Zuordnung

des Parameters zu einem bestimmten Bediensystem i ausdrücken. Bei offenen Bediensystemnetzwerken wird die Netzwerkumgebung durch den Index „0“ repräsentiert.

Der Zustand eines Bediensystemnetzwerkes wird im einfachsten Fall mit Hilfe eines Vektors k mit M Elementen beschrieben, in dem die Elemente k_i die Zahl der Kunden, die sich im Bediensystem $i = 1, \dots, M$ befinden, bezeichnet.

Die Flüsse $\lambda_{i,j}$ zwischen je zwei Bediensystemen werden durch Indexierung mit i und j bezeichnet, die wobei der erste die Quelle und der zweite die Senke des Flusses angibt. Die Gesamtheit λ_j aller an einem Bediensystem j eintreffenden Flüsse entsteht durch Summierung aller eintreffenden Flüsse als

$$\lambda_j = \sum_{i=1}^M \lambda_{i,j} \quad \forall j = 1, \dots, M \quad (3.1)$$

für geschlossene Bediensystemnetzwerke und durch

$$\lambda_j = \lambda_{0,j} + \sum_{i=1}^M \lambda_{i,j} \quad \forall j = 1, \dots, M \quad (3.2)$$

für offene Bediensystemnetzwerke.

Um die Struktur der Flüsse unabhängig von konkreten Ankunftsdaten darstellen zu können, wird die Matrix der Übergangswahrscheinlichkeiten Q eingeführt. Ihre Elemente $q_{i,j}$ geben an, mit welcher Wahrscheinlichkeit im Mittel die Kunden vom Bediensystem i zum Bediensystem j übergehen. Die Matrixelemente $q_{i,j}$ entstehen für geschlossene Bediensysteme also aus:

$$q_{i,j} = \frac{\lambda_{i,j}}{\sum_{j=1}^M \lambda_{i,j}} \quad \forall i = 1, \dots, M \quad (3.3)$$

$$\forall j = 1, \dots, M$$

und für offene Bediensystemnetzwerke entsprechend als:

$$q_{i,j} = \frac{\lambda_{i,j}}{\sum_{j=0}^M \lambda_{i,j}} \quad \forall i = 0, \dots, M \quad (3.4)$$

$$\forall j = 0, \dots, M$$

Die einzelnen Übergangswahrscheinlichkeiten werden zur Matrix der Übergangswahrscheinlichkeiten

$$Q = (q_{i,j}) \quad (3.5)$$

zusammengefaßt. Ausgehend von der Matrix der Übergangswahrscheinlichkeiten Q können Kennwerte für Bediensystemnetzwerke bestimmt werden. Die Vorgehensweise bei offenen und geschlossenen unterscheidet sich, da in offenen Netzwerken der Fluß im Netz von außen vorgegeben wird, woraus ein mittlerer Bestand von Kunden im Netz resultiert, während in einem geschlossenen Netz der Fluß durch die Zahl der im Netzwerk befindlichen Kunden bestimmt wird.

Zunächst sollen offene Bediensystemnetzwerke behandelt werden, wobei zuerst unterstellt sei, daß alle Zwischenankunfts- und Bedienzeiten exponentialverteilt sind. Im nachfolgenden Abschnitt werden die Modelle auf generelle Zeitverteilungen ausgedehnt.

3.2 Offene Bediensystemnetzwerke mit exponential verteilten Zeiten

3.2.1 Die Produktformlösung

Unter der Voraussetzung, daß an einem Bediensystem keine neuen Kunden entstehen, bzw. keine Kunden vernichtet werden, gilt für jedes Bediensystem, daß der Fluß in das Bediensystem hinein genau so groß ist wie der Fluß aus dem Bediensystem hinaus. Langfristig muß deshalb

$$\lambda_j = \lambda_{0,j} + \sum_{i=1}^M \lambda_{i,j} \quad \forall j = 1, \dots, M \quad (3.6)$$

gelten (siehe Formel (3.4)). Mit $\lambda_{i,j} = \lambda_i q_{i,j}$ resultieren M Flußgleichungen der Form:

$$\lambda_j = \lambda_{0,j} + \sum_{i=1}^M \lambda_i q_{i,j} \quad \forall j = 1, \dots, M \quad (3.7)$$

Diese Gleichungen können sukzessive gelöst werden, sofern die Flüsse im Netz keine Schleifen aufweisen, Die Netzwerke sind allerdings selten frei von Schleifen, so zum Beispiel wenn Werkstattfertigungen mit Nacharbeitszyklen modelliert werden müssen.

Aufgrund der in diesen Fällen bestehenden der Abhängigkeiten zwischen den Flüssen im Netz muß in diesem Fall das lineare Gleichungssystem simultan gelöst werden, das durch die Gleichungen 3.7 definiert ist.

Hierzu müssen die Gleichungen (3.7) umgeformt werden. Ausgeschrieben wird (3.7) zu:

- die Ankunft des nächsten Kunden unabhängig von der seit der letzten Ankunft verstrichenen Zeit ist und
- dies auch für die Zeit bis zur Beendigung der Bedienung an einem der Bediensysteme gilt und
- die Übergangswahrscheinlichkeiten $q_{i,j}$ unabhängig vom aktuellen Zustand sind.

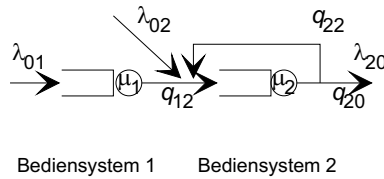


Abbildung 3.1: Beispiel eines offenen Bediensystemnetzwerkes

Für das in Abbildung 3.1 gezeigte Bediensystemnetzwerk ergibt sich nachstehend gezeigter Zustandsraum.

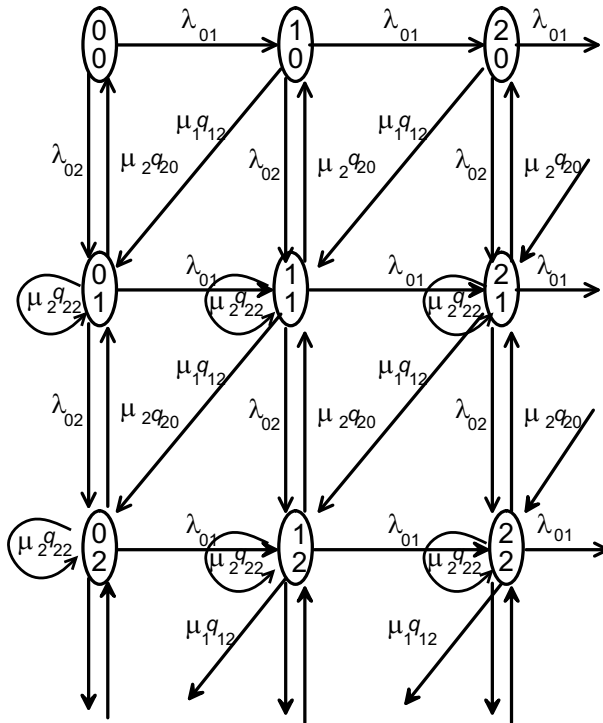


Abbildung 3.2: Zustandsraum eines offenen Bediensystemnetzwerkes

Wie bei Einzelbediensystemen können nun Gleichungen aufgestellt werden, die zu erfüllen sind, wenn sich das gesamte Netzwerk im statistischen Gleichgewicht

befindet. Im Gleichgewicht erreicht ein System einen Zustand genau so oft wie dieser Zustand wieder verlassen wird. Hierfür werden die Ankunfts- und Bedienraten mit den Zustandswahrscheinlichkeiten des Bediensystemnetzwerkes bewertet, um so einen *Wahrscheinlichkeitsfluß* zu erhalten.

Da alle Bewegungen von Kunden in diesen Zustand hinein und hinaus erfaßt werden, nennt man diese Gleichungen die *globalen* Gleichgewichtsgleichungen. Für die Wahrscheinlichkeit $p \binom{1}{1}$ des Zustandes $\binom{1}{1}$ lauten die Gleichungen wie folgt:

$$p \binom{1}{1} (\lambda_{0,1} + \mu_2 q_{2,0} + \lambda_{0,2} + \mu_2 q_{2,2} + \mu_1 q_{1,2}) = \quad (3.11)$$

$$p \binom{0}{1} \lambda_{0,1} + p \binom{1}{0} \lambda_{0,2} + p \binom{2}{0} \mu_1 q_{1,2} + p \binom{1}{2} \mu_2 q_{2,0} + p \binom{1}{1} \mu_2 q_{2,2}$$

Obige Gleichung kann noch vereinfacht werden, indem der auf beiden Seiten vorkommende Term $p \binom{1}{1} \mu_2 q_{2,2}$ gestrichen wird.

Aus den globalen Gleichgewichtsgleichungen sind nur selten direkt numerische Lösungen der Zustandswahrscheinlichkeiten zu gewinnen, da die Zahl der möglichen Zustände und damit die Zahl der aufzustellenden Gleichungen unbegrenzt ist.

Im Unterschied zu den globalen Gleichgewichtsgleichungen stellen die *lokalen* Gleichgewichtsgleichungen nur den Wahrscheinlichkeitsfluß, der durch die Ankunft eines Kunden an *einem bestimmten* Bediensystem entsteht, dem Wahrscheinlichkeitsfluß, der dem Abgang eines Kunden an genau diesem Bediensystem entspricht, gegenüber.

Die lokalen Gleichgewichtsgleichungen sollen zuerst an einem einzelnen $M|M|1$ -Bediensystem (siehe Abschnitt 2.5) erläutert werden. Mit Hilfe der Formeln (2.25) und (2.26) konnten die Zustandswahrscheinlichkeiten $p(k)$ hergeleitet werden. Betrachtet man Gleichung (2.26), so ist festzustellen, daß in dieser Gleichung sowohl die Raten, mit der ein Zustand k aufgrund der Ankunft eines Kunden, als auch die Raten, mit der der Zustand aufgrund des Abgangs eines Kunden nach beendeter Bedienung verlassen wird, enthalten sind.

Für jeden der zwei Schnitte (a) und (b) durch den Zustandsgraphen (siehe Abbildung 3.3) kann eine Gleichgewichtsgleichung aufgestellt werden, die erfüllt sein muß, wenn sich das Bediensystem im Gleichgewicht befindet. Hierzu werden die Wahrscheinlichkeiten der benachbarten Zustände $p(k), p(k-1), p(k+1)$ und die gegebenenfalls zustandsabhängigen Ankunfts- und Bedienraten $\lambda(k), \mu(k), \lambda(k-1)$ und $\mu(k+1)$ verwendet.

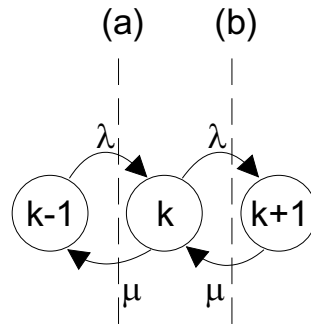


Abbildung 3.3: Graph zur Herleitung der lokalen Gleichgewichtsgleichungen für ein $M|M|1$ -Bedienungssystem

$$(a) \quad p(k)\mu(k) = p(k-1)\lambda(k-1) \quad (3.12)$$

$$(b) \quad p(k)\lambda(k) = p(k+1)\mu(k+1) \quad (3.13)$$

Die erste Gleichung stellt den Wahrscheinlichkeitsfluß über den Schnitt (a) dar, also daß der Zustand k von Zuständen mit einer kleineren Zahl von Kunden $k' < k$ aus genau so oft erreicht wird, wie er durch Abgang eines Kunden den Zustand k erreicht.

Die zweite Gleichung behandelt Schnitt (b), also den Wahrscheinlichkeitsfluß für das Verlassen des Zustands k durch Ankunft eines Kunden und den Wahrscheinlichkeitsfluß für das Erreichen des Zustandes k ausgehend von einem Kunden mehr im System.

Definitionsgemäß werden in einer lokalen Gleichgewichtsgleichung nur die Rate, betrachtet, mit der ein Zustand durch Ankunft eines Kunden erreicht bzw. durch Abgang eines Kunden verlassen wird (Schnitt (a)).

Durch Addition der beiden Gleichungen kann die globale Gleichgewichtsgleichung (2.26) hergestellt werden. Lösungen der Gleichungen (3.13) und (3.12) sind deshalb auch immer Lösungen der Gleichung (2.26). Umgekehrt muß dies allerdings nicht der Fall sein.

Für Bediensystemnetzwerke können nach demselben Muster ebenfalls lokale Gleichgewichtsgleichungen erstellt werden. Hier wird zusätzlich unterschieden, bei welchem Bediensystem Kunden ankommen oder abgehen. So entstehen bei einem Bediensystemnetzwerk mit M Knoten M lokale Gleichgewichtsgleichungen pro Systemzustand. Damit die nachfolgende Darstellung übersichtlicher wird, wird in Abbildung 3.4 und in Ausdruck (3.12) der Fluß $\mu_2 q_{2,2}$ nicht dargestellt, da dieser auf beiden Seiten der Gleichungen als abgehender und ankommender Strom auftreten würde.

Durch Aufstellung der lokalen Gleichgewichtsgleichungen entsteht pro Zustand eine größere Zahl von Gleichungen als bei der Aufstellung der globalen Gleichgewichtsgleichungen. Sie sind jedoch aufgrund ihrer Struktur leichter zu lösen. Au-

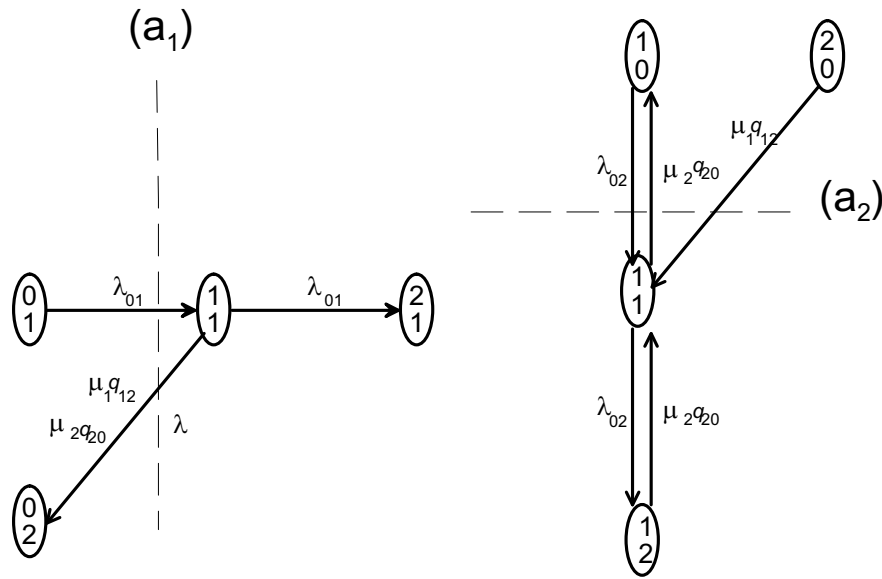


Abbildung 3.4: Graph zur Herleitung der lokalen Gleichgewichtsgleichungen für ein offenes Bediensystemnetzwerk mit zwei Bediensystemen gemäß Abbildung 3.1

ßerdem sind die Eigenschaften der lokalen Gleichgewichtsgleichungen eine Grundlage für effiziente Berechnungsverfahren, die in den nächsten Abschnitten dargestellt werden.

Für den Zustand $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$ können in obigem Beispiel zwei lokale Gleichgewichtsgleichungen aufgestellt werden, jeweils eine für das Bediensystem 1 und eine für das Bediensystem 2. Sie bestehen aus Teilgleichungen, die an jedem der Schnitte entstehen. Sie lauten in diesem Fall (siehe auch Abbildung (3.4)):

Für Bediensystem 1:

$$\underbrace{p \begin{pmatrix} 0 \\ 1 \end{pmatrix}}_{a_1} \lambda_{0,1} = \underbrace{p \begin{pmatrix} 1 \\ 1 \end{pmatrix}}_{a_1} \mu_1 q_{1,2} \quad (3.14)$$

Für Bediensystem 2:

$$\underbrace{p \begin{pmatrix} 1 \\ 0 \end{pmatrix} \lambda_{0,2} + p \begin{pmatrix} 2 \\ 0 \end{pmatrix} \mu_1 q_{1,2}}_{a_2} = \underbrace{p \begin{pmatrix} 1 \\ 1 \end{pmatrix} \mu_2 q_{2,0}}_{a_2} \quad (3.15)$$

Auch hier kann überprüft werden, daß sich aus den lokalen Gleichgewichtsgleichungen die globale Gleichgewichtsgleichung durch Addition aller betroffenen Gleichungen ergibt. Dabei ist zu beachten, daß die Gleichungen der Schnitt (b_1)

und (b_2) erst bei der Betrachtung der Zustände mit einem zusätzlichen Kunden aufgenommen werden.

Wie bereits bei den einzelnen Bediensystemen gilt auch hier, daß die Existenz einer Lösung der lokalen Gleichgewichtsgleichungen nicht gewährleistet ist. Sofern es jedoch eine solche Lösung gibt, ist sie gleichzeitig auch Lösung der globalen Gleichgewichtsgleichungen, denn aus den lokalen Gleichgewichtsgleichungen lassen sich stets die globalen Gleichgewichtsgleichungen gewinnen, indem die zu einem Zustand gehörenden lokalen Gleichgewichtsgleichungen addiert werden.

Bei genauerer Betrachtung der lokalen Gleichgewichtsgleichungen kann festgestellt werden, daß sie den Gleichgewichtsgleichungen eines Einzelbediensystems ähneln. Die Ankunftsrate an dem betrachteten Bediensystem λ_i setzt sich in einem Netz allerdings aus den anteiligen Strömen der von anderen Bediensystemen oder von außen eintreffenden Kunden zusammen. Die die Bedienrate betreffenden Terme bleiben unverändert, sieht man davon ab, daß die Zustandswahrscheinlichkeit der anderen Bediensysteme mitgeführt wird.

Eine auf den lokalen Gleichgewichtsgleichungen beruhende Berechnung der Zustandswahrscheinlichkeiten basiert also letztlich auf einer unabhängigen Berechnung von Zustandswahrscheinlichkeiten für die einzelnen Bediensysteme. Dies ist möglich, da ein Bediensystemnetzwerk mit ausschließlich exponentialverteilten Zwischenankunfts- und Bedienzeiten folgende Eigenschaften aufweist:

1. Der Abgangsprozeß eines $M|M|m$ -Bediensystems ist exponentialverteilt (Siehe Burke [7]).
2. Die zufällige Ausdünnung mit Wahrscheinlichkeit p eines Stromes mit exponentialverteilten Zwischenankunftszeiten und Ankunftsrate λ führt zu einem Strom mit exponentialverteilten Zwischenankunftszeiten und Ankunftsrate $p\lambda$.
3. Die Überlagerung von Strömen mit jeweils exponentialverteilter Zwischenankunftszeit ergibt einen Strom mit exponentialverteilten Zwischenankunftszeiten.

Die Gesamtheit dieser Eigenschaften führt dazu, daß die Kennwerte und insbesondere Zustandswahrscheinlichkeiten der einzelnen Bediensysteme nach der Bestimmung der Ankunftsraten unabhängig voneinander mit Hilfe der Formeln für $M|M|1$ bzw. $M|M|m$ -Bediensysteme berechnet werden können. Das heißt, die Bediensysteme können quasi freigeschnitten und isoliert betrachtet werden. Auf dieser Erkenntnis basieren die sogenannten *Dekompositionsmethoden*, die sowohl zur exakten als auch zur näherungsweise Berechnung von Kennwerten entwickelt wurden.

Gelegentlich genügen bereits isoliert berechnete Kennwerte, um Aufschluß über das Verhalten eines einzelnen, möglicherweise kritischen, Materialflußelementes zu erhalten. Häufig werden jedoch Zustandswahrscheinlichkeiten für das Gesamtsystem benötigt, beispielsweise um gemeinsam genutzte Puffer auszulegen (siehe zum Beispiel Furmans in [23] und [24], sowie fortführend Furmans und Zillus [27]).

Jackson ([37]) hat gezeigt, daß die nachstehende sogenannte *Produktformlösung* eine Lösung der lokalen Gleichgewichtsgleichungen ist.

$$p(k = \begin{pmatrix} k_1 \\ k_2 \\ \vdots \\ k_M \end{pmatrix}) = (1 - \rho_1)\rho_1^{k_1} \cdot (1 - \rho_2)\rho_2^{k_2} \cdots (1 - \rho_M)\rho_M^{k_M} \quad (3.16)$$

Dieses Ergebnis besagt, daß die Zustandswahrscheinlichkeiten für ein Bediensystemnetzwerk berechnet werden können, *als ob* die einzelnen Knoten eines Bediensystemnetzwerkes unabhängig voneinander als $M|M|1$ -Bediensysteme betrachtet werden können und das Ergebnis als Produkt der einzelnen Zustandswahrscheinlichkeiten $p(k_1), p(k_2), \dots, p(k_M)$ der Bediensysteme

$$p(k = \begin{pmatrix} k_1 \\ k_2 \\ \vdots \\ k_M \end{pmatrix}) = p(k_1)p(k_2) \cdots p(k_M) \quad (3.17)$$

bestimmt werden könnte.

Daraus läßt sich jedoch nicht ableiten, daß die Zwischenankunftszeiten an den einzelnen Bediensystemen tatsächlich exponentialverteilt sind. Dies ist, siehe Gross und Harris [32], S. 230ff., nur dann der Fall, wenn der Fluß im Netzwerk keine Schleifen aufweist.

Diese Eigenschaft der Dekomponierbarkeit ist Grundlage vieler Näherungsverfahren, die im Analogieschluß die Berechnung von Kennwerten für ein Netzwerk auf die Kennwerte eines einzelnen freigeschnittenen Bediensystems zurückführen (siehe Abschnitt 3.3).

3.2.2 Beispiel: Analyse eines Bediensystemnetzwerkes

Das Bediensystemnetzwerk aus Abbildung 3.1 soll mit Hilfe der in diesem Abschnitt eingeführten Formeln analysiert werden. Hierzu sind die Parameter wie folgt gegeben:

Die Ankunftsraten betragen $\lambda_{0,1} = 5$, $\lambda_{0,2} = 2$, die zugehörigen Bedienraten sind $\mu_1 = 10$ und $\mu_2 = 10$. Die Übergangswahrscheinlichkeiten seien $q_{1,2} = 1$, $q_{2,0} = 0,9$ und $q_{2,2} = 0,1$.

Die Ankunftsrate am ersten Bediensystem beträgt $\lambda_1 = \lambda_{0,1} = 5$. Der Fluß λ_2 durch das zweite Bediensystem setzt sich zusammen aus dem Abgangsfluß des ersten Bediensystems λ_1 (der auch gleichzeitig mit dessen Ankunftsrate $\lambda_{0,1}$ von extern beschrieben werden kann), den externen Ankünften $\lambda_{0,2}$ und dem rückgeführten Teil des abgehenden Flusses $\lambda_2 q_{2,2}$. Für das zweite Bediensystem gilt deshalb:

$$\lambda_2 = \lambda_{0,1} + \lambda_{0,2} + \lambda_2 q_{2,2}$$

woraus

$$\lambda_2 = \frac{\lambda_{0,1} + \lambda_{0,2}}{1 - q_{2,2}} = \frac{5 + 2}{0,9} = \frac{70}{9}$$

folgt.

Damit können die Auslastungsgrade für Bediensystem 1 als $\rho_1 = \lambda_1/\mu_1 = 1/2$ und Bediensystem 2 als $\rho_2 = \lambda_2/\mu_2 = \frac{70}{9 \cdot 10} = 7/9$ bestimmt werden.

Mit Hilfe des Ausdrucks 2.32 sind nun für jedes Bediensystem die Zustandswahrscheinlichkeiten berechenbar als:

k_i	$p(k_1)$	$p(k_2)$
0	1/2	2/9
1	1/4	14/81
2	1/8	98/729
⋮	⋮	⋮

Nun sind durch Verwendung von Ausdruck (3.17) Zustandswahrscheinlichkeiten für das Netzwerk bestimmbar. So ist beispielsweise die Wahrscheinlichkeit, daß das gesamte Netzwerk leer ist:

$$p(k = \begin{pmatrix} 0 \\ 0 \end{pmatrix}) = p(k_1 = 0) \cdot p(k_2 = 0) = \frac{1}{2} \cdot \frac{2}{9} = \frac{1}{9} \quad (3.18)$$

3.2.3 BCMP-Netzwerke

Die Bedingungen, unter denen exakte Ergebnisse für Bediensystemnetzwerke berechenbar sind, wurden von Baskett, Chandy, Muntz und Palacios in [4] erheblich erweitert. Die Netzwerke, für die diese Ergebnisse gelten, sind deshalb als *BCMP*-Netzwerke bekannt.

Sie zeigten, daß für ein Netz eine Produktformlösung existiert, falls es folgende Eigenschaften besitzt:

- An den Bediensystemen wird nach den Disziplinen FCFS, PS bzw. LCFS abgefertigt oder das Bediensystem verfügt über eine unendliche Zahl von Bedienstationen (sogenannter IS, „Infinite Server“).
- Ist die Bediendisziplin FCFS, so sind als Bedienzeitverteilungen nur Exponentialverteilungen zugelassen, für LCFS, PS oder eine unendliche Zahl von Bediensystemen können Bedienzeitverteilungen eingesetzt werden, die eine ganzrationale Laplace-Transformierte besitzen.
- Die Zwischenankunftszeiten sind
 - entweder exponentialverteilt und alle Kunden stammen aus einer Quelle, von der aus sie mit gegebenen Wahrscheinlichkeiten $p_{0,i}$ auf die Bediensysteme verteilt werden,
 - oder stammen aus mehreren unabhängigen Quellen, die jeweils Kunden mit exponentialverteilten Zwischenankunftszeiten erzeugen.

Eine genauere Darstellung findet sich in Baskett et al. [4] und bei Bolch ([5], Seite 81 ff.). Die Ergebnisse von Baskett et al. gelten auch für die unter Abschnitt 3.4 behandelten geschlossenen Bediensystemnetzwerke.

In Materialflußsystemen sind sehr häufig FCFS Bediendisziplinen anzutreffen, was häufig bereits in technischen Gegebenheiten begründet ist. Außerdem sind bei FCFS Bediendisziplinen lediglich exponentialverteilte Bedienzeiten zur Erzielung exakter Ergebnisse zulässig weswegen sich BCMP-Netzwerke nur selten einsetzen lassen.

Aus diesem Grund wurden zahlreiche Näherungsverfahren entwickelt, am Beispiel des Näherungsverfahrens „Queueing Network Analyzer“ (QNA), das am IFL für die Analyse von Materialflußsystemen eingesetzt wurde, sollen nachfolgend die wichtigsten Elemente eines Näherungsansatzes vorgestellt werden.

3.3 Näherungsverfahren für offene Bediensystemnetzwerke

Die Berechnung von Kennwerten für Bediensystemnetzwerke, die ausschließlich von Exponentialverteilungen charakterisiert werden, beruht, wie unter Abschnitt 3.2.1 gezeigt, auf dem „Freischneiden“ der einzelnen Bediensysteme, nachdem mit Hilfe der Lösung der Flußgleichungen die Ankunftsströme an den Bediensystemen ermittelt wurden. Diese Vorgehensweise ist als Dekompositionsverfahren bekannt und aufgrund der speziellen Eigenschaften der Produktformlösung möglich.

Für Bediensystemnetzwerke, bei denen zumindest eine der M Zwischenankunftsverteilungen oder eine der M Bedienzeitverteilungen nicht die Exponentialverteilungseigenschaft besitzt, existiert keine Produktformllösung. Um dennoch approximativ Kennwerte zu bestimmen, werden die Bediensysteme ebenfalls freigeschnitten. Diese Näherung geht davon aus, daß die Bediensysteme sich nur über die Flüsse $\lambda_{i,j}$ gegenseitig beeinflussen.

Im Fall generell verteilter Zeiten ist neben der Ankunftsrate, die aus der Lösung der Flußgleichungen resultiert, auch die Bestimmung der Varianz (oder der Variationskoeffizienten) der Zeitlücken zwischen eintreffenden bzw. abgehenden Kunden notwendig. Die Flußgleichungen können wie in Abschnitt 3.2 beschrieben gelöst werden. Die Berechnung der Varianzen bzw. Variationskoeffizienten erfordert zusätzliche Betrachtungen. Insbesondere ist zur Charakterisierung der Flüsse innerhalb des Bediensystemnetzwerkes die Bestimmung der Varianz der Abgangsprozesse der Bediensysteme sowie der Varianz, die bei der Aufspaltung von Flüssen oder bei deren Überlagerung entsteht, notwendig.

Die Berechnung von Netzwerken mit generell verteilten Zeiten zerfällt deshalb in die drei Teilaufgaben:

- Berechnung der Abgangsprozesse an einem $G|G|m$ -Bediensystem
- Aufteilung eines Abgangsstromes in mehrere Teilströme, die verschiedene Ziele haben
- Überlagerung mehrerer Ströme zu einem gemeinsamen Ankunftsstrom

Die Berechnung der Charakteristika des Abgangsstromes¹ eines $G|G|m$ -Bediensystems in Abhängigkeit von Zwischenankunfts- und Bedienzeitverteilung kann mit Hilfe der bereits in Abschnitt 2.12 und 2.13 vorgestellten Formeln erfolgen. Weitere Verfeinerungen sind möglich und werden unter Abschnitt 3.3.2 vorgestellt.

Das Hauptaugenmerk dieses Abschnittes liegt deshalb auf den Formeln, die zur Charakterisierung von Strömen dienen, die aufgeteilt und überlagert werden. Hierfür sind prinzipiell zwei Ansätze bekannt:

Kühn [43] nutzt Erneuerungsprozesse und modelliert die Ströme durch eine geeignete Hyper- bzw. Hypoexponentialverteilung mit identischen Mittelwerten und Variationskoeffizienten. Aufbauend auf diesen Verteilungen werden die beschreibenden Verteilungen der resultierenden Ströme bestimmt.

Whitts Ansatz [62] baut unter anderem auf der Arbeit von Kühn auf. Er nähert ebenfalls die Ströme zwischen den Bediensystemen mit Hilfe von Erneuerungsprozessen an und nutzt deren bekannte Eigenschaften, um die ersten beiden Momente

¹Nachfolgend werden die Begriffe „Fluß“ und „Strom“ synonym verwendet. Dies ist darin begründet, daß „Abgangsstrom“ vertrauter als „Abgangsfluß“ klingt, das selbe gilt auch für „Ankunftsströme“ bzw. „-flüsse“

und damit die Kennwerte Variabilität und Mittelwert der resultierenden Ströme zu bestimmen. Eine Bestimmung von Verteilungsformen findet in diesem Ansatz nicht statt.

3.3.1 Exkurs: Erneuerungsprozesse und Punktprozesse

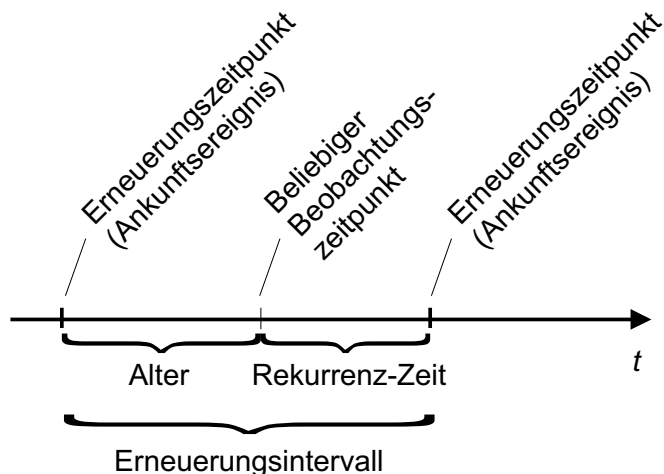


Abbildung 3.5: Wichtige Begriffe zur Beschreibung von Erneuerungsprozessen

Die wichtigste Eigenschaft der bisher verwendeten Markov-Prozesse ist die sogenannte Gedächtnislosigkeit (siehe Abschnitt 2.5). Diese Eigenschaft ist mit der Exponentialverteilung verknüpft. In Anwendungsbereichen, in denen die Exponentialverteilung nicht zur Beschreibung der vorliegenden Verteilungen geeignet ist, kann mit der Markoveigenschaft nur eingeschränkt gearbeitet werden.

Eine Klasse von stochastischen Prozessen mit breiterem Anwendungsbereich sind die sogenannten *Erneuerungsprozesse*. Mit dem Namen „Erneuerungsprozess“ ist die Vorstellung verbunden, daß mit Hilfe des Prozesses beispielsweise der Verbrauch von Glühbirnen beschrieben werden kann. Die Lebenszeit der Glühbirnen kann mit einer für alle Glühbirnen gültigen Verteilung beschrieben werden und sei für alle Glühbirnen auch in den Parametern identisch. Nach jedem Ausfall der Glühbirne wird die Birne „erneuert“ und die Zeit bis zum nächsten Ausfall wird wiederum mit Hilfe der gleichen Verteilung beschrieben.

Charakteristisch für Erneuerungsprozesse ist die Eigenschaft, daß nach jedem Eintritt eines Ereignisses die Zeit bis zum nächsten Ereignis aus der gleichen Verteilung stammt. Wird ein Ankunftsprozeß als Erneuerungsprozeß abgebildet und ist die Verteilung der Zwischenankunftszeiten bekannt, so kann im Moment einer Ankunft die Zeit bis zum nächsten Ankunftsereignis mit Hilfe dieser Verteilung beschrieben werden. Die Zeit zwischen zwei aufeinanderfolgenden Ereignissen,

bezeichnet man als *Erneuerungsintervall*, den Zeitpunkt eines Ereignisses als *Erneuerungszeitpunkt*, da sich hier der stochastische Prozeß „wie neu“ präsentiert (siehe Abbildung 3.5).

Die Verteilung der Zeit von einem beliebigen Zeitpunkt t bis zur nächsten Erneuerung (z.B. Ankunft, Glühbirnenausfall) bezeichnet man als *Rekurrenz-Zeit* oder *Vorwärts-Rekurrenzzeit*. Sie gehorcht bei einem Erneuerungsprozeß im Unterschied zu einem Markov-Prozeß *nicht* der gleichen Verteilung wie die Zeit zwischen zwei Ankünften.

Die Zeit, die seit der letzten Erneuerung vergangen ist, nennt man konsequenterweise *Rückwärts-Rekurrenzzeit* bzw. das *Alter* des Prozesses .

Punktprozesse sind eine Generalisierung der Erneuerungsprozesse, die nicht mehr die Unabhängigkeit der Intervallängen zwischen aufeinanderfolgenden Ankunftsereignissen voraussetzt. Ankunftsströme, die sich aus der Überlagerung von Ankünften verschiedener Quellen zusammensetzen sind im allgemeinen nur als Punktprozesse zu beschreiben. Nur die Überlagerung von Poisson-Prozessen resultiert in einem Erneuerungsprozeß, der wiederum auch ein Poisson-Prozeß ist.

3.3.2 „QNA“, das Berechnungsverfahren von Whitt

QNA steht für *Queueing Network Analyzer*, also Analyseinstrument für Bediensystemnetzwerke und wurde von Ward Whitt für AT&T entwickelt, um Telekommunikationsnetzwerke und Fertigungssysteme zu modellieren.

Whitt hat mit QNA seine Arbeiten im Zusammenhang mit der approximativen Berechnung von Bediensystemnetzwerken in [62] und [61] zusammengefaßt.

Da die Berechnung von Kennwerten für Punktprozesse grundsätzlich möglich, in der Praxis aber aufgrund der Schwierigkeiten, die notwendigen Informationen zu beschaffen (siehe auch Whitt in [60]) nicht durchführbar ist, wird von Whitt vorgeschlagen, Punktprozesse mit Erneuerungsprozessen zu approximieren.

Die Ankunfts- und Bedienprozesse werden über die Ankunfts- bzw. Bedienraten und die Variabilitäten der Zwischenankunfts- und Bedienzeiten beschrieben. Zusätzlich zu den bereits in Abschnitt 3.2 ermittelten Flüssen $\lambda_{i,j}$ zwischen den Bediensystemen und Ankunftsrate λ_i der Bediensysteme müssen deshalb die Variabilitäten $c_{an,i}^2$ der Ankunfts- und $c_{d,i}^2$ der Abgangsströme sowie als Hilfsgröße die Variabilitäten der Ströme $c_{i,j}^2$ zwischen den Bediensystemen bestimmt werden (siehe auch Abbildung 3.6).

Aufteilung von Strömen

Hat ein Bediensystem i mehrere Nachfolger j , zu denen die Kunden nach Beendigung ihrer Bedienung bei i mit Wahrscheinlichkeit $q_{i,j} < 1$ übergehen, so

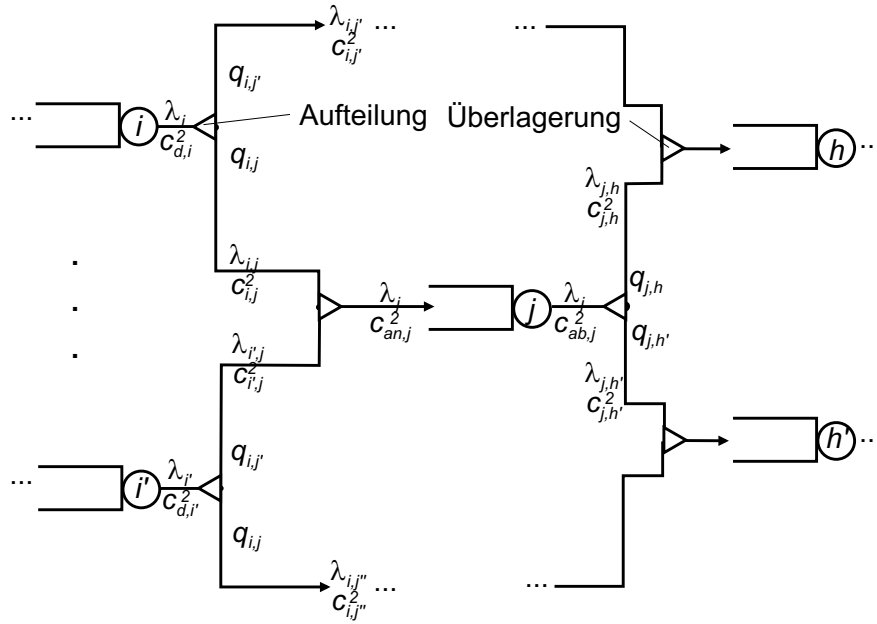


Abbildung 3.6: Ankunfts- und Abgangsraten λ_i der Bediensysteme, Flüsse $\lambda_{i,j}$ zwischen den Bediensystemen und zugehörige Variabilitäten $c_{an,i}^2$ des Ankunftsstromes, $c_{d,i}^2$ des Abgangsstromes und $c_{i,j}^2$ der Ströme zwischen den Bediensystemen in einem Netzwerk mit generellen Verteilungen

wird ein Kunde aus dem Abgangsstrom von Bediensystem i nur mit der Wahrscheinlichkeit $p_1 = q_{i,j}$ in den Ankunftsstrom von j übernommen. Die Wahrscheinlichkeit p_2 daß genau der übernächste Kunde übernommen wird beträgt $p_2 = (1 - q_{i,j}) \cdot q_{i,j}$. Für den dritten Kunden ist sie $p_3 = (1 - q_{i,j})^2 \cdot q_{i,j}$ und für den n -ten $p_n = (1 - q_{i,j})^{n-1} \cdot q_{i,j}$. Zur Ermittlung der Zwischenankunftszeit im Strom von i nach j sind deshalb mit einer Wahrscheinlichkeit p_n n Zwischenabgangszeitintervalle zu addieren.

Die aus diesem Prozeß entstehende Verteilung ist die Verteilung einer Summe aus einer zufälligen Zahl (Anzahl von Abgängen) von Zufallsvariablen (Zeit der Zwischenabgangszeiten).

Zur Bestimmung der Variabilität im Fluß von i nach j basiert auf der Variabilität der Zwischenabgangszeiten bei i und der Verteilung der Zahl der Abgänge bei i , die zwischen zwei Kunden liegen, die aufeinanderfolgend in Richtung j verzweigt werden. Die mittlere Zahl der Abgänge bei i , die zwischen zwei Kunden stattfinden, die in Richtung j verzweigt werden ist mit $N_{i,j}$ bezeichnet. Sie kann aus

$$\begin{aligned}
 E(N_{i,j}) &= \sum_{n=1}^{\infty} n \cdot q_{i,j} \cdot (1 - q_{i,j})^{(n-1)} \\
 &= \frac{1}{1 - q_{i,j}}
 \end{aligned} \tag{3.19}$$

bestimmt werden. Zur Bestimmung der Varianz wird das leichter zu berechnende zweite Moment der Zeitlücken im Strom von i nach j verwendet. Analog zu (3.19) ist das zweite Moment:

$$\begin{aligned}
 E(N_{i,j}^2) &= \sum_{n=1}^{\infty} n^2 \cdot q_{i,j} \cdot (1 - q_{i,j})^{(n-1)} \\
 &= \frac{2 - (1 - q_{i,j})}{(1 - q_{i,j})^2}
 \end{aligned} \tag{3.20}$$

$$\begin{aligned}
 \Leftrightarrow \text{Var}(N_{i,j}) &= E(N_{i,j}^2) - E(N_{i,j})^2 \\
 &= \frac{1 - q_{i,j}}{q_{i,j}^2}
 \end{aligned} \tag{3.21}$$

Der Mittelwert der Zwischenabgangszeit an Bediensystem i sei $E(t_{d,i})$, die Varianz der Zeitlücken des Abgangsstromes kann dann aus der bekannten Variabilität des Abgangsstromes $c_{d,i}^2$ (zu berechnen mit Hilfe von Ausdrücken wie in Tabelle 2.3) von Bediensystem i und dem Mittelwert der Zwischenabgangszeit $E(t_{d,i})^2$ zu

$$\text{Var}(t_{d,i}) = c_{d,i}^2 \cdot E(t_{d,i})^2 \tag{3.22}$$

bestimmt werden.

Für Mittelwert und Varianz einer derart zusammengesetzten Verteilung gelten nach Feller [21], S. 301:

$$\begin{aligned}
 E(t_{i,j}) &= E(N_{i,j}) \cdot E(t_{d,i}) \\
 &= \frac{E(t_{d,i})}{q_{i,j}}
 \end{aligned} \tag{3.23}$$

$$\begin{aligned}
 \text{Var}(t_{i,j}) &= E(N_{i,j}) \cdot \text{Var}(t_{d,i}) + \text{Var}(N_{i,j}) \cdot E(t_{d,i})^2 \\
 &= \frac{E(t_{d,i})^2 \cdot (c_{d,i}^2 \cdot q_{i,j} + 1 - q_{i,j})}{q_{i,j}^2}
 \end{aligned} \tag{3.24}$$

Die gesuchte Variabilität $c_{i,j}^2$ der Zeitlückenverteilung im Strom von i nach j entsteht durch nachfolgende Normierung mit Hilfe des quadrierten Erwartungswertes.

$$c_{i,j}^2 = \frac{\text{Var}(t_{i,j})}{E(t_{i,j})^2} = c_{d,i}^2 \cdot q_{i,j} + 1 - q_{i,j} \tag{3.25}$$

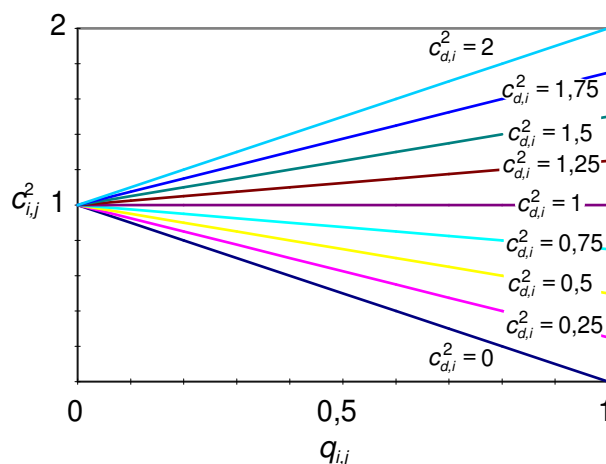


Abbildung 3.7: Variabilität $c_{i,j}^2$ des Stromes von i nach j in Abhängigkeit von der Verzweigungswahrscheinlichkeit $q_{i,j}$ für verschiedene Werte der Abgangsvariabilität $c_{d,i}^2$

Trägt man wie in Abbildung 3.7 die Variabilität $c_{i,j}^2$ des aus einer Verzweigung abgehenden Stromes über der Verzweigungswahrscheinlichkeit $q_{i,j}$ in Richtung j für verschiedene Werte $c_{d,i}^2$ der Variabilität des Abgangsstromes des Bediensystems i auf, so ist zu erkennen, daß für kleine Verzweigungswahrscheinlichkeiten die Variabilität des abgehenden Stromes gegen 1 strebt, unabhängig davon, welchen Wert $c_{d,i}^2$ besitzt. Es kann gezeigt werden, daß der Abgangsprozeß nicht nur die Variabilität 1 besitzt, sondern tatsächlich für kleine Werte von $q_{i,j}$ und für zustandsunabhängige Verzweigungswahrscheinlichkeiten gegen einen Markovprozeß strebt.

Strebt die Verzweigungswahrscheinlichkeit $q_{i,j}$ gegen den Wert 1, geht die Variabilität im Strom von i nach j gegen die Variabilität des bei dem Bediensystem i abgehenden Stromes $c_{d,i}^2$.

Überlagerung von Strömen

Ausgangspunkt sind Kundenströme von mehreren Quellen $i = 1, \dots, M$ zu einem gemeinsamen Ziel j . Die Ströme seien durch ihre Ankunftsrate $\lambda_{i,j}$ und die zugehörigen Variabilitäten $c_{i,j}^2$ der Zwischenankunftszeiten $t_{i,j}$ des Stromes von i nach j beschrieben, die mit Hilfe der im vorigen Abschnitt beschriebenen Verfahren bestimmt werden können.

Gesucht ist die Rate des zusammengeführten Ankunftsstromes λ_j sowie die Variabilität der Zwischenankunftszeiten an Bediensystem j , $c_{an,j}^2$.

Die Ankunftsrate λ_j ergibt sich durch einfache Addition der Ankunftsrate und

kann durch Lösung der Flußgleichungen bestimmt werden, falls das Bediensystemnetzwerk keine eindeutige Flußrichtung besitzt.

$$\lambda_j = \sum_{i=0}^M \lambda_{i,j} \quad \forall j = 1, \dots, M \quad (3.26)$$

Die Bestimmung der Variabilität oder gleichwertiger Maße ist schwieriger. Im Gegensatz zu der exakt möglichen Berechnung der Variabilität aufgeteilter Ströme, ist die Bestimmung dieses Kennwertes für die Überlagerung von Ankunftsströmen nur näherungsweise möglich, da aus der Überlagerung von Erneuerungsprozessen nur im Sonderfall der Überlagerung von Markovprozessen ein Erneuerungsprozeß resultiert. Im allgemeinen wird jedoch ein Punktprozeß erzeugt, der nachfolgend mit Hilfe von Erneuerungsprozessen näherungsweise dargestellt werden soll.

Whitt schlägt eine Kombination von zwei Approximationsmethoden vor, der sogenannten stationären Intervallmethode² und der asymptotischen Methode³. In der Veröffentlichung [60] hat Whitt die Basisverfahren dargestellt und Albin führt in [2] die beiden Vorgehensweisen zu einem hybriden Approximationsverfahren zusammen.

Der Grundgedanke der stationären Intervallmethode besteht darin, die Zeit bis zum nächsten Eintreffen eines Kunden zu betrachten während die asymptotische Approximationsmethode unendlich viele Ankünfte betrachtet.

Die asymptotische Methode bestimmt die Variabilität des überlagerten Stromes als mit der relativen Ankunftsrate gewichtete Summe der einzelnen Ankunftsströme.

$$c_{an,j}^2 = \sum_{i=1}^M \frac{\lambda_{i,j} \cdot c_{i,j}^2}{\lambda_j} \quad (3.27)$$

Die stationäre Intervallmethode führt zu keinem einfachen Ausdruck, da abhängig von der Variabilität der Ströme verschiedene Approximationen mit Hilfe von verschobenen Exponentialverteilungen und zweiphasigen Hyperexponentialverteilungen vorgenommen werden.

Da keine der beiden Approximationen für sich alleine befriedigende Ergebnisse liefert, wurde von Albin ein hybrides Verfahren vorgeschlagen, das mit Hilfe von Simulationsläufen parametrisiert wurde.

Wenn c_a^2 die Variabilität ist, die mit Hilfe der asymptotischen Methode bestimmt wurde und die Variable c_s^2 diejenige beschreibt, die aus der stationären Intervallmethode resultiert, so wird unter Zuhilfenahme der Hilfsvariablen w_j als Kombination

²Im Original Stationary Interval Method

³Im Original Asymptotic Method

$$c_{an,j}^2 = w_j c_a^2 + (1 - w_j) c_s^2 \quad (3.28)$$

vorgeschlagen. Zahlreiche Simulationen zeigten, daß es möglich ist, $c_s^2 = 1$ zu setzen. Die Genauigkeit verringert sich dabei nur unerheblich, im Gegenzug fällt jedoch die aufwendige Ermittlung von c_s^2 fort.

$$c_{an,j}^2 = w_j c_a^2 + (1 - w_j) = w_j \sum_{i=1}^M \left[\frac{\lambda_{i,j} \cdot c_{i,j}^2}{\lambda_j} \right] + 1 - w_j \quad (3.29)$$

Es muß nun noch der Gewichtungsfaktor w_j bestimmt werden. Albin schlägt vor, den Gewichtungsfaktor w_j von den Ankunftsdaten $\lambda_{i,j}$ und der Auslastung des nachfolgenden Bediensystems j abhängig zu machen. Hierfür wird zuerst die Hilfsvariable v_j als

$$v_j = \frac{1}{\sum_{i=1}^M \left[\frac{\lambda_{i,j}}{\lambda_j} \right]^2} \quad (3.30)$$

bestimmt. Damit wird w_j als

$$w_j = \frac{1}{1 + 4(1 - \rho_j)^2 (v_j - 1)} \quad (3.31)$$

festgelegt.

Der Hilfsparameter v_j ist ein Maß für die Zahl der Ankunftsströme, dies stimmt genau dann, wenn alle ankommenden Ströme gleich stark sind.

Die Hilfsparameter w_j und v_j sind Bestandteil des Näherungsverfahrens und könnten durch geeignetere Funktionen ersetzt werden, sobald bessere Funktionen bekannt sind.

Approximation der Abgangsprozesse

Der Kennwert $c_{d,i}$ des Abgangsprozesses kann mit Hilfe der in Tabelle 2.3 angegebenen Näherungsverfahren approximiert werden. Es ist anzustreben Näherungsverfahren zu verwenden, die nicht Berechnungsergebnisse für die Bediensysteme (wie z.B. die Wartezeit) verwenden, um die Kennwertberechnung selber von der Bestimmung der Eingangsparameter zu entkoppeln.

Whitt schlägt eine einfache Linearkombination der Bedienzeit- und Zwischenankunftszeitvariabilitäten vor, die mit dem Auslastungsgrad ρ gewichtet werden (siehe Tabelle 2.3, Zeile 6).

Für Mehrkanalsysteme wird die bereits vorgestellte Formel 2.173 vorgeschlagen. Die Näherungsgüte beim Einsatz in der Kennwertberechnung für Netzwerke kann noch verbessert werden, in dem Kennwerte vor- und nachgelagerter Bediensysteme in die Berechnung miteinbezogen werden. Auch hierfür liefern die asymptotische und die stationäre Intervallmethode die Grundlagen. In [63] wird gezeigt, daß die in Tabelle 2.3 angegebene Approximation das Ergebnis einer auf der stationären Intervallmethode beruhenden Überlegung ist. Die asymptotische Methode liefert als Ergebnis für die Variabilität $c_{d,i}^2$ des Abgangsprozesses nichts anderes als die Variabilität $c_{an,i}^2$ des Ankunftsprozesses.

Der Vergleich der Näherungsformeln mit Simulationsergebnissen zeigte, daß die Reduktion der Variabilität, die durch einen getakteten Bedienprozeß zu erzielen ist, nicht so stark ausfällt, wie Formel 2.173 dies prognostiziert. Um diesen Einfluß zu reduzieren, schlägt Whitt eine Begrenzung des Einflusses getakteter Bedienzeiten vor, so daß sich nachstehender Ausdruck für die Variabilität des Abgangsprozesses ergibt.

$$c_{d,i}^2 = 1 + (1 - \rho_i^2)(c_{an,i}^2 - 1) + \frac{\rho_i^2}{\sqrt{m}}(\max\{c_{ab,i}^2; 0, 2\} - 1) \quad (3.32)$$

Bei hohen Auslastungen $\rho_i \rightarrow 1$ (sogenannter „heavy traffic“) nimmt der Einfluß des Bedienungsprozesses vorgelagerter Bediensysteme ab, der Einfluß der Ankunftsprozesse nimmt dagegen zu. Deshalb soll der Abgangsprozeß nach Bestimmung der durch die Verzweigung erzeugten Variabilitäten noch einmal mit Hilfe der Auslastungsinformation abgestimmt werden, um die Ergebnisse der stationären Intervall-Betrachtung durch die bei hohen Auslastungen bessere asymptotische Methode zu ergänzen.

Ausgehend von dem Ausdruck 3.25 wird unter Einbeziehung des Auslastungsanteils die Variabilität wie folgt erweitert:

$$c_{i,j}^2 = v_{i,j}(q_{i,j}c_{an,i}^2 + 1 - q_{i,j}) + (1 - v_{i,j}) \cdot \left[q_{i,j} \left(1 + (1 - \rho_i^2)(c_{an,i}^2 - 1) + \frac{\rho_i^2}{\sqrt{m_i}}(c_{ab,i}^2 - 1) \right) + 1 - q_{i,j} \right] \quad (3.33)$$

Mit der Korrektur aus Ausdruck 3.32 erhält man stattdessen:

$$c_{i,j}^2 = v_{i,j}(q_{i,j}c_{an,i}^2 + 1 - q_{i,j}) + (1 - v_{i,j}) \cdot \left[q_{i,j} \left(1 + (1 - \rho_i^2)(c_{an,i}^2 - 1) + \frac{\rho_i^2}{\sqrt{m_i}}(\max\{c_{ab,i}^2; 0, 2\} - 1) \right) + 1 - q_{i,j} \right] \quad (3.34)$$

Variabilitätsgleichungen

Die vorangehenden Abschnitte haben die wechselseitige Abhängigkeit der Variabilität von den Eingangsgrößen dargestellt. Aufgrund dieser Abhängigkeiten ist

eine sukzessive Bestimmung der Variabilitäten nur dann möglich, wenn das Netz keine Rückflüsse besitzt.

Falls dies nicht der Fall ist, kann ein Gleichungssystem aufgestellt werden, dessen Lösung die Variabilitäten $c_{an,i}^2$ im Ankunftsstrom der Bediensysteme sind.

Aus der Überlagerung der Ankunftsprozesse (siehe 3.29) resultiert nachstehendes Gleichungssystem

$$c_{an,j}^2 = 1 - w_j + w_j \sum_{i=0}^M q_{i,j} c_{i,j}^2 \quad (3.35)$$

Die $c_{i,j}$ sind hierin noch zu bestimmen, wofür die bei der Bestimmung der Abgangsprozesse und der nachfolgenden Überlagerung angestellten Überlegungen von Bedeutung sind. Die Lösung der Flußgleichungen für gegebene externe Ankunftsdaten muß bereits vorliegen, da die Auslastung ρ_j bei der Bestimmung der Variabilität der Abgangsprozesse von Bedeutung ist. Aus diesem Grunde sind hier die Formeln 3.35 bzw. 3.33 einzusetzen, woraus die umfangreichere Formulierung

$$\begin{aligned} c_{an,j}^2 = & 1 - w_j + w_j \left[\frac{\lambda_{0,j}}{\lambda_j} c_{0,j}^2 + \sum_{i=1}^M \frac{\lambda_{i,j}}{\lambda_j} (v_{i,j} [q_{i,j} c_{an,i}^2 + 1 - q_{i,j}] \right. \\ & + (1 - v_{i,j}) \{ q_{i,j} [1 + (1 - \rho_i^2)(c_{an,i}^2 - 1) \\ & \left. + \frac{\rho_i^2}{\sqrt{m_i}} (\max\{c_{ab,i}^2; 0, 2\} - 1)] + 1 - q_{i,j} \} \right] \end{aligned} \quad (3.36)$$

resultiert.

Die in diesem Gleichungssystem zusammengefaßten Abhängigkeiten lassen sich in einfacherer Form bei geeigneter Umformung darstellen, denn die einzelnen Gleichungen sind lineare Zusammenhänge der Form:

$$c_{an,j}^2 = a_j + \sum_{i=1}^M c_{an,i}^2 b_{i,j} \quad \forall j = 1, \dots, M \quad (3.37)$$

Der Vektor a und die Matrix B sind anhand der bereits aufgeführten Formeln vor der Lösung des Gleichungssystems zu bestimmen. Die Komponenten a_j werden mit Hilfe der bereits bekannten Hilfsvariablen v_j (siehe 3.30) und w_j (siehe 3.31), bestimmt, die zur Parametrierung der Überlagerung von Flüssen dienen.

Die Komponenten des Vektors a bilden im Wesentlichen den Einfluß der Variabilität der externen Ankünfte und den Einfluß der Bedienzeit ab.

$$\begin{aligned} a_j = & 1 + w_j \left\{ \left(\frac{\lambda_{0,j}}{\lambda_j} c_{0,j}^2 - 1 \right) + \right. \\ & \left. \sum_{i=1}^M \frac{\lambda_{i,j}}{\lambda_j} \left[(1 - q_{i,j}) + (1 - v_{i,j}) q_{i,j} \rho_i^2 \left(1 + \frac{1}{\sqrt{m_i}} (\max\{c_{ab,i}^2; 0, 2\} - 1) \right) \right] \right\} \end{aligned} \quad (3.38)$$

Die Matrix B der $b_{i,j}$ beschreibt die Transformation der Variabilitäten der Eingangsströme an den Bediensystemen i in die Variabilitäten der Ausgangsströme einschließlich ihrer Überlagerung.

$$b_{i,j} = w_j \frac{\lambda_{i,j}}{\lambda_j} q_{i,j} (v_{i,j} + (1 - v_{i,j})(1 - \rho_i^2)) \quad (3.39)$$

Die Vorgehensweise der Kennwertberechnung mit Hilfe des „QNA“ ist nachfolgend zusammengefaßt.

Die Berechnungsschritte des „QNA“

1. Löse die Flußgleichungen für ein offenes Bediensystemnetzwerk mit Hilfe von 3.10
2. Bestimme die Auslastungsgrade ρ_i der Bediensysteme
3. Löse das Gleichungssystem der Variabilitätsgleichungen mit Hilfe von Ausdruck 3.36 bis 3.39
4. Berechne Kennwerte für die Bediensysteme mit Hilfe der Ausdrücke aus den Abschnitten 2.12 bzw. 2.13

3.4 Geschlossene Bediensystemnetzwerke mit exponentialverteilten Bedienzeiten

Für Materialflußsysteme, in denen die Zahl K der Kunden im System konstant ist, wie z.B. bei einem flexiblen Fertigungssystem die Zahl der Paletten oder bei einem Fahrerlosen Transportsystem die Zahl der Fahrzeuge, bieten sich geschlossene Bediensystemnetzwerke als Modelle an.

In Abbildung 3.8 ist ein geschlossenes Bediensystemnetzwerk dargestellt, das demjenigen aus Abbildung 3.1 gleicht. Allerdings können im geschlossenen Netzwerk keine Zugänge von außen an den Bediensystemen 1 und 2 stattfinden.

Im Gegensatz zu den offenen Bediensystemnetzwerken ist der Zustandsraum \mathcal{Z} des geschlossenen Netzwerkes begrenzt. Die tatsächliche Größe hängt von der Zahl der insgesamt in dem Netzwerk befindlichen Zahl von Kunden ab. Die Größe Z des Zustandsraumes \mathcal{Z} ist die Zahl der Möglichkeiten, mit der K Kunden auf M Bediensysteme verteilt werden können und kann mit Hilfe des Binomialkoeffizienten

$$Z = \binom{M + K - 1}{M - 1} \quad (3.40)$$

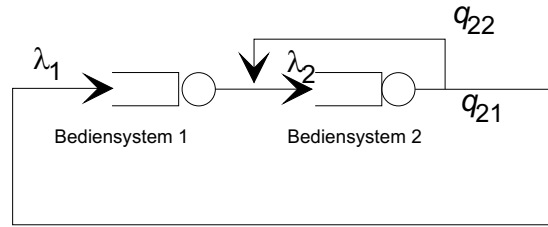


Abbildung 3.8: Beispiel eines geschlossenen Bediensystemnetzwerkes mit $M = 2$ Bediensystemen

berechnet werden.

In Abbildung 3.9 ist der Zustandsraum des Bediensystemnetzwerkes aus Abbildung 3.8 gezeigt, wie er sich im Falle von zwei im Netzwerk befindlichen Kunden darstellt. In diesem Fall sind nur 3 verschiedene Zustände möglich. Ein Vergleich mit Abbildung 3.2 zeigt, daß der nun vorliegende Zustandsraum einen Ausschnitt aus demjenigen für ein offenes Bediensystemnetzwerk darstellt. Allerdings kommt die neue Übergangswahrscheinlichkeit $q_{2,1}$ hinzu.

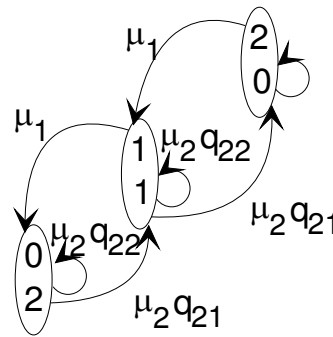


Abbildung 3.9: Zustandsraum des geschlossenen Bediensystemnetzwerkes aus Abbildung 3.8

Deshalb haben die globalen Gleichgewichtsgleichungen auch ein ähnliches Aussehen. Wird wie bei den offenen Bediensystemnetzwerken der Ausdruck $p \begin{pmatrix} 1 \\ 1 \end{pmatrix} \mu_2 q_{2,2}$ auf beiden Seiten der Gleichung gestrichen, so lautet die Gleichung für den Zustand $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$ wie folgt:

$$p \begin{pmatrix} 1 \\ 1 \end{pmatrix} (\mu_1 + \mu_2 q_{2,1}) = p \begin{pmatrix} 2 \\ 0 \end{pmatrix} \mu_1 + p \begin{pmatrix} 0 \\ 2 \end{pmatrix} \mu_2 q_{2,1} \quad (3.41)$$

Ebenso lassen sich lokale Gleichgewichtsbedingungen angeben, die wie für offene Bediensystemnetzwerke abgeleitet werden können.

3.4.1 Produktformlösung für geschlossene Bediensystemnetzwerke

Deshalb ist es nicht erstaunlich, daß es Gordon und Newell in [28] gelang zu zeigen, daß auch für geschlossene Bediensystemnetzwerke Produktformlösungen existieren. Ihre Form ist derjenigen für offene Bediensystemnetzwerke (siehe Ausdruck (3.17)) ähnlich.

$$p(k = \begin{pmatrix} k_1 \\ k_2 \\ \vdots \\ k_M \end{pmatrix}) = \frac{1}{G(K)} \prod_{i=1}^M F_i(k_i) \quad (3.42)$$

Darin gibt K die Anzahl der Kunden im Bediensystemnetzwerk an.

Für den Fall, daß mit Hilfe der Funktion $F_i(k_i)$ die Zustandswahrscheinlichkeiten eines Bediensystems i bestimmt werden, ist keine Normalisierungskonstante $G(K)$ notwendig. Die Einführung der Normalisierungskonstanten $G(K)$ erlaubt jedoch, bei der Durchführung der Berechnung alle in die Gleichung als konstantes Produkt eingehenden Faktoren zu eliminieren und in der Normalisierungskonstante so zu berücksichtigen, daß sich die Summe der Zustandswahrscheinlichkeiten des begrenzten Zustandsraums zu 1 ergänzt.

Die Normalisierungskonstante $G(K)$ ist wie folgt zu bestimmen:

$$G(K) = \sum_{\sum_{i=1}^M k_i = K} \prod_{i=1}^M F_i(k_i) \quad (3.43)$$

Die Terme $F_i(k_i)$ spielen die gleiche Rolle wie die Zustandswahrscheinlichkeiten der Bediensysteme in offenen Bediensystemnetzwerken. Im Unterschied zu offenen Bediensystemnetzwerken können für geschlossene Netze nicht ohne weiteres Ankunftsdaten angegeben werden, da die Ankunftsdaten an den Bediensystemen eines geschlossenen Bediensystemnetzwerkes von den noch zu berechnenden Zustandswahrscheinlichkeiten des Netzes abhängen, die ihrerseits durch die Zahl der Kunden und die Übergangswahrscheinlichkeiten bestimmt werden.

Aus diesem Grunde wird die sogenannte *Besuchshäufigkeit* e_j eingeführt, die den relativen Anteil des Gesamtflusses angibt, der durch das Bediensystem j fließt. e_j ist mit Hilfe der Übergangsmatrix Q zu berechnen.

$$e_j = \sum_{i=1}^M e_i q_{i,j} \quad j = 1, \dots, M \quad (3.44)$$

Das durch Ausdruck (3.44) gegebene Gleichungssystem enthält jedoch nur $M - 1$ unabhängige Gleichungen, da ein geschlossenes Bediensystemnetzwerk vorliegt.

Da die Besuchshäufigkeiten nur in ihrer Relation zueinander von Bedeutung sind, kann eine der Variablen frei bestimmt werden. Eine häufig gewählte Möglichkeit besteht darin, die Summe der e_j auf 1 zu normieren. Als zusätzliche Bedingung wird also die Gleichung

$$\sum_{j=1}^M e_j = 1 \quad (3.45)$$

eingefügt. In diesem Fall geben die e_j den Anteil am Gesamtfluß an, der durch das Bediensystem j fließt.

Nimmt man dagegen an, daß das Bediensystem 1 eine besondere Funktion hat, z.B. als „Zählknoten“ dient, an dem der Gesamtdurchsatz des Netzes gemessen werden soll (es soll also $\lambda = \lambda_1$ gelten), so bietet sich an, $e_1 \equiv 1$ zu setzen. In diesem Fall geben die e_j dann das Verhältnis des Durchsatzes λ_j am Bediensystem j im Verhältnis zum Durchsatz λ_1 am Bediensystem 1 an. Es gilt also:

$$e_j = \frac{\lambda_j}{\lambda_1} \quad (3.46)$$

Diese Annahme soll im weiteren verwendet werden.

Wird als zusätzliche Gleichung

$$e_1 \equiv 1 \quad (3.47)$$

gewählt, und stellt man sich Bediensystem 1 als ein Bediensystem vor, welches bei jedem Durchlauf eines Kunden durch das Netz genau einmal besucht wird, so geben die e_j an, wie oft ein Kunde im Durchschnitt bei einem Durchlauf durch das Netz das Bediensystem j besucht. Mit $\lambda = \lambda_1$ gilt für den Durchsatz λ_j an Knoten j :

$$\lambda_j = \lambda e_j \quad j = 1, \dots, M \quad (3.48)$$

Mit Hilfe der Besuchshäufigkeiten kann bereits ein wichtiger Kennwert für das Bediensystemnetzwerk bestimmt werden, der *Sättigungsdurchsatz*. Der Sättigungsdurchsatz ist der Grenzwert des Netzwerkdurchsatzes bei gegen unendlich strebender Kundenzahl K .

Der Sättigungsdurchsatz wird durch das am höchsten ausgelastete Bediensystem bestimmt, an dem im Sättigungsfall keine Brachzeiten mehr auftreten.

Da das Verhältnis der Durchsätze an den einzelnen Bediensystemen durch die Besuchshäufigkeiten beschrieben wird, und diese in einem festen Verhältnis zueinander stehen, ist eine Durchsatzerhöhung nur proportional zu den Besuchshäufigkeiten möglich. Sobald sich ein Bediensystem dem Auslastungsgrad 1 nähert, erhöhen zusätzlich ins System eingebrachte Kunden nur den Bestand vor der Engpaßstation, führen aber nicht zu wesentlich erhöhtem Durchsatz an diesem und damit an allen Bediensystemen.

Auf der Basis der Besuchshäufigkeiten, der Bedienraten und der Zahl der Bedienstationen an den Bediensystemen kann der Sättigungsdurchsatz einfach bestimmt werden. Besitzen mehrere Bediensysteme die gleiche Leistungsfähigkeit $m_i\mu_i$ (gemessen als Produkt aus Anzahl Bedienstationen und deren Bedienrate) in Relation zu den Besuchshäufigkeiten e_i , so wird der Sättigungsdurchsatz durch diese Bediensysteme gemeinsam beschränkt.

Zur Bestimmung der Engpaßknoten werden also diejenigen Bediensysteme gesucht, die den Gesamtdurchsatz λ_{max} entscheidend beschränken, ihre Indizes seien in der Menge \mathcal{E} enthalten. Gehen wir davon aus, daß am Zählknoten, dem Bediensystem 1, die Besuchshäufigkeit auf 1 gesetzt wurde. An allen anderen Bediensystemen liegt bei maximalem Netzwerkdurchsatz ein Durchsatz von $\lambda_{i,max} = \lambda_{max} \cdot e_i$ vor. Die Engpässe zeichnen sich dadurch aus, daß an ihnen der dem Sättigungsdurchsatz entsprechende Durchsatz $\lambda_{i,max}$ der Summe der Bedienraten aller Bedienstationen entspricht:

$$\lambda_{i,max} = \mu_i \cdot m_i \quad \forall i \in \mathcal{E} \quad (3.49)$$

Für jedes Bediensystem $i = 1, \dots, M$ kann ein hypothetischer Sättigungsdurchsatz $\lambda_{i,max}^{hyp}$ bestimmt werden, der erreicht würde, falls das Bediensystem i zur Menge der Engpaßstationen gehören würde. Der Sättigungsdurchsatz ergibt sich dann als Minimum aller $\lambda_{i,max}^{hyp} = \frac{\mu_i m_i}{e_i}$.

$$\lambda_{max} = \min_{i=1, \dots, M} \frac{\mu_i m_i}{e_i} \quad (3.50)$$

Die Menge der Engpaßstationen \mathcal{E} ist damit wie folgt bestimmt:

$$\mathcal{E} = \left\{ i \mid \lambda_{max} = \frac{\mu_i m_i}{e_i} \right\} \quad (3.51)$$

Zur Berechnung von Durchsätzen unterhalb des Sättigungsdurchsatzes und zur Berechnung des Durchsatzes für eine bestimmte Kundenzahl K sind weitere Betrachtungen notwendig, die ebenfalls auf der Produktformlösung basieren.

Die Produktformlösung für offene Bediensystemnetzwerke beruht auf der Berechnung von Zustandswahrscheinlichkeiten für die einzelnen Bediensysteme, die mit Hilfe der Ausdrücke 2.32 für Einkanalsysteme und 2.35 für Mehrkanalsysteme zu bestimmen sind.

Soll der Ausdruck 2.32 in einem geschlossenen Bediensystemnetzwerk für das Bediensystem i eingesetzt werden, so kann er wie folgt erweitert werden:

$$\begin{aligned}
 p(k_j) &= (1 - \rho_j)\rho_j^{k_j} \\
 &= \left(1 - \frac{\lambda_j}{\mu_j}\right) \left(\frac{\lambda_j}{\mu_j}\right)^{k_j} \\
 &= \left(1 - \lambda \frac{e_j}{\mu_j}\right) \left(\lambda \frac{e_j}{\mu_j}\right)^{k_j}
 \end{aligned} \tag{3.52}$$

Werden vorstehende Ausdrücke in 3.42 eingesetzt, so entsteht:

$$\begin{aligned}
 p(k = \begin{pmatrix} k_1 \\ k_2 \\ \vdots \\ k_M \end{pmatrix}) &= \frac{p(k_1)p(k_2)\cdots p(k_M)}{G(K)} \\
 &= \frac{\left(1 - \lambda \frac{e_1}{\mu_1}\right) \left(\lambda \frac{e_1}{\mu_1}\right)^{k_1} \cdot \left(1 - \lambda \frac{e_2}{\mu_2}\right) \left(\lambda \frac{e_2}{\mu_2}\right)^{k_2} \cdots \left(1 - \lambda \frac{e_M}{\mu_M}\right) \left(\lambda \frac{e_M}{\mu_M}\right)^{k_M}}{G(K)}
 \end{aligned} \tag{3.53}$$

Da $G(K)$ in jedem Falle berechnet werden muß, bietet es sich an, aus Formel 3.53 alle Teile zu entfernen, die für alle Zustände gleich sind, da diese über die Normalisierungskonstante $G(K)$ kompensiert werden können. In Ausdruck 3.53 sind zwei Vereinfachungen möglich:

- Da die Summe aller Kunden im System konstant und gleich $\sum_{i=1}^M k_i = K$ ist, ist auch das Produkt $\prod_{i=1}^M \lambda^{k_i}$ konstant. Das heißt, λ kann aus Ausdruck 3.53 entfernt werden.
- Das Produkt $\left(1 - \lambda \frac{e_1}{\mu_1}\right) \cdot \left(1 - \lambda \frac{e_2}{\mu_2}\right) \cdots \left(1 - \lambda \frac{e_M}{\mu_M}\right)$ ist ebenfalls für alle Zustände konstant und kann weggelassen werden.

Mit Hilfe dieser Überlegungen kann der den Zustand des Bediensystems j beschreibende Bestandteil

$$F_j(k_j) = \left(\frac{e_j}{\mu_j}\right)^{k_j} \tag{3.54}$$

isoliert werden.

Für Mehrkanalbediensysteme können die gleichen Überlegungen angestellt werden, woraus der nachstehende Ausdruck für $F_j(k_j)$ resultiert (siehe [5]).

$$F_j(k_j) = \left(\frac{e_j}{\mu_j}\right)^{k_j} \cdot \frac{1}{b_j(k_j)} \tag{3.55}$$

Der Faktor $b_j(k_j)$ dient der Berücksichtigung der Zahl paralleler Bedienstationen und ist deshalb wie folgt zu bestimmen:

$$b_j(k_j) = \begin{cases} k_j! & \text{für } k_j \leq m_j \\ m_j! m_j^{k_j - m_j} & \text{für } k_j \geq m_j \\ 1 & \text{für } m_j = 1 \end{cases} \quad (3.56)$$

Mit Hilfe der Zustandswahrscheinlichkeiten des Netzwerkes können nun die Zustandswahrscheinlichkeiten $p_j(k_j = n)$ für einen bestimmten Zustand n eines Bediensystems bestimmt werden. Hierzu werden alle Wahrscheinlichkeiten von Netzwerkzuständen addiert, die den gleichen Zustand n des Bediensystems j beinhalten.

$$p_j(k_j = n) = \sum_{\substack{k_j = n \\ \sum_{i=1}^M k_i = K}} p(k) \quad (3.57)$$

Die notwendigen Schritte zur Berechnung der Zustandswahrscheinlichkeiten in einem geschlossenen Bediensystemnetzwerk mit Hilfe der Produktformlösung sind nachstehend zusammengefaßt:

Berechnungsschritte der Produktformlösung für geschlossene Netze

1. Bestimme die Besuchshäufigkeiten e_j mit Hilfe von 3.44
2. Bestimme für jedes Bediensystem $j = 1, \dots, M$ und mögliche Kundenzahl $k_j = 0, \dots, K$ den Faktor $F_j(k_j)$
3. Berechne die Normalisierungskonstante $G(K)$ mit 3.43
4. Berechne die Zustandswahrscheinlichkeiten der Netzwerkzustände mit Hilfe von 3.42
5. Berechne die Zustandswahrscheinlichkeiten einzelner Bediensysteme mit Hilfe von 3.57

Die vorstehend beschriebene Vorgehensweise ist eine scheinbar einfache Möglichkeit, Kennwerte für geschlossene Bediensystemnetzwerke zu berechnen. Bei genauerer Betrachtung der dabei vorzunehmenden Rechenoperationen stellt sich jedoch heraus, daß eine praktische Durchführung der numerischen Berechnungen sehr schnell unmöglich wird.

K	1	2	10	100	10	50	100
M	1	2	10	10	100	50	50
Z	1	3	92.378	$\approx 4,2 \cdot 10^{12}$	$\approx 4,2 \cdot 10^{13}$	$\approx 5,0 \cdot 10^{28}$	$\approx 6,7 \cdot 10^{39}$

Tabelle 3.1: Größe Z des Zustandsraumes eines geschlossenen Bediensystemnetzwerkes für K Kunden und M Bediensysteme.

Dies ist in der großen Zahl möglicher Zustände begründet, die ein geschlossenes Bediensystem annehmen kann. Tabelle 3.1 gibt die Größe des Zustandsraumes für einige Kombinationen von K und M an. Die Berechnung erfolgte mit Hilfe von Formel 3.40.

In Folge der großen Zahl von Zuständen, die angenommen werden können, ist auch die Wahrscheinlichkeit, mit der ein bestimmter Zustand eintritt, sehr klein. Die Größenordnung für die Wahrscheinlichkeit eines Zustand ist im Mittel ungefähr so groß wie $1/Z$, also z.B. für $K = 100$ und $M = 50$ liegen die Zustandswahrscheinlichkeiten in der Größenordnung von 10^{-39} . Dadurch können bei der numerischen Berechnung erhebliche Fehler auftreten.

Außerdem ist die Bestimmung der zulässigen Zustände aufgrund ihrer großen Zahl sehr zeitaufwendig. Könnten beispielsweise auf einem Rechner pro Sekunde 1.000 zulässige Zustände ermittelt und deren Wahrscheinlichkeit berechnet werden, so wären in diesem Beispiel etwa $3 \cdot 10^{27}$ Jahre zur Berechnung aller Zustandswahrscheinlichkeiten notwendig.

Dennoch sind die hier dargestellten Ergebnisse von Bedeutung, da aufbauend auf der Produktformlösung einige effiziente Berechnungsverfahren entwickelt wurden. Hierzu zählen:

- Der Faltungsalgorithmus, entwickelt von Buzen [10]
- LBANC (Local Balance Algorithm for Normalizing Constants) entwickelt von Chandy und Sauer [54]
- SCAT (Self Correcting Approximation Technique) von Neuse und Chandy [52]

Eine gute Übersicht über diese Verfahren geben die Monographie von Bolch [5], sowie das Buch von King [39].

Die Mittelwertanalyse bietet eine gute Basis für Näherungsverfahren für die Kennwertberechnung von Bediensystemnetzwerken mit generell verteilten Bedienzeiten zur Analyse von Materialflusssystemen. Deshalb wird im nachfolgenden Abschnitt die Mittelwertanalyse dargestellt.

3.4.2 Mittelwertanalyse, ein Berechnungsverfahren für geschlossene Bediensystemnetzwerke

Die von Reiser und Lavenberg [53] entwickelte Mittelwertanalyse ermittelt Kennwerte unter Verwendung von Little's Gesetz und einer Betrachtung der Zahl der Kunden, die ein Kunde im Moment seiner Ankunft an einem Bediensystem dort vorfindet. Eine direkte Berechnung der normalisierenden Konstanten wird nicht vorgenommen. Aus diesem Grund ist die Mittelwertanalyse sehr effizient, wenn die Bestimmung von Mittelwerten für die Lösung der Dimensionierungs- oder Analyseaufgabe ausreichend ist.

Sevcik und Mitrani ([55]) sowie Reiser und Lavenberg ([44]) haben gezeigt, welche Zahl von Kunden in einem geschlossenen Bediensystemnetzwerk mit exponentialverteilten Bedienzeiten im Moment einer Kundenankunft erwartet werden kann. Ein gut verständlicher, auf der Operationellen Analyse basierender Beweis kann in [14] nachgelesen werden. Dieses Ergebnis ist als Ankunftstheorem, oder genauer als Satz über die Verteilung der Kundenzahl im Ankunftszeitpunkt bekannt.

Satz 3.1 (Über die Verteilung der Kundenzahl im Ankunftszeitpunkt)

Die Wahrscheinlichkeit, daß ein am Bediensystem i eintreffender Kunde das Bediensystemnetzwerk mit K Kunden im Zustand $(k_1, k_2, \dots, k_i - 1, \dots, k_M)^T$ vorfindet, ist gleich der Gleichgewichtswahrscheinlichkeit des Zustandes $(k_1, k_2, \dots, k_i - 1, \dots, k_M)^T$ im gleichen Bediensystemnetzwerk mit $K - 1$ Kunden im Umlauf.

Mit Hilfe des Ankunftstheorems kann die Verweilzeit $t_{v,i}$ eines gerade am Bediensystem i eingetroffenen Kunden in derselben Weise bestimmt werden, wie es für das $M|G|1$ -Bediensystem durchgeführt wurde. Die Verweilzeit setzt sich zusammen aus der Bearbeitungszeit des gerade in Bearbeitung befindlichen Kunden und der Summe der Bearbeitungszeiten aller bereits wartender Kunden. Hinzu kommt die erwartete Bedienzeit des eingetroffenen Kunden $t_{ab,i}$. Da die Bedienzeit exponential verteilt ist, besteht kein Unterschied zwischen der Verteilung der Restbedienzeit des gerade bedienten Kunden und der Verteilung der Bedienzeit. Da die erwartete Verweilzeit an einem Bediensystem von der Zahl im Netzwerk umlaufender Kunden K abhängt, soll ein von K abhängiger Wert der erwarteten Verweilzeit $t_{v,i}(K)$ bestimmt werden. Für den Mittelwert der Verweilzeit $t_{v,i}(K)$ an einem Bediensystem i mit einer Bedienstation gilt:

$$t_{v,i}(K) = \frac{1}{\mu_i} \left[1 + \underbrace{N_i(K-1)}_{\text{Ankunftstheorem}} \right] \quad (3.58)$$

Der von der Zahl der Kunden abhängende Gesamtdurchsatz des Netzes $\lambda(K)$ wird mit Hilfe der Verweilzeit der Kunden K an den einzelnen Bediensystemen ermittelt.

Hierzu verwenden wir die bereits mit Hilfe von 3.44 und der Normierung 3.46 berechneten Besuchshäufigkeiten und gewichten mit ihnen die an den Bediensystemen anfallenden Verweilzeiten.

Dividiert man nun die Gesamtzahl der Kunden K durch die im Mittel für den Umlauf eines Kunden durch das Bediensystemnetzwerk benötigte Zeit $\sum_{i=1}^M e_i \cdot t_{v,i}(K)$, so kann der Gesamtdurchsatz berechnet werden.

$$\lambda(K) = \frac{K}{\sum_{i=1}^M e_i \cdot t_{v,i}(K)} \quad (3.59)$$

Mit Hilfe von Little's Gesetz kann außerdem ein Zusammenhang zwischen der mittleren Zahl von Kunden $N_i(K)$ an dem Bediensystem i mit Ankunftsrate $\lambda_i(K)$ sowie Verweilzeit $t_{v,i}(K)$ an diesem Bediensystem hergestellt werden. Die ermittelten Werte gelten immer für eine feste Zahl K von Kunden im System.

$$N_i(K) = \lambda_i(K)t_{v,i}(K) \quad (3.60)$$

Dabei ist $\lambda_i(K) = \lambda(K) \cdot e_i$.

Offensichtlich ist $N_i(0) = 0$. Damit und mit den Ausdrücken 3.58 bis 3.60 ist die Voraussetzung geschaffen, iterativ Kennwerte für das Bediensystemnetzwerk zu berechnen, die ausschließlich aus Einkanalbediensystemen bestehen.

Bei Mehrkanalsystemen ist die Bestimmung der Verweilzeit $t_{v,i}(K)$ aufwendiger, da 3.58 nicht in dieser einfachen Form gilt. Zur Berechnung wird die bedingte Wahrscheinlichkeit $p_i(j|l)$ herangezogen, die angibt, mit welcher Wahrscheinlichkeit j Kunden am Bediensystem i gleichzeitig bedient werden, wenn sich im Gesamtnetzwerk l Kunden befinden. Es gilt (siehe [5], S. 105ff.):

$$p_i(j|l) = \frac{\lambda_i(l)}{\mu_i \cdot j} \cdot p_i(j-1|l-1) \quad j = 0, \dots, m_i - 1 \quad (3.61)$$

Für ein leeres Netzwerk ist offensichtlich

$$p_i(0|0) = 1 \quad (3.62)$$

zutreffend.

Die Wahrscheinlichkeit eines leeren Systems kann als Residuum aller Zustandswahrscheinlichkeiten $p_i(j|l)$ berechnet werden.

$$p_i(0|l) = 1 - \sum_{k=1}^l p_i(k|l) \quad (3.63)$$

Die Verweilzeit läßt sich damit für Mehrkanalsysteme wie folgt berechnen:

$$t_{v,i}(K) = \frac{1}{\mu_i \cdot m_i} \left[1 + N_i(K-1) + \sum_{j=0}^{m_i-2} (m_i - j - 1) p_i(j|K-1) \right] \quad (3.64)$$

Damit sind alle notwendigen Ausdrücke für die Berechnung von Mittelwerten für Leistungskenngrößen von geschlossenen Bediensystemen mit exponentialverteilten Bedienzeiten zusammengestellt und es resultiert die nachfolgend dargestellte Vorgehensweise.

Berechnungsschritte der Mittelwertanalyse

1. Bestimme die Besuchshäufigkeiten e_i mit Hilfe von 3.44 und 3.46
2. Initialisiere die Kennwerte

$$\begin{aligned} N_i(0) &= 0 & i &= 1, \dots, M \\ p_i(0|0) &= 1 & i &= 1, \dots, M \\ p_i(j|0) &= 0 & j &= 1, \dots, m_i - 1 \end{aligned}$$

3. Iteriere über der Zahl der Kunden $l = 1$ bis K :
 - (a) Bestimme $t_{v,i}$ nach Formel 3.58 bzw. 3.64 für alle Bediensysteme $i = 1, \dots, M$
 - (b) Berechne den Durchsatz des Netzwerkes für l Kunden mit 3.59
 - (c) Berechne den Durchsatz an den Bediensystemen aus $\lambda_i = \lambda \cdot e_i$
 - (d) Bestimme die Zahl der Kunden an jedem Bediensystem mit Hilfe von Formel 3.60
4. Abschließend können für die Bediensysteme die aktuellen Werte anderer Kenngrößen, wie z.B. Auslastungsgrade ρ_i , Wartezeiten $t_{w,i}$ bestimmt werden.

Als Nachteil der Mittelwertanalyse wird in der Literatur der hohe Speicherbedarf genannt, da alle Größen während der Iteration über die Auftragszahlen gespeichert werden müssen, was insbesondere bei den für Mehrkanalsysteme zu berechnenden $p_i(j|l)$ ins Gewicht fällt. Für praxisrelevante Fälle aus dem Materialflußbereich stellen jedoch die Speicheranforderungen (Größenordnung K^2 Werte) auf modernen Arbeitsplatzrechnern kein Hindernis dar.

Die Anwendung der Mittelwertanalyse auf große Netzwerke mit Mehrkanalbediensystemen wird durch die Formel 3.63 begrenzt, wenn die Berechnung der Differenz an die Grenzen der Rechengenauigkeit von Digitalrechnern stößt.

Ein Vorzug der Mittelwertanalyse besteht darin, daß nicht nur der Durchsatz für K Kunden, sondern auch für alle kleineren Werte bestimmt wird. Da der Durchsatz geschlossener Bediensystemnetzwerke mit unbeschränkten Warteräumen wächst, wenn Kunden hinzugefügt werden, können mit Hilfe der Mittelwertanalyse solange Kunden dem System hinzugefügt werden, bis der erforderliche Durchsatz erbracht wird oder der Sättigungsdurchsatz erreicht wird. Bei bekanntem Maximaldurchsatz λ_{max} kann das typische Durchsatzverhalten eines geschlossenen Bediensystemnetzwerkes (siehe Abbildung 3.10) iterativ ermittelt werden.

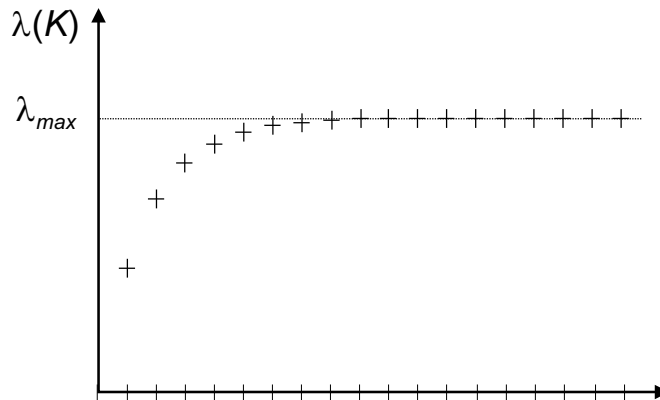


Abbildung 3.10: Durchsatz $\lambda(K)$ und Maximaldurchsatz λ_{max} eines geschlossenen Bediensystemnetzwerkes

Die Bedeutung der Mittelwertanalyse für Anwendungen in der Materialflußplanung liegt vor allem darin, daß sie als Grundlage zahlreicher Näherungsverfahren verwendet werden kann. Als Beispiel hierfür werden in den nächsten Abschnitten die „Antwortzeiterhaltung“⁴ und das Verfahren von Marie vorgestellt.

3.5 Geschlossene Bediensystemnetzwerke mit generell verteilten Bedienzeiten

Für geschlossene Bediensystemnetzwerke mit generell verteilten Bedienzeiten sind keine exakten Kennwerte berechenbar. Deshalb wurden einige Näherungsverfahren entwickelt. Am Beispiel der „Antwortzeiterhaltung“ läßt sich gut darstellen, wie Näherungsverfahren auf Verfahren zur Ermittlung exakter Ergebnisse aufgebaut werden können. Mit dem Verfahren von Marie wird anschließend ein Verfahren vorgestellt, das eine bessere Näherungsqualität als die „Antwortzeiterhaltung“ ermöglicht.

⁴Im Englischen Response Time Preservation genannt und deshalb unter der Abkürzung RTP bekannt

3.5.1 Das Verfahren der Antwortzeiterhaltung

Das Verfahren der Antwortzeiterhaltung entwickelt von Agrawal et al. [1] sei zuerst für Bediensystemnetzwerke dargestellt, die ausschließlich aus Einkanal-Bediensystemen bestehen. Ziel des Verfahrens ist vor allem die Approximation des Gesamtdurchsatzes. Bei konstanter Kundenzahl K hängt der Gesamtdurchsatz davon ab, wie groß die Verweilzeit der Kunden an den Bediensystemen jeweils ist. Deshalb steht die Abschätzung der mittleren Verweilzeiten $t_{v,i}$ im Mittelpunkt dieses Verfahrens. Die Verweilzeit eines Kunden in einem Bediensysteme wird im Zusammenhang mit geschlossenen Bediensystemnetzwerken häufig dazu verwendet, das Antwortverhalten von Host-Systemen abzubilden. Deshalb ist dieses Verfahren in der Literatur als „Antwortzeiterhaltung“ (im Englischen *response time preservation*) bekannt.

Das Verfahren basiert auf einer Aufteilung des Gesamtproblems in eine „Mikrosicht“, in der die einzelnen Bediensysteme betrachtet werden und eine „Makrosicht“, in der das gesamte Netzwerk abgebildet ist.

In der Mikrosicht wird ausgenutzt, daß die exakte Berechnung von Kennwerten für $M|G|1$ -Bediensystemen möglich ist, wobei vereinfacht angenommen wird, daß sich die Ankunftsprozesse an einem Knoten zu exponentialverteilten Zwischenankunftszeiten überlagern.

Hierfür ist jedoch zumindest die Kenntnis der Ankunftsraten λ_i notwendig, wenn sich K Kunden im Bediensystemnetzwerk befinden. Die Ankunftsraten werden in der Makrosicht mit Hilfe eines geschlossenen Bediensystemnetzwerkes mit exponentialverteilten Bedienzeiten ermittelt.

Im ersten Schritt werden Kennwerte, insbesondere die Ankunftsraten λ_i für alle Bediensysteme z.B. mit Hilfe der Mittelwertanalyse bestimmt. Dabei werden die Bedienraten μ_i der Bediensysteme wie vorgegeben verwendet. Mit Hilfe der bekannten Ankunftsraten λ_i können nun Kennwerte der freigeschnittenen Bediensysteme mit Hilfe der Ausdrücke 2.99 bis 2.101 für $M|G|1$ -Bediensystem-Modelle berechnet werden. Von besonderem Interesse sind die Verweilzeiten $t_{v,i}$, die mit

$$t_{v,i}^{M|G|1} = \frac{\rho_i \cdot E(t_{ab,i})^2}{2 \cdot E(t_{ab,i})(1 - \rho_i)} \quad (3.65)$$

berechnet werden. Die so ermittelten Verweilzeiten werden im allgemeinen nicht mit denjenigen übereinstimmen, die für ein $M|M|1$ -Bediensystem mit den gleichen Bedien- und Ankunftsraten ermittelt worden wären. Deshalb werden im nächsten Schritt äquivalente Bedienraten μ'_i bestimmt, die bei einem $M|M|1$ -Bediensystem zur gleichen Verweilzeit $t_{v,i}^{M|G|1}$ wie für das isolierte $M|G|1$ -Bediensystem führen.

$$\begin{aligned}
 t_{v,i}^{M|M|1} &\stackrel{!}{=} t_{v,i}^{M|G|1} \\
 \Leftrightarrow t_{v,i}^{M|G|1} &= \frac{1/\mu'_i}{1 - \lambda_i/\mu'_i} \\
 \Rightarrow \mu'_i &= \frac{1 + \lambda_i \cdot t_{v,i}^{M|G|1}}{t_{v,i}^{M|G|1}} \quad (3.66)
 \end{aligned}$$

Mit Hilfe dieser angepassten Bedienraten wird wieder eine Berechnung für das Bediensystemnetzwerk vorgenommen und neue Ankunftsraten λ_i bestimmt. Das Verfahren bricht ab, sobald sich der Neuberechnete Durchsatz des Netzes nur noch marginal von dem im vorigen Schritt berechneten unterscheidet.

Zusammengefaßt läßt sich der Ablauf der Berechnung wie folgt beschreiben:

Berechnungsschritte der Antwortzeiterhaltung

1. Schätze den Gesamtdurchsatz des Netzwerkes λ und die Ankunftsraten λ_i mit Hilfe eines geschlossenen Bediensystemnetzwerkes mit gleichen jedoch exponentialverteilten Bedienzeiten und ansonsten identischer Struktur. Merke den ermittelten Gesamtdurchsatz $\lambda^{vor} = \lambda$.
2. Analysiere die Knoten unter der Annahme exponentialverteilter, zustandsunabhängiger Ankunftsraten λ_i und den vorgegebenen Werten für $E(t_{ab,i})$ und $E(t_{ab,i}^2)$. Bestimme die Verweilzeiten $t_{v,i}^{M|G|1}$ für alle Knoten nach 3.65.
3. Bestimme die äquivalenten Bedienraten, die für $M|M|1$ -Bediensysteme zu gleichen Verweilzeiten führen würden mit Hilfe von Formel 3.66.
4. Analysiere das Netz als geschlossenes Bediensystemnetzwerk mit exponentialverteilten Bedienzeiten $1/\mu'_i$. Bestimme den neuen Wert für den Gesamtdurchsatz λ^{neu} .
5. Vergleiche den neu berechneten Netzdurchsatz λ^{neu} mit dem im vorigen Durchlauf ermittelten λ^{vor} . Falls sich die Veränderung des Durchsatzes $|\lambda^{vor} - \lambda^{neu}| > \epsilon$ noch über einer vorgegebenen Schwelle ϵ befindet, dann führe eine erneute Iteration ab Schritt 2 durch.

Eine Erweiterung des Verfahrens der Antwortzeiterhaltung auf Netzwerke mit $M|G|m$ -Bediensystemen ist möglich, indem die Approximationen aus Abschnitt 2.11 zu Hilfe genommen werden.

Das Verfahren der Antwortzeiterhaltung kann zu schlechten Approximationsergebnissen führen, die insbesondere dann auftreten, wenn sich wenige Kunden im Bediensystemnetzwerk befinden. Der Grund hierfür liegt in der Behandlung der Ankunftsprozesse als zustandsunabhängige Markov-Prozesse. Es ist jedoch einsichtig, daß insbesondere in Netzen mit kleiner Kundenzahl K eine starke Zustandsabhängigkeit bei den Ankunftsrate besteht. Ein Selbstregulierungseffekt tritt ein, da die Ankunftsrate nachläßt, wenn sich an einem Knoten bereits viele Kunden angesammelt haben und deshalb die Wahrscheinlichkeit von Ankünften geringer wird. In der Materialflußplanung kann dieses Verhalten zum Beispiel von Bedeutung sein, wenn geschlossene Fahrzeugsysteme abgebildet werden.

Die Zustandsabhängigkeit der Ankunftsrate ist in dem Verfahren von Marie, das im nächsten Abschnitt vorgestellt wird, wesentlich besser abgebildet.

3.5.2 Das Verfahren von Marie

Im von Raymond Marie (siehe [49], [50]) entwickelten Verfahren wird ebenfalls eine Mikro- und ein Makrosicht benutzt, um Kennwerte iterativ zu bestimmen. In der Mikrosicht werden $\lambda(k)|C_n|1$ -Bediensysteme (siehe Abschnitt 2.7.2) verwendet, die es erlauben, zustandsabhängige Ankunftsrate in die Betrachtung miteinzubeziehen.

Die zustandsabhängigen Ankunftsrate an einem Bediensystem werden für jedes einzelne Bediensystem bestimmt, indem jedes der Bediensysteme aus dem Netz genommen und durch einen Kurzschluß überbrückt wird. In Abbildung 3.11 ist das Bediensystemnetzwerk zu sehen (a) und die Überbrückung von Bediensystem i (b) zur Berechnung der zustandsabhängigen Ankunftsrate $\lambda_i(k_i)$.

Für das kurzgeschlossene Bediensystemnetzwerk sind nun die Durchsätze für jede mögliche Zahl von Kunden $K - k_i$ für $k_i = 0, \dots, K - 1$ zu bestimmen. Der Fluß auf der kurzgeschlossenen Verbindung, die sich an der Stelle des Bediensystem i befindet, ist äquivalent zur zustandsabhängigen Ankunftsrate im Zustand k_i an diesem Bediensystem i , wenn sich in dem Ersatznetzwerk $K - k_i$ Kunden befinden. Es wird dabei wie bei dem Verfahren der Antwortzeiterhaltung unterstellt, daß die Bedienzeiten an den Bediensystemen des Netzes exponentialverteilt sind. Deshalb können effiziente Berechnungsverfahren eingesetzt werden, wobei sich für das Verfahren von Marie die Mittelwertanalyse (siehe Abschnitt 3.4.2) anbietet, da der iterative Berechnungsalgorithmus die Durchsätze sukzessive für $K' = 1, \dots, K$ bestimmt. Bei Abschluß der Mittelwertanalyse liegen bereits für alle möglichen Zustände k_i die entsprechenden zustandsabhängigen Ankunftsrate vor.

Die Mittelwertanalyse muß allerdings modifiziert werden, da die Bedienrate der $\lambda(k)|C_n|1$ -Bediensysteme ebenfalls zustandsabhängig sind. Dies kann durch eine

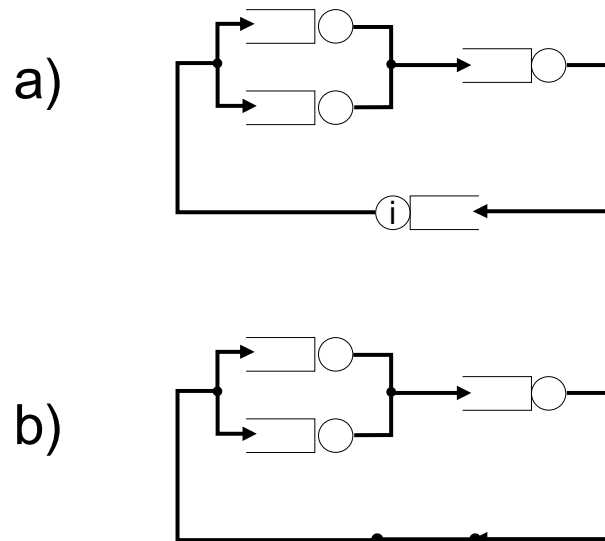


Abbildung 3.11: Kurzschluß des Bediensystems i zur Kennwertberechnung nach dem Verfahren von Marie

Erweiterung der Mittelwertanalyse erreicht werden, die analog zu der in Abschnitt 3.4.2 kurz dargestellten Erweiterung in Bezug auf Mehrkanalbediensysteme vorgeht.

Ausdruck 3.58 wird um Bedienraten $\nu_i(k)$ erweitert, die zu Verweilzeiten abbilden die spezifisch für eine bestimmte Kundenanzahl im System sind.

$$t_{v,i}(K) = \sum_{j=0}^{K-1} \left[p_i(j|K-1)(j+1) \cdot \frac{1}{\nu_i(j+1)} \right] \quad (3.67)$$

Die Ausdrücke 3.59 und 3.60 können aus der Mittelwertanalyse übernommen werden. Die iterative Berechnung der Zustandswahrscheinlichkeiten (siehe 3.61) kann für zustandsspezifische Bedienraten zu

$$p_i(j|l) = \frac{\lambda_i(l)}{\nu_i(j)} \cdot p_i(j-1|l-1) \quad j = 1, \dots, K \quad (3.68)$$

erweitert werden. Die Werte $p_i(0|0)$ und $p_i(0|l)$ sind entsprechend 3.62 und 3.63 zu bestimmen.

Die Hilfsnetze werden für jedes Bediensystem, das keine exponentialverteilten Bedienzeiten besitzt, erzeugt und mit ihrer Hilfe werden die zugehörigen Ankunftsrate bestimmt.

Danach wird in die Mikrosicht übergewechselt, in der die einzelnen Bediensysteme in Isolation, jedoch bei zustandsabhängigen Ankunftsrate analysiert werden. Hierfür werden die in Abschnitt 2.7.2 dargestellten Ergebnisse verwendet.

Nach jedem Berechnungsschritt ist zu überprüfen, ob die Ergebnisse der Mikrobetrachtungen mit den Ergebnissen des Makromodells übereinstimmen. Hierfür schlägt Marie zwei Kennwerte vor. Zum einen ist zu überprüfen, ob die Gesamtzahl von Kunden, die sich aus der Summation der Einzelanalysen ergibt, mit der vorgegebenen festen Kundenzahl bis auf eine vorzugebende relative Differenz ϵ übereinstimmt.

$$\left| \frac{K - \sum_{i=1}^M N_i}{K} \right| \stackrel{?}{<} \epsilon_1 \quad (3.69)$$

Zum anderen dient die Übereinstimmung der Durchsätze als Überprüfungs-kriterium.

$$\left| \frac{\frac{\lambda_j}{e_j} - \frac{1}{M} \sum_{i=1}^M \frac{\lambda_i}{e_i}}{\frac{1}{M} \sum_{i=1}^M \frac{\lambda_i}{e_i}} \right| \stackrel{?}{<} \epsilon_2 \quad \forall j = 1, \dots, M \quad (3.70)$$

Berechnungsschritte des Verfahrens von Marie

1. Erzeuge für jeden Knoten $i = 1, \dots, M$ ein Hilfsnetz, das diesen Knoten nicht enthält. Setze die zustandsabhängigen Bedienraten in den Bediensystemen auf:

$$\mu_j(k_j) = \begin{cases} 0 & \text{für } k_j = 0 \\ \mu_j & \text{für } k_j > 0 \end{cases} \quad \forall j = 1, \dots, M \quad (3.71)$$

2. Berechne bei allen Hilfsnetzen die zustandsabhängigen Ankunfts-raten $\lambda_i(k_i)$ mit Hilfe der Ausdrücke 3.67, 3.68, sowie 3.62, 3.59 und 3.63 mit Hilfe der angepaßten Mittelwertanalyse und der Beziehung $\lambda_i = e_i \cdot \lambda$.
3. Analysiere jeden Knoten mit Hilfe der im vorigen Schritt berechneten zustandsabhängigen Ankunfts-raten unter Nutzung der Ergebnisse aus Abschnitt 2.7.2.
4. Überprüfe die Konvergenzkriterien 3.69 und 3.70. Falls die Ergebnisse unbefriedigend sind, setze die im vorigen Schritt bestimmten zustandsabhängigen Bedienraten ein

$$\mu_j(k_j) = \nu_j(k_j) \quad \forall j = 1, \dots, M \quad (3.72)$$

und fahre mit Schritt 2 fort.

3.6 Modellierung und Optimierung von Produktionssystemen mit Hilfe von bedientheoretischen Modellen

Im Verlauf der Planung eines Fertigungssystems und des zugehörigen Materialflusssystems sind bedientheoretische Modelle dann sinnvoll einzusetzen, wenn statische Auslastungsberechnungen keine aussagekräftigen Ergebnisse (z.B. bezüglich der Durchlaufzeiten und Fertigungsbestände) liefern können, die Zahl der zu betrachtenden Alternativen jedoch noch so groß ist, daß eine Analyse aller Möglichkeiten mit Hilfe von Simulationsmodellen zu aufwendig ist.

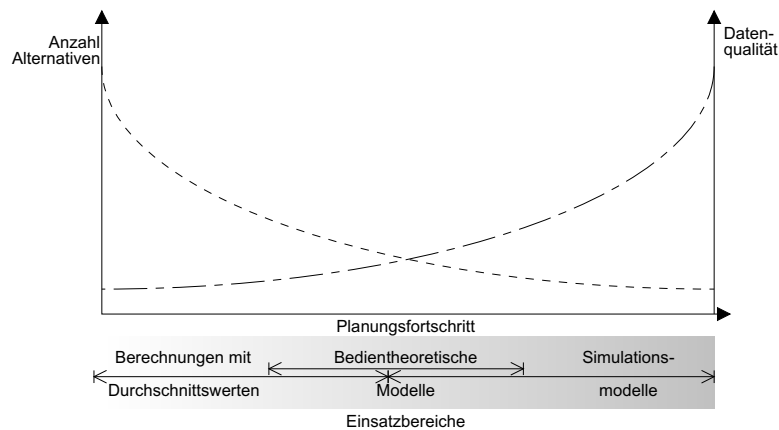


Abbildung 3.12: Anwendungsbereiche verschiedener Modellierungsarten (nach [57])

Da sich Kennwerte für bedientheoretische Modelle analytisch berechnen oder mit Hilfe eines Simulationsmodelles ermitteln lassen, ist eine analytische Kennwertermittlung nicht zwingend zum Einsatz bedientheoretischer Modelle notwendig (siehe auch Abbildung 3.12). Das Verständnis für die Haupteinflussfaktoren, daß sich aus der Herleitung analytischen Berechnungsverfahren ergibt, hilft bei der Analyse von Simulationsergebnissen. Hieraus wiederum können Hinweise auf Verbesserungsmöglichkeiten gezogen werden.

3.6.1 Hinweise zur Gestaltung von Materialflusssystemen

Wie bereits im vorigen Abschnitt deutlich wurde, ist die Reduzierung der Variabilitäten in Bedienungs- und Zwischenankunftszeiten eine Möglichkeit, ohne Kapazitätserweiterung Durchlaufzeiten in Materialflusssystemen zu reduzieren (siehe auch Fromm in [22]).

Neben der Möglichkeit der Losgrößenharmonisierung im operativen Bereich, die im folgenden Abschnitt behandelt werden wird, beeinflusst die Gestaltung der Fertigungsstrukturen die Variabilität von Zwischenankunfts- und Bedienzeiten erheblich.

Der Schwerpunkt des QNA liegt auf der Berechnung der Variabilität, die durch Bedienprozesse, Verzweigungen und Zusammenführungen erzeugt wird. Aus den Formeln des QNA können folgende Schlüsse gezogen werden:

- Falls eine Maschine viele verschiedene Nachfolger hat, führt dies dazu, daß die Variabilität der Ankunftsprozesse der Nachfolger gegen 1 strebt, selbst wenn der Bedienprozeß der Maschine getaktet ist (siehe Abbildung 3.7). Wird z.B. eine Werkstattfertigung mit einer Waschmaschine ausgestattet, die nach einigen Arbeitsschritten die notwendige Reinigung gemeinsam für alle Teile vornimmt, so führt dies aufgrund der großen Zahl möglicher Nachfolger und der damit verbundenen kleinen Verzweigungswahrscheinlichkeit in den partiellen Fluß $\lambda_{i,j}$ zu einem dieser Nachfolger dazu, daß die Variabilität $c_{i,j}^2$ in diesem partiellen Fluß ungünstig beeinflusst wird, falls die Variabilität der Bearbeitungszeiten deutlich kleiner als 1 ist.
- Der Abgangsprozeß einer Maschine mit niedriger Variabilität der Bedienzeiten hat eine geringe Variabilität, wenn die Maschine
 - hoch ausgelastet ist ($\rho \rightarrow 1$) oder
 - einen Ankunftsprozeß niedriger Variabilität besitzt.

Aus diesem Grund können auf das Gesamtsystem bezogen Verbesserungen erreicht werden, wenn Maßnahmen zur Verstetigung der Bedienzeit getroffen werden, selbst wenn dadurch die Auslastung der Maschine erhöht wird. Beispiele für solche Maßnahmen sind: Vorbeugende Instandhaltung, oder standardisierte, vorbereitete Umrüstungsabläufe.

In Abschnitt 2.10.4 wurde darüberhinaus deutlich, daß die Reduzierung der Fertigungsunterbrechungen durch Störungen, sowie die Begrenzung der Ausfallzeitdauern ebenfalls einen wesentlich stärkeren Effekt hat, als alleine aus der Betrachtung der Auswirkungen auf die Auslastung geschlossen werden kann.

3.6.2 Optimierung der Losgrößen in Werkstattfertigungen

Optimierungsverfahren

In Fortsetzung der Überlegungen in Abschnitt 2.9.2 ist eine Übertragung der Vorgehensweise auf vernetzte Werkstattfertigungen möglich.

Der mittlere Gesamtbestand einer Werkstattfertigung kann mit Hilfe eines offenen Bediensystemnetzwerkes abgeschätzt werden. Sowohl die Methode von Kühn (siehe [43]) wie auch QNA schneiden die einzelnen Bediensysteme frei und behandeln die Maschinen und Maschinengruppen als $G|G|1$ - bzw. $G|G|m$ -Bediensysteme.

Die Betrachtung von Verzweigungen und Zusammenführungen in Abschnitt 3.3.2 hat gezeigt, daß die Variabilität der Zwischenankunftszeiten bei starker Vernetzung in der Größenordnung von 1, also in der Nähe der Variabilität von Exponentialverteilungen liegt.

Deshalb bietet es sich an, den Materialfluß eines Fertigungssystems mit mehreren Maschinen als Kumulation unabhängiger $M|G|1$ - bzw. $M|G|m$ -Bediensysteme zu betrachten. Unter Nutzung dieser Annahme können die Ergebnisse aus Abschnitt 2.9.2 weiterverwendet werden.

Die Losgrößen l_h für die Produkte $h = 1, \dots, P$ seien auf allen Maschinen $i = 1, \dots, M$ gleich und werden zu einem Losgrößenvektor $l = (l_1, l_2, \dots, l_P)^T$ zusammengefaßt. Die Stückbearbeitungszeiten $t_{b,i,h}$ und die Rüstzeiten $t_{r,i,h}$ können produkt- und maschinenbezogen verschieden sein. Aus dieser Modellierung resultieren M Teilmodelle mit Kosten K_i , $K_{l,i}$, $K_{r,i}$ und $K_{f,i}$ aus denen die Gesamtkosten für einen gegebenen Losgrößenvektor l durch

$$K = \sum_{i=1}^M K_i \quad (3.73)$$

durch Summation erhalten werden.

Unterstellt man die Konvexität der einzelnen Kostenverläufe, so ist die Gesamtkostenfunktion als Summe konvexer Funktionen ebenfalls konvex. Sie läßt sich für zwei Produkte graphisch in einem dreidimensionalen Gebilde darstellen. Für ein Beispiel mit drei Maschinen und zwei Produkten ist ein Teil des Lösungsraumes in Abbildung 3.13 dargestellt.

Wie in Abschnitt 2.9.2 wurde für die Durchführung der Optimierung das Gradientenverfahren verwendet, was zu den in Abbildung 3.13 dargestellten Suchweg führte.

Der Vorzug dieser Vorgehensweise liegt in der stark vereinfachten Berechnung der Eingangsgrößen für die Bestimmung der Kostenkomponente $K_{f,i}$. Aufgrund der in der Modellierung unterstellten gegenseitigen Unabhängigkeit der Bediensysteme und damit ihrer Ankunftsprozesse wäre auch eine parallelisierte Durchführung dieser Berechnung möglich. Da die Bestimmung der Kosten $K_{f,i}$ den zeitaufwendigsten Teil des Optimierungsverfahrens darstellt, würde daraus eine weitere Geschwindigkeitssteigerung resultieren.

Die Durchführung der Optimierung mit Hilfe des Gradientenverfahrens für Probleme mit ca. 100 verschiedenen Produkten und 60 Maschinen ist jedoch auch auf einem PC in Zeiten unterhalb von einer Minute durchführbar, deshalb ist eine

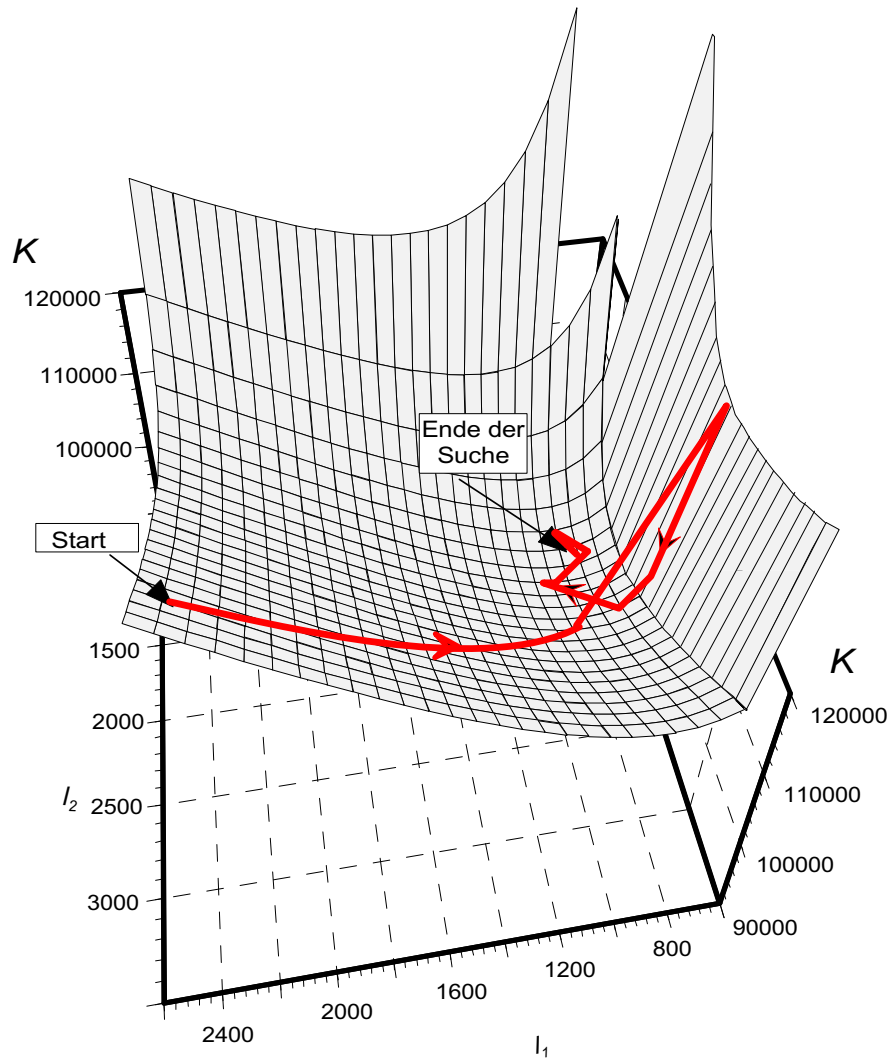


Abbildung 3.13: Verlauf der Kostenfunktionen K , bei Fertigung auf drei Maschinen und Variation der Losgrößen l_1 und l_2 für zwei Produkte (nach Greiling [30])

Parallelisierung nur bei deutlich größeren Problemstellungen sinnvoll.

Am Institut für Fördertechnik und Logistiksysteme (IFL) der Universität Karlsruhe wurde auch untersucht, welche Einbußen in Bezug auf die Lösungsqualität bei dem stark vereinfachten Modell auf $M|G|1$ -Basis im Vergleich zu einer Optimierung mit Hilfe des QNA von Whitt zu verzeichnen sind.

Hierfür wurde die Implementierung des QNA im Rahmen des Programmpaketes PEPSY (siehe [6]) in Verbindung mit einem übergeordneten, am IFL implementierten Gradientenverfahren genutzt. Die mit beiden Optimierungsmethoden erzielten Ergebnisse wurden anschließend mit Hilfe eines Simulationsmodells

miteinander verglichen (siehe Greiling in [30]). Dabei konnte nicht festgestellt werden, daß mit Hilfe der aufwendigeren Modellierung unter Berücksichtigung der Zwischenabgangs- und -ankunftszeiten ein systematisch besseres Ergebnis zu erzielen ist.

Einsatz in der Praxis

Der Einsatz der Optimierungsmethode wurde in Kooperation mit einem Unternehmen versuchsweise zur Optimierung der Losgrößen in einer Gangräderfertigung für Nutzfahrzeuggetriebe eingesetzt. Zum Zeitpunkt der Durchführung der Optimierung wurden 152 verschiedene Produkte auf 66 verschiedenen Maschinen hergestellt. Die abzubildenden Arbeitspläne enthielten bis zu 10 Produktionsschritte, im Durchschnitt waren 4,8 Arbeitsschritte für den Durchlauf eines Produktes durch das Fertigungssegment notwendig.

Die Fertigung der Produkte erfolgt im untersuchten Fertigungsabschnitt programmbezogen, weshalb eine Zusammenfassung der Produktionsmengen zu Losen sinnvoll ist. Vor der Einführung des bedientheoretischen Optimierung erfolgte die Losgrößenbestimmung aufgrund eines an der ersten Bearbeitungsmaschine ausgerichteten Andler'schen Losgrößenmodells. Aus der Anwendung dieses Losgrößenmodells resultieren erhebliche Schwankungen in den Losgrößen und Bearbeitungszeiten, so daß, wie mit Hilfe einer Istaufnahme festgestellt wurde, die Variabilitäten der Bearbeitungszeiten an einigen Maschinen deutlich über dem Wert „1,“ lagen.

Die Losgrößenoptimierung erfolgt im Rahmen einer rollierenden Planung. Hierzu werden aus der Programmplanung (Horizont bis zu einem Jahr) die zu fertigen Mengen und aus den Arbeitsplänen die benötigten Ressourcen, Stück- und Rüstzeiten entnommen.

Die Daten der Fertigungsstruktur, zu denen u.a. die Anzahl der Maschinen pro Ressourcengruppe, Verfügbarkeiten und Standzeiten der Werkzeuge gehören, werden manuell erfaßt und, da es sich um selten ändernde Daten handelt, manuell gepflegt.

Wie im vorhergehenden Abschnitt dargestellt, werden die Daten aus Produktionsprogramm, Stücklisten und Fertigungsstruktur zu einem bedientheoretischen Modell verdichtet. Für die Losgrößen, die als Ergebnis des Verfahrens zu bestimmen sind, muß eine gültige Startlösung vorgegeben werden, die sicherstellt, daß die Auslastung ρ_i an keinem der Bediensysteme größer als „1“ ist.

Die Optimierung auf der Basis des bedientheoretischen Modells mit Hilfe der Gradientenmethode wird durch das Software-Paket KOLA (**K**ostenoptimierende **L**osgrößen**a**nalyse) durchgeführt und hat als Ergebnis einen Losgrößenvektor.

Nicht alle Gegebenheiten der abgebildeten Gangräderfertigung konnten mit Hilfe des bedientheoretischen Modells abgebildet werden. Die wichtigsten Aspekte,

die im bedientheoretischen Modell vereinfacht abgebildet werden, sind:

Rüstzeiten Die Rüstzeiten an den Maschinen in der Fertigung können reihenfolgeabhängig sein. Gehören zwei Produkte zur gleichen Produktfamilie, da sie gemeinsame geometrische Eigenschaften besitzen, so kann die Rüstzeit von einem Produkt der Produktfamilie zu einem anderen Produkt aus der Produktfamilie kürzer sein, als wenn die Maschine für ein Produkt aus einer anderen Produktfamilie gerüstet werden muß.

Für das bedientheoretische Modell wurden mittlere Rüstzeiten ermittelt, die unabhängig von der tatsächlichen Reihenfolge der Lose zur Bearbeitungszeit eines Loses hinzuaddiert werden. Hierdurch ist zumindest eine gute Abbildung der Maschinenauslastung gewährleistet.

Werkzeugtausch Bei einigen der Maschinen muß während der Bearbeitung des Loses das Bearbeitungswerkzeug ausgetauscht werden, da die Standzeit des Werkzeuges bei der zu erzielenden Qualität kürzer als die Losbearbeitungszeit ist. Dieser zusätzliche Zeitbedarf muß ebenfalls über eine Istaufnahme ermittelt und zur Losbearbeitungszeit hinzuaddiert werden.

Überlappende Fertigung Da die Bearbeitungszeiten der Lose insbesondere vor der Anwendung der bedientheoretischen Optimierung sehr groß sein können, werden die ersten Teile eines Loses bereits zur Bearbeitungsmaschine für den nächsten Bearbeitungsschritt weitergeleitet, bevor die letzten Teile an der Vorgängermaschine bearbeitet sind.

Aus dieser überlappenden Fertigung kann das sogenannte „Lossplitting“ resultieren, wenn die Teillose nicht mehr zusammengeführt werden. Mit Hilfe des Lossplittings können Durchlaufzeitreduzierungen erreicht werden, allerdings muß eine Erhöhung der Rüstvorgänge in Kauf genommen werden, wenn es nicht gelingt, die an einer Maschine getrennten Teile eines Loses wieder zusammenzuführen.

Reihenfolgeplanung Die Reihenfolgeplanung an den Bearbeitungsmaschinen wird vor Ort durch die Meister und Einrichter vorgenommen. Grundsätzlich wird nach FCFS vorgegangen, knappe Produkte können jedoch vorgezogen werden. Die Priorisierung knapper Produkte kann so weit gehen, daß die Fertigung des laufenden Loses unterbrochen wird, um ein Produkt mit dringendem Bedarf zu fertigen.

Auch auf diese Weise kommt es zu dem bereits beschriebenen Lossplitting. Weiterhin können Werker aus verschiedenen Gründen (z.B. Verfügbarkeit von Werkzeugen, Material und Hilfsstoffen) von der geplanten Reihenfolge abweichen.

Beim praktischen Einsatz bedientheoretischer Optimierungsverfahren muß sichergestellt werden, daß die erzielten Ergebnisse auch tatsächlich zu Verbesserungen im Ablauf führen. Da das für die Optimierung verwendete bedientheoretische Modell der Fertigung ein vereinfachtes Abbild der realen Fertigung darstellt, wird der praktische Einsatz des Verfahrens in validierende Simulationsexperimente eingebettet (siehe auch Dörrsam, Furmans und Greiling [19] und Dörrsam und Furmans [18]), die es ermöglichen, Lossplittung und Werkzeugtausch detaillierter abzubilden.

Der Gesamtablauf der monatlichen Überprüfung der Losgrößen wird in den in Abbildung 3.14 dargestellten Schritten durchgeführt:

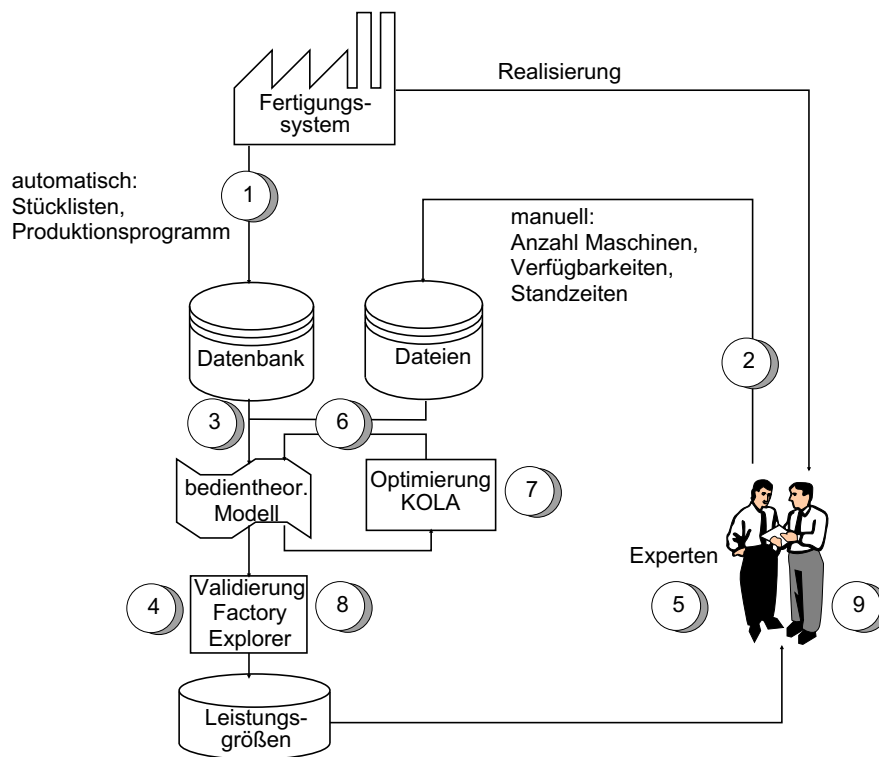


Abbildung 3.14: Ablauf des Praxiseinsatzes der Losgrößenoptimierung

1. Abfrage der aktuellen Mengenplanung und Arbeitsplanung aus den PPS-Systemen.
2. Überprüfung und ggfs. Überarbeitung der Fertigungsstrukturbeschreibung.
3. Computergestützte Erstellung des Simulationsmodells
4. Durchführung der Referenzsimulation und Ergebnisauswertung mit bisherigen Losgrößen.

5. Überprüfung der Ergebnisse durch Produktionsplaner und Meister.
6. Computergestützte Erstellung des Optimierungsmodells.
7. Durchführung des Optimierungslaufes.
8. Übertragung der Losgrößen in das Simulationsmodell und Durchführung des Simulationsexperimentes mit anschließender Ergebnisauswertung
9. Überprüfung und Freigabe der Ergebnisse durch Produktionsplaner und Meister.

Trotz der starken Vereinfachungen konnten mit Hilfe der Ergebnisse des vorgeschlagenen Optimierungsverfahrens die Durchlaufzeiten und damit auch der Umlaufbestand um 40 % gegenüber dem vorherigen Betriebszustand reduziert werden. Damit einher gingen weitere positive Effekte wie z.B. :

- Niedrigere Varianz der Durchlaufzeiten, da die erreichten Durchlaufzeiten zusätzliche manuelle Eingriffe und Priorisierungen sowie Lossplitting überflüssig machten.
- Das Ergebnis des Optimierungsverfahrens war ein Losgrößenvektor, der im Vergleich zu den vorher verwendeten Startlosgrößen zu erhöhtem Rüstaufwand geführt hätte. Durch den Wegfall der zuvor vorgenommenen ungeplanten Rüstvorgänge erhöhte sich der Rüstaufwand insgesamt jedoch nicht, so daß in der Summe eine Kostenreduzierung in der Fertigung durch Verringerung der Bestandskosten, der zur Zwischenlagerung notwendigen Fläche und des verringerten Suchaufwandes erreicht wurde.

Das Ergebnis des Piloteinsatzes führte dazu, daß auch benachbarte Fertigungsbereiche in die Optimierung miteinbezogen wurden.

Der Vorteil des bedientheoretischen Ansatzes liegt darin, daß der Datenbedarf aus den vorhandenen Produktionsplanungssystemen und vertretbarem manuellem Zusatzaufwand gedeckt werden kann. Die Optimierungsergebnisse sind robust gegenüber Abweichungen von den Planungsdaten. Aus diesem Grund genügt in diesem Anwendungsfall eine monatliche Überprüfung der Losgrößen. Die daraus resultierende Stabilität der ermittelten Losgrößen macht es möglich, die berechneten Losgrößen auch für die Beschaffung zu verwenden.

3.6.3 Weitere Anwendungsbereiche

Modellierung von Halbleiterfertigungen

Ein bedientheoretischer Analysemethoden werden häufig zur Modellierung, Analyse und Optimierung von Halbleiterfertigungen eingesetzt, da Halbleiterfertigungen sehr viele Prozessschritte umfassen und der Einfluß stochastischer Größen sehr hoch ist. Eine detaillierte Planung von Halbleiterfertigungseinrichtungen ist häufig erst sehr spät möglich, da die Produkt- und Produktionsentwicklung parallel verlaufen und viele Dimensionierungsaufgaben aufgrund von Datenabschätzungen vorgenommen werden müssen.

Weiterhin sind Kenntnisse über bedientheoretische Modelle in der Computerbranche aufgrund des erfolgreichen Einsatzes bei der Modellierung von Betriebssystemen und Netzwerken weit verbreitet, was die Akzeptanz für den Einsatz dieser Modelle im Planungsprozeß erhöht. Vor diesem Hintergrund wurde z.B. von Whitt die in Abschnitt 3.3.2 vorgestellte QNA entwickelt.

Im wesentlichen auf den gleichen Grundlagen beruht das von Connors et al. (siehe [12]) um die Berücksichtigung von Störereignissen verfeinerte Modell. Dieses Modell liegt auch dem kombinierten Simulations- und bedientheoretischen Analyse-Werkzeug „Factory Explorer“ zugrunde, das auch im Rahmen des in Abschnitt 3.6.2 dargestellten Einsatzes für die Simulationsexperimente herangezogen wurde.

Flexible Fertigungssysteme

Ein weiterer Anwendungsbereich, für den bedientheoretische Modelle verstärkt entwickelt wurden, ist die Leistungsbewertung von Planungsalternativen in der Planungsphase von flexiblen Fertigungssystemen. Einen Überblick hierüber gibt die Monographie von Tempelmeier und Kuhn [57]. Im Mittelpunkt der bedientheoretischen Modelle stehen hier geschlossene Netzwerke mit exponentialverteilten Bedienzeiten, wie sie in Abschnitt 3.4 behandelt wurden. Das Grundmodell wird dabei unter anderem um die Behandlung von Blockierungszuständen, Maschinenausfällen und mangelnder Materialverfügbarkeit ergänzt. Hierfür können häufig die besonderen Strukturen der Übergangsmatrix eines flexiblen Fertigungssystem genutzt werden (siehe auch [8]).

Die Güte der Ergebnisse wird mit entsprechenden Simulationsexperimenten nachgewiesen. *Suri* hat in [56] mit Hilfe der operationellen Analyse gezeigt, daß Mittelwerte für Durchlaufzeiten und Wartezeiten, die mit Hilfe geschlossener Bediensystemnetzwerke unter der Annahme von exponentialverteilten Bedienzeiten berechnet wurden, gegen Verletzung der Verteilungsannahme robust sind, weshalb die Verfeinerung bezüglich der Verteilungsannahmen bei geschlossenen Bediensystemnetzwerken, z.B. durch den Einsatz des Verfahrens von Marie nur in

beschränktem Zugewinn an Genauigkeit resultiert.

Zur Dimensionierung und Konfiguration von Materialfließpuffern (siehe Furmans in [23] und [24]) werden Wahrscheinlichkeiten für die Zahl der Kunden in einem Puffe benötigt. Für Zustandswahrscheinlichkeiten gelten die von Suri nachgewiesenen Robustheitseigenschaften nicht, so daß eine Pufferplatzbestimmung für geschlossene Bediensystemnetzwerke auf dem geeigneten mehrphasigen Bedienzeitmodellen und dem Verfahren von Marie basieren muß.

Eine einfache Steuerungsstrategie: CONWIP

Die explizite Abbildung von Steuerungsstrategien wurden bei den bisher vorgestellten Modellen außer acht gelassen. Eine implizite Modellierung der CONWIP-Steuerung ist mit Hilfe von geschlossenen Bediensystemen möglich.

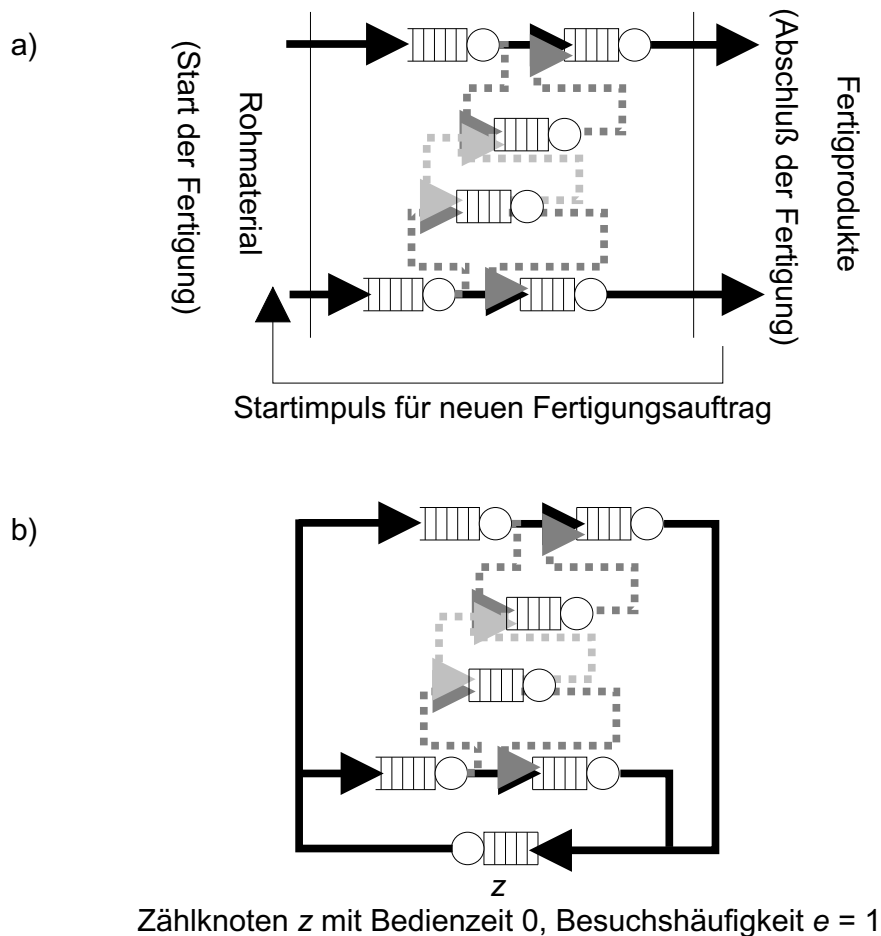


Abbildung 3.15: Fertigungssystem mit CONWIP-Steuerung (a) und äquivalentes geschlossenes Bediensystemnetzwerk (b)

CONWIP steht für ‘constant work in progress’, also für eine Steuerungsstrategie, die gleichbleibende Umlaufbestände sicherstellt. Dies wird dadurch erreicht, daß die Arbeit an einem neuen Los erst dann beginnt, wenn ein anderes Los die Fertigung nach Abschluß des letzten Bearbeitungsschrittes verlassen hat.

Physisch handelt es sich bei dem zu modellierenden Fertigungssystem um ein offenes System, logisch wird das System durch die Bedingung, daß ein neuer Auftrag nur dann in das System gelangen darf, wenn ein bereits begonnener Auftrag abgeschlossen ist, zu einem geschlossenen System.

Durch Abbildung des Fertigungssystems mit Hilfe eines geschlossenen Bediensystemnetzwerkes, bei dem die Zahl der Kunden der konstanten Zahl im System befindlicher Aufträge entspricht, wird die Einhaltung des vorgegebenen Umlaufbestandes sichergestellt. Der Abschluß der Bearbeitung eines innerhalb der Fertigung befindlichen Auftrages führt zum Start eines neuen Auftrages an einer der erstbearbeitenden Maschinen und entspricht damit dem Durchlaufen des Zählknotens z im geschlossenen Bediensystemnetzwerk (siehe auch Abbildung 3.15).

Die Parametrierung zueinander äquivalenter offener und geschlossener Bediensystemnetzwerke wird von Whitt in [64] vertieft behandelt.

Eine aufwendigere Steuerungsstrategie: Kanban

Kanban-Systeme verknüpfen den Fluß von Nachfrageinformationen mit der Steuerung des Materialflusses. Ihre formale Regelung erlaubt eine Abbildung und Untersuchung mit Hilfe bedientheoretischer Modelle. In Abbildung 3.16 ist der einfachste Fall eines Kanbansystemes dargestellt.

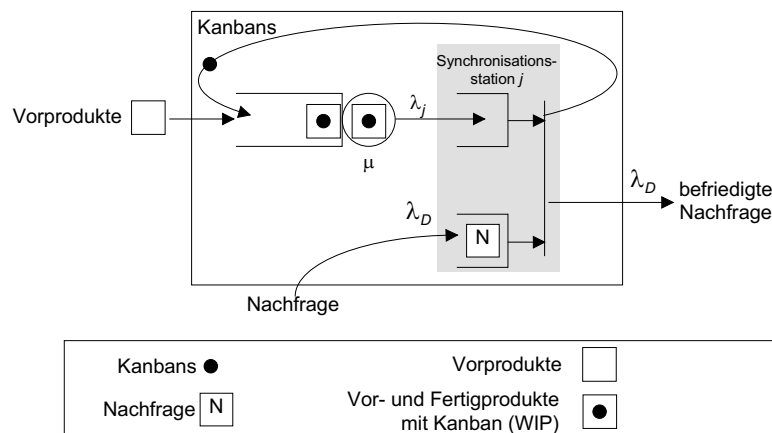


Abbildung 3.16: Einstufiges Kanban-System

Innerhalb des Systems läuft eine begrenzte Zahl von Kanbans um. Sie werden zu Beginn der Produktion zusammen mit Vorprodukten in den Warteraum des Bediensystem gebracht. In der Bedienstation werden die Vorprodukte bearbeitet

und gehen mit den Kanbans in den dem Bediensystem nachgelagerten Ausgangspuffer über.

Vorprodukte stehen ständig zur Verfügung. Die Kanbans stellen jedoch eine „Produktionsberechtigung“ dar. Stehen am Systemeingang keine freien Kanbans mehr zur Verfügung, so können keine Vorprodukte entnommen und bearbeitet werden, d.h. die Kapazität der Bedienstation liegt brach.

Trifft eine Nachfrage ein, so wird diese sofort erfüllt, wenn im Ausgangspuffer ein Bestand vorhanden ist. Das Produkt verläßt als befriedigte Nachfrage das System und der Kanban wird freigesetzt. Der freigesetzte Kanban geht zusammen mit einem neu bezogenen Vorprodukt in das Bediensystem über.

Ist im Ausgangspuffer kein Bestand vorhanden, so geht die Nachfrage nicht verloren, sondern wartet im Nachfragepuffer in nach der FCFS-Regel auf ein Fertigprodukt. Trifft dieses ein, wird ebenfalls der Kanban zurückgeleitet und die befriedigte Nachfrage verläßt das System.

Das Bediensystem ist um eine Synchronisationsstation erweitert worden, die die Eigenschaft besitzt, daß an ihr entweder Fertigprodukte mit Kanbans oder unbefriedigte Nachfragen warten.

Für die Analyse eines Kanban-gesteuerten Fertigungssystems schlagen Dallery und DiMascolo (siehe z.B. [17] [11], [46]) eine Erweiterung des Verfahrens von Marie vor. Die Grundprinzipien der Vorgehensweise sollen an einem einstufigen, Kanban-gesteuerten Fertigung dargestellt werden.

Das Basismodell von Dallery und Di Mascolo hat folgende Voraussetzungen:

- Es gibt einen Produkttyp.
- Es befindet sich eine konstante Zahl K Kanbans im System.
- Jeder Kanban ist entweder mit genau einem Teil oder einem Behälter mit einheitlicher Menge verbunden.
- Kanbans kehren sofort nach Freiwerden am Synchronisationsknoten zum Eingang zurück.
- Die Fertigung erfolgt in sequentiellen Stufen.
- Jede Fertigungsstufe wird als ein Bediensystem abgebildet.
- Vorprodukte sind unbegrenzt vorhanden.
- Jede nicht sofort befriedigte Nachfrage wird gespeichert.
- Die Zwischenankunftszeit der Nachfrage ist exponentialverteilt mit Rate λ_D .

- Die Bedienzeiten können mit Phasenverteilungen abgebildet werden.

Unter diesen Voraussetzungen ist nachstehend beschriebene Vorgehensweise zur Kennwertberechnung möglich.

Der maximale Durchsatz des Systems wird dann erreicht, wenn der Nachfragepuffer ständig besetzt ist. Der Durchsatz für dieses “ gesättigte“ System kann als Maximaldurchsatz des geschlossenen Bediensystemnetzwerkes bei kurzgeschlossener Synchronisationsstation ermittelt werden.

1. Bildung eines äquivalenten Bediensystemnetzwerkes. An Stelle des Synchronisationsknotens tritt eine Bediensystem, welches in der Folge ebenfalls mit lastabhängigen Bedien- und Ankunftsdaten analysiert wird.
2. Die Bearbeitungsstationen werden in der Mikro-Sicht als $\lambda(k)|C(n)|1$ -Bediensysteme abgebildet und analysiert.
3. Für den Synchronisationsknoten (in obigem Beispiel j) werden durch Analyse seines Markovprozesses Zustandswahrscheinlichkeiten und die daraus resultierenden lastabhängigen Bedienraten $\mu(k)$ ermittelt

Wird der Synchronisationsknoten j als Einzelknoten betrachtet, so ist sein Zustand durch einen Vektor \vec{k}_j beschreibbar. Das erste Element $k_{j,1}$ des Vektors beschreibt die Zahl der Fertigprodukte im Synchronisationsknoten, das zweite Element $k_{j,2}$ die Anzahl wartender Nachfragen. $k_{j,1}$ und $k_{j,2}$ können nicht gleichzeitig größer als Null sein (siehe Abbildung 3.17).

Die zustandsabhängigen Bedienraten des Synchronisationsknoten werden mit nachstehenden Ausdrücken bestimmt:

$$\begin{aligned}
 \mu_j(0) &= 0 \\
 \mu_j(1) &= \lambda_D(0) \frac{p(0,0)}{p(1,0)} = \lambda_D \frac{p(0,0)}{p(1,0)} \\
 \mu_j(2) &= \lambda_D(1) \frac{p(1,0)}{p(2,0)} = \lambda_D \frac{p(1,0)}{p(2,0)} \\
 \mu_j(k) &= \lambda_D(k) \frac{p(k-1,0)}{p(k,0)} = \lambda_D \frac{p(k-1,0)}{p(k,0)} \quad (3.74)
 \end{aligned}$$

4. Mit den aus der Analyse des Synchronisationsknoten erzielten Ergebnissen $\mu(k)$ kann eine erneute Kennwertberechnung für das Gesamtsystem durchgeführt werden, bis Konvergenz eintritt.

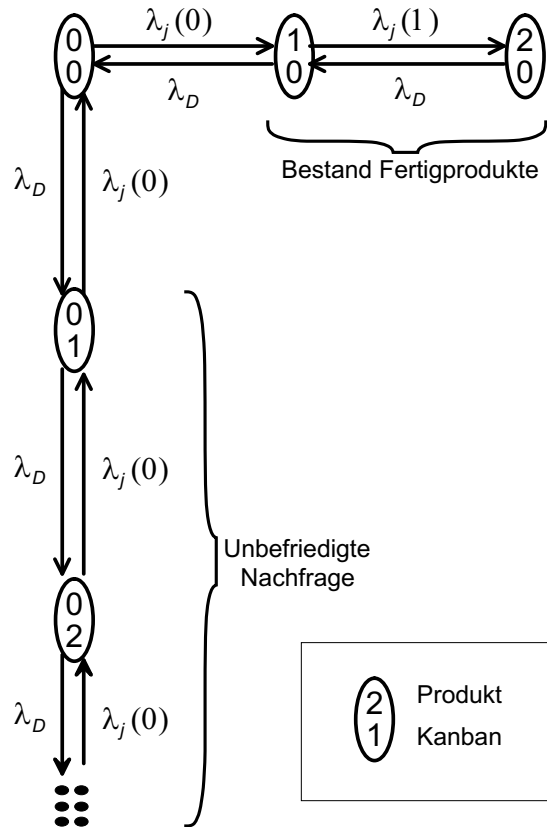


Abbildung 3.17: Zustandsraum des Synchronisationsknotens

Der Grenzdurchsatz eines Kanban-gesteuerten Materialflusssystems wird erreicht, wenn jedes Fertigprodukt sofort auf eine Nachfrage trifft, so daß der assoziierte Kanban umgehend mit neuem Rohmaterial zum Beginn der Fertigung zurückkehrt. In diesem Fall führt der Synchronisationsknoten zu keiner Verzögerung des Kanban-Umlaufes.

Zur Ermittlung des Grenzdurchsatzes wird deshalb der Synchronisationsknoten kurzgeschlossen. Anschließend erfolgt eine Berechnung des Sättigungsdurchsatzes mit Hilfe der für geschlossene Bediensysteme beschriebenen Vorgehensweise und Ausdruck 3.50.

Für ein einstufiges System können unter der Annahme exponentialverteilter Zwischenankunftszeiten der Nachfrage die nachfolgend angegebenen Kennwerte mit geschlossenen Ausdrücken berechnet werden.

Durchschnittlicher Fertigungsbestand: Mittelwert der Zahl von Kunden in den Bediensystemen und im Puffer für Fertigprodukte der Synchronisationsstation.

Anteil nicht sofort befriedigter Nachfragen: Da wegen der exponentialverteilt-

ten Zwischenankunftszeiten das PASTA-Theorem (zur Anwendung siehe Abschnitt 2.9 , Herleitung siehe Wolff [66] Seite 293 ff.) gilt, ist der Anteil nicht sofort befriedigter Nachfrage p_B identisch mit der Wahrscheinlichkeit, den Puffer für Fertigprodukte zu einem beliebigen Zeitpunkt leer anzutreffen:

$$p_B = \sum_{k_2=0}^{\infty} p(k_1, k_2)^T \quad (3.75)$$

Durchschnittliche Zahl wartender Nachfragen:

$$N_{B,w} = p_B \frac{1}{\frac{\lambda_j(0)}{\lambda} - 1} \quad (3.76)$$

Mittlere Wartezeit der Nachfragen:

$$T_{D,w} = \frac{N_{B,w}}{\lambda} \quad (3.77)$$

Mittlere Wartezeit der nicht sofort bedienten Nachfragen:

$$T_{B,w} = \frac{N_{B,w}}{p_B \lambda} \quad (3.78)$$

Ein Erweiterung der vorstehend beschriebenen Vorgehensweise auf die Modellierung von mehrstufigen Kanban Systemen mit mehreren Produkten durch Duri, Dallery und DiMascolo ist in [20] zu finden, eine Anpassung auf sogenannte „Base stock Systeme“ in [45]. Ein Vergleich zwischen Kanban und CONWIP-System findet sich bei [33].

Shantikumar und Buzacott stellen in [9], S. 490 ff. das PAC-Modell vor, mit dem verschiedene Modelle der Verknüpfung von Informations- und Materialfluß abgebildet werden können. Für Kanban- und Base-stock Systeme sind diese Modelle mit denen von Dallery, Di Mascolo und Frein (s.o.) zu vergleichen, wobei mit letztgenannten Ansätzen Kennwerte für einen größeren Anwendungsbereich berechnen werden können.

4 Zusammenfassung und Ausblick

Die in den Kapiteln 2 und 3 beschriebenen Modelle und Berechnungsverfahren bilden eine Grundlage für die weitere Beschäftigung mit bedientheoretischen Modellen und deren Anwendung bei der Analyse und Planung von Materialflusssystemen.

Ausgangspunkt ist die Operationellen Analyse, die ausgehend von der Beobachtung eines Bediensystems über einen begrenzten Zeitraum, Aussagen über Leistungskenngrößen ermöglicht. Die Operationelle Analyse kann bei der Auswertung von Simulationsexperimenten und von Beobachtungen realer Systeme genutzt werden, um die Beobachtungsdaten denjenigen Ergebnissen gegenüberzustellen, die mit äquivalenten stochastischen Modellen ermittelt wurden.

Für die Vorhersage von Leistungsdaten sind die Ansätze der operationellen Analyse jedoch nur eingeschränkt geeignet, weshalb nachfolgend stochastische Modelle eingeführt werden.

Mit dem Übergang auf das stochastische Basismodell des $M|M|1$ -Bediensystem in dem Bedien- und Zwischenankunftszeiten ausschließlich durch Exponentialverteilungen beschrieben werden, ist es möglich, sowohl Zustandswahrscheinlichkeiten als auch Mittelwerte von Leistungskenngrößen des Systems zu bestimmen. Es zeigt sich, daß die Annahme exponentialverteilter Bedien- und Zwischenankunftszeiten zu den gleichen Ergebnissen führt, wie die Annahme zustandsunabhängiger Ankunfts- und Bedienraten bei der operationellen Analyse. Insbesondere die Bedienzeiten einzelner Bausteine von Materialflusssystemen können jedoch selten mit Hilfe von Exponentialverteilungen beschrieben werden.

Eine genauere Abbildung praxisrelevanter Bedienzeitverteilungen ist mit Hilfe der Phasenverteilungen möglich, die durch Kombination mehrerer Exponentialverteilungen entstehen. Neben der Erlangverteilung wurde die Cox-Verteilung vorgestellt, da diese für die Darstellung nicht-exponentialverteilter Bedienzeiten in geschlossenen Netzwerken von Bedeutung ist.

Soll die Bedienzeit eines Materialflußelementes lediglich mit Mittelwert und Varianz beschrieben werden, so können Modelle von $M|G|1$ -Bediensystemen ein-

gesetzt werden. Es wurde gezeigt, daß für diese Modelle die Mittelwerte von Leistungskenngrößen wie Wartezeit und mittlere Anzahl von Kunden im Bediensystem exakt bestimmt werden können. Damit sind die ersten Ergebnisse erzielt, die zur Analyse von Materialflußelementen genutzt werden können.

Littles Gesetz erlaubt es, einen Zusammenhang zwischen Verweilzeit eines Kunden im System, der Anzahl von Kunden im System und der Ankunftsrate herzustellen.

Die durch Littles Gesetz beschriebenen Zusammenhänge gelten jedoch nicht nur für $M|G|1$ -Bediensysteme, sondern haben für fast alle bedientheoretischen Modelle Geltung, die zur Abbildung von Materialflußsystemen verwendet werden.

Die Bedeutung der $M|G|1$ -Bediensysteme und von Littles Gesetz wird durch die Anwendung des Modells auf Fragestellungen der Losgrößenplanung gezeigt.

Aufbauend auf den $M|G|1$ -Bediensystemen können die Auswirkungen von Prioritäten auf verschiedene Kundenklassen untersucht werden. Mit $M|G|1$ -Prioritätssystemen können Zusammenführungselemente mit verschiedenen Vorfahrtsstrategien und Störungen abgebildet und untersucht werden.

Für Mehrkanalsysteme mit generellen Bedienzeiten sowie für Ein- und Mehrkanalbediensysteme bei denen Zwischenankunfts- und Bedienzeiten nur mit Hilfe von Mittelwert und Varianz beschrieben sind, können lediglich Näherungswerte und Obergrenzen angegeben werden. Diese bilden den Abschluß des Kapitels 2.

Aufbauend auf den Modellen für Einzelbediensysteme werden in Kapitel 3 die Grundlagen für die Untersuchung von Bediensystemnetzwerken gelegt. Sowohl offene als auch geschlossene Bediensystemnetzwerke werden modelliert. Bei beiden Klassen von Netzwerken steht die Betrachtung von Netzwerken am Anfang, deren Bedienzeiten und (für offene Netzwerke) Zwischenankunftszeiten Exponentialverteilungen gehorchen.

Für diese Netzwerke wird die Bedeutung, der Gültigkeitsbereich und der Aufbau der Produktformlösungen dargestellt. Während für offene Bediensystemnetzwerke die direkte Berechnung von Leistungskenngrößen für das Gesamtsystem und die Einzelbediensysteme durch Dekomposition möglich ist, ist die Produktformlösung für die Kennwertberechnung für geschlossene Bediensystemnetzwerke schon bei kleinen Problemgrößen nicht mehr einsetzbar. Repräsentativ für die Gruppe der effizienteren Berechnungsverfahren für geschlossene Bediensystemnetzwerke wird die Mittelwertanalyse vorgestellt.

Für offene Bediensysteme deren Bedien- und Zwischenankunftszeiten mit Mittelwert und Varianz beschrieben werden, wird das Berechnungsverfahren QNA dargestellt. Durch eine Approximation der gegebenen Punktprozesse durch Erneuerungsprozesse und Lösung von Gleichungssystemen für Flüsse und Variabilitäten wird eine Approximation der Kennwerte berechnet.

Für geschlossene Netzwerke werden aufbauend auf der Mittelwertanalyse die Aus-

wirkungen von Bedienzeiten, die von der Exponentialverteilung abweichen, durch zwei Näherungsverfahren abgeschätzt.

Das einfache Verfahren der Antwortzeiterhaltung wird zuerst erläutert, nach Diskussion der Schwachpunkte des Näherungsverfahrens wird als aufwendigere aber auch genauere Approximation das Verfahren von Marie dargestellt. Dieses Verfahren berücksichtigt zustandsabhängige Ankunftsdaten und verwendet dazu die bereits im Kapitel 2 eingeführten Bediensysteme mit Cox-Verteilungen und zustandsabhängigen Ankunftsdaten.

Zum Abschluß des Kapitels 3 werden die Ansätze zur Losgrößenoptimierung für $M|G|1$ -Bediensysteme auf Bediensystemnetzwerke übertragen.

Die vorgestellten bedientheoretischen Modelle und Methoden bilden nur einen kleinen Ausschnitt aus der Vielzahl veröffentlichter Modelle und Berechnungsverfahren ab. Sie bilden jedoch eine Grundlage, um Materialflußsysteme in ihren Grundzügen abbilden zu können. Weiterhin können auf ihnen aufbauende Modelle für speziellere Fragestellungen verstanden und eingeordnet werden.

Die Anwendbarkeit der vorgestellten Methoden ist häufig durch die mit ihnen erzielbare Genauigkeit der quantitativen Aussagen gegeben. Die Berechnung von Zustandswahrscheinlichkeiten ist oft durch restriktive Annahmen hinsichtlich der Verteilung von Bedien- und Zwischenankunftszeiten eingeschränkt.

Allen vorgestellten Modellen gemeinsam ist die Annahme, daß die Zeit kontinuierlich abgebildet wird, d.h. Zeiten können mit Ausnahme von Dirac-Verteilungen grundsätzlich jeden realen Wert annehmen.

Mit Hilfe von zeitdiskreten Modellen, wie sie bei Tran-Gia in [58] eingeführt werden, können neben Mittelwerten auch Verteilungen von Kenngrößen bestimmt werden. Die Anwendung von zeitdiskreten Berechnungsverfahren auf Bediensystemnetzwerke zur Modellierung von Materialflußsystemen ist bis heute die Ausnahme.

Für die Materialflußplanung wichtig ist die Möglichkeit mit Hilfe zeitdiskreter Modelle die Verteilung der Anzahl Kunden im System oder die Verteilung der Wartezeiten zu ermitteln. Mit diesen Angaben können Puffer in Materialflußsystemen dimensioniert werden.

Ein Beispiel hierfür zeigt die Modellierung von Montagelinien mit Hilfe zeitdiskreter Verfahren durch den Autor in [25] und die Abbildung von Materialflußpufferkonfigurationen durch Furmans und Zillus in [27].

Aufgrund des vorhandenen großen Potentials zeitdiskreter, bedientheoretischer Modelle können in der Anwendung dieser Verfahren auf Fragestellungen des Materialflusses noch erhebliche Fortschritte erzielt werden. Hierfür muß jedoch noch die Brücke zwischen Theorie und Anwendung geschlagen werden.

Symbolverzeichnis

c_{an}	Variationskoeffizient des Ankunftsprozesses
c_{an}^2	Variabilität des Ankunftsprozesses
c_{ab}	Variationskoeffizient des Abfertigungsprozesses
c_{ab}^2	Variabilität des Abfertigungsprozesses
d_h	Nachfragemenge des Produktes h pro Periode
k	Zustand \equiv Anzahl Kunden in einem Bediensystem
\vec{k}	Zustandsvektor eines Bediensystemnetzwerkes
k_i	Anzahl Kunden am Bediensystem i eines Bediensystemnetzwerkes
k_f	Kostensatz für Fertigungsbestandsabhängige Kosten eines Loses
k_l	Lagerkostensatz pro Los
k_r	Rüstkostensatz
K	Anzahl Kunden in einem geschlossenen Bediensystemnetzwerk
K_f	Summe aller vom Fertigungsbestand abhängigen Kosten
K_l	Summe aller Lagerkosten
K_r	Summe aller Rüstkosten
l_h	Losgröße des Produktes h
m	Anzahl Bedienstationen eines Bediensystems
m_i	Anzahl Bedienstationen des Bediensystems i
M	Anzahl Bediensysteme in einem Bediensystemnetzwerk
n	Anzahl Phasen in mehrphasigen Verteilungen
N	Mittlere Anzahl Kunden im Bediensystem
$N(t)$	Mittlere Anzahl Kunden im Bediensystem bis zum Zeitpunkt t
N	Mittlere Anzahl Kunden im Bediensystem i
N_w	Mittlere Anzahl wartender Kunden am Bediensystem
p_h	Anteil der Lose des Produktes h an der Gesamtzahl vpn Produktionslosen

$p(k)$	Wahrscheinlichkeit, daß sich das Bediensystemnetzwerk im Zustand k befindet
$p(k_i)$	Wahrscheinlichkeit, daß sich das Bediensystem i im Zustand k befindet
P	Anzahl verschiedener Produkte
$t_{r,h}$	Rüstzeit für Produkt h
$t_{b,h}$	Stückbearbeitungszeit für Produkt h
$t_{an,i}$	Mittlere Zwischenankunftszeit an Bediensystem i (im Kontext von Bediensystemnetzwerken), oder von Kunden der Prioritätsklasse i (im Kontext von Bediensystemnetzwerken mit Prioritäten)
$t_{ab,i}$	Mittlere Bedienzeit an Bediensystem i (im Kontext von Bediensystemnetzwerken), oder von Kunden der Prioritätsklasse i (im Kontext von Bediensystemnetzwerken mit Prioritäten)
$T_{ab,n}$	Bedienzeit des n -ten Kunden
$T_{an,n}$	Zwischenankunftszeit zwischen n -tem Kunden und Kunde $n + 1$
$T_{d,n}$	Zeit zwischen dem Abgang des Kunden n und dem Kunden $n + 1$
X_n	Differenz zwischen hinzukommender und geleisteter Arbeit
Y_n	Dauer der Brachzeitperiode nach Abgang des Kunden n
y_i	Wahrscheinlichkeit für eine Brachzeitperiode von $i \cdot t_{inc}$
λ_i	Ankunftsrate am Bediensystem i
$\lambda_i(k_i)$	Ankunftsrate eines Bediensystems im Zustand k
$\lambda_{i,j}$	Fluß vom Bediensystem i zum Bediensystem j
λ_0	Rate der externen Ankünfte eines Bediensystemnetzwerkes
λ_h	Ankunftsrate des Produktes h
μ_i	Bedienrate des Bediensystems i
ρ_i	Auslastungsgrad des Bediensystems i
τ_n	Ankunftszeitpunkt des Kunden n am Bediensystem

Literaturverzeichnis

- [1] Subhash C. Agrawal, Jeffrey P. Buzen und Annie W. Shum. Response Time Preservation: A General Technique for Developing Approximate Algorithms for Queueing Networks. *ACM Sigmetrics Performance Evaluation Review*, 12(3), August 1984.
- [2] Susan L. Albin. Approximating a Point Process by a Renewal Process, II: Superposition Arrival Processes to Queues. *Operation Research*, 32:1133–1162, 1984. Rutgers University, Piscataway, New Jersey.
- [3] Dieter Arnold. *Materialflußlehre*. Vieweg Verlag, Braunschweig, 1995.
- [4] Forest Baskett, K. Mani Chandy, Richard R. Muntz und Fernando Palacios. Open, Closed and Mixed Networks of Queues with Different Classes of Customers. *Journal of the Association for Computing Machinery*, 22(2):248–260, April 1975.
- [5] Gunter Bolch. *Leistungsbewertung von Rechensystemen mittels analytischer Warteschlangenmodelle*. Teubner, Stuttgart, 1989.
- [6] Gunter Bolch und Matthias Kirschnick. PEPSY-QNS. Performance Evaluation and Prediction System for Queueing Networks. Forschungsbericht, TR-I4-21-92, Institut für mathematische Maschinen und Datenverarbeitung an der Friedrich-Alexander-Universität in Erlangen-Nürnberg, 1992. Überarbeitete Auflage.
- [7] P.J. Burke. The Output of Queueing Systems. *Operations Research*, 4:699–704, 1956.
- [8] J.A. Buzacott und David D. Yao. Flexible Manufacturing Systems: A Review of Analytical Models. *Management Science*, 32(7):890–905, Juli 1986.
- [9] John A. Buzacott und J. George Shantikumar. *Stochastic Models of Manufacturing Systems*. Prentice-Hall, Englewood Cliffs, 1993.
- [10] Jeffrey P. Buzen. Computational Algorithms for Closed Queueing Networks with Exponential Servers. *Communications of the ACM*, 16(9):527–531, 1973.

- [11] Christian Commault, Maria Di Mascolo und Yannick Frein. Un outil de mise en Oeuvre du juste a Temps, le Kanban. Forschungsbericht, Laboratoire d'Automatique de Grenoble, ENSIEG, B.P. 46, 38402 Saint-Martin- d'Hères, France, Januar 1991.
- [12] Dan P. Connors, Gerald D. Feigin und David D. Yao. A Queueing Network Model for Semiconductor Manufacturing. *IEEE Transactions on Semiconductor Manufacturing*, 9:412 – 427, 1996.
- [13] D.J. Daley. Inequalities for Moments of Tails of Random Variables, with Queueing Applications. *Z. Wahrsch.*, 41:139–143, 1977.
- [14] Yves Dallery und Xi-Ren Cao. Operational analysis of stochastic closed queueing networks. *Performance Evaluation*, 14:43 –61, 1992.
- [15] Peter Denning und Jeffrey P. Buzen. Operational Analysis of Queueing Networks. In H. Beilner und E. Gelembé, Herausgeber, *Measuring, Modelling and Evaluating Computer Systems*, Seiten 151–173. North-Holland Publishing Company, 1977.
- [16] Peter Denning und Jeffrey P. Buzen. The Operational Analysis of Queueing Network Models. *Computing Surveys*, 10(3):225–261, September 1978.
- [17] Maria DiMascolo, Yannick Frein und Yves Dallery. An Analytical Method for Performance Evaluation of Kanban Controlled Production Systems. Forschungsbericht, Laboratoire d' Automatique de Grenoble, August 1993. Laboratoire MASI (URA 818 CNRS) und Institute Blaise Pascal, Université Pierre et Marie Curie, 4, place Jussieu, 75252 paris cedex 05.
- [18] Volker Dörrsam und Kai Furmans. Application Case Study of a Queueing Network Simulation Tool for Analyzing and Optimizing a Manufacturing System. In *Proceedings of the ESS '95*, Seiten 669 – 673. Society of Computer Simulations, 1995.
- [19] Volker Dörrsam, Kai Furmans und Markus Greiling. Reducing WIP in Gear Parts Manufacturing Using Queueing Networks. In *Proceedings of CESA 96*, 1996.
- [20] C. Duri, C. Frein und M. Di Mascolo. Performance evaluation of kanban multiple-product production systems. In *INRIA/IEEE Conference on Emerging Technologies and Factory Automation, ETFA '95*, Oktober 1995.
- [21] William Feller. *An Introduction to Probability Theory and its Applications*, volume 1. Wiley International, New York, dritte Auflage, 1968.
- [22] H. Fromm. Das Management von Zeit und Variabilität in Geschäftsprozessen. *CIM Management*, 5:7–14, 1992.

-
- [23] Kai Furmans. *Ein Beitrag zur theoretischen Behandlung von Materialflupuffern in Bediensystemnetzwerken*. Doktorarbeit, Universität Karlsruhe, Fakultät für Maschinenbau, Institut für Fördertechnik, 1992.
- [24] Kai Furmans. An Open Queueing Network Model of Manufacturing Systems with Independent Production Buffers. In W. Krug und A. Lehmann, Herausgeber, *Proceedings of the 1992 European Simulation Symposium, Dresden*, Seiten 560 – 565, San Diego, November 1992. Society for Computer Simulation.
- [25] Kai Furmans. A Discrete Time Model of a Car Assembly System. In *INRIA/IEEE Conference on Emerging Technologies and Factory Automation, ETFA'95*, Oktober 1995.
- [26] Kai Furmans. Verfahren zur Grobplanung von Materialflußsystemen. In H. Grabowski und G. Zülch, Herausgeber, *Proceedings zur Workshopreihe Innovative Produktentwicklung und Produktsystemplanung*, Seiten 331 – 344. Veröffentlichungen des Sonderforschungsbereiches 346, Universität Karlsruhe (TH), Bd. 1, 1995.
- [27] Kai Furmans und Andrea Zillus. Modeling Independent Production Buffers in Discrete Time Queueing Networks. In *Proceedings of Rensselaer's Fifth International Conference Computer Integrated Manufacturing and Automation Technology*, Seiten 275–280, 1996.
- [28] William J. Gordon und Gordon F. Newell. Closed Queueing Systems With Exponential Servers. *Operations Research*, 15(2):254–265, April 1967.
- [29] Winfried K. Grassmann und Joti L. Jain. Numerical Solutions of the Waiting Time Distribution and Idle Time Distribution of the Arithmetic $GI|G|1$ Queue. *Operations Research*, 37(1):141–150, Januar–Februar 1989.
- [30] Markus Greiling. *Optimierung der Produktionslogistik durch Integration mittelfristiger Materialflußplanung - Ein losgrößenorientierter Ansatz*. Doktorarbeit, Universität Karlsruhe, Fakultät für Maschinenbau, Institut für Fördertechnik und Logistiksysteme, Dezember 1997.
- [31] Werner Großeschallau. *Materialflußrechnung*. Springer-Verlag, Berlin ; Heidelberg, 1984.
- [32] Donald Gross und Carl Harris. *Fundamentals of Queueing Theory*. John Wiley Sons, New York, zweite Auflage, 1985.

- [33] Stefan Gstettner und Heinrich Kuhn. Analysis of Production Control Systems Kanban and CONWIP. Forschungsbericht, Universität zu Köln, Seminar für allgemeine Betriebswirtschaftslehre, Industriebetriebslehre und Produktionswirtschaft, Univ.-Prof.Dr.Horst Tempelmeier, Albertus-Magnus-Platz, D-50923 Köln, Mai 1995.
- [34] T. Gudehus. Staueffekte von Transportknoten. *Zeitschrift für Operations Research*, 20:209–252, 1976.
- [35] Hans-Otto Günther und Horst Tempelmeier. *Produktion und Logistik*. Springer-Verlag, Berlin ; Heidelberg, 2. Auflage, 1995.
- [36] Frederick S. Hillier und Gerald J. Lieberman. *Introduction to Operations Research*. Holden Day, Oakland, Ca., 4. Auflage, 1986.
- [37] James R. Jackson. Jobshop-Like Queueing Systems. *Management Sciences*, 10(1):131–142, Oktober 1963.
- [38] Joti L. Jain und Winfried K. Grassmann. Numerical Solution for the Departure Process from the $GI|G|1$ Queue. *Comput. Opns. Res.*, 15(3):293–296, 1988.
- [39] Peter J. B. King. *Computer and Communication Systems Performance Modelling*. International Series in Computer Science. Prentice Hall, 1990.
- [40] Leonard Kleinrock. *Queueing Systems, Volume 1: Theory*. Wiley, New York, 1975.
- [41] Leonard Kleinrock. *Queueing Systems, Volume 2: Computer Applications*. Wiley, New York, 1976.
- [42] W. Krämer und M. Langenbach-Belz. Approximate Formula for General Single Server Systems with Single and Batch Arrivals. Forschungsbericht, Institut für Nachrichtenvermittlung und Datenverarbeitung, Universität Stuttgart, Oktober 1977. Seidenstraße 36, 7000 Stuttgart 1.
- [43] Paul J. Kühn. Approximate Analysis of General Queueing Networks by Decomposition. *IEEE Transactions on Communications*, 27(1):113–126, January 1979.
- [44] S.S. Lavenberg und M. Reiser. Stationary State probabilities at Arrival Instants for Closed queueing Networks with Multiple Types of Customers. *J.Appl.Prob.*, 17:1048–1061, 1982.

-
- [45] George Liberopoulos und Yves Dallery. Extended Kanban Control System: A New Kanban-Type Pull Control Mechanism for Multi-Stage Manufacturing Systems. Forschungsbericht, Institute Blaise Pascal, Dezember 1995. Laboratoire MASI (URA 818 CNRS), Institute Blaise Pascal, Université Pierre et Marie Curie , 4, place Jussieu , 75252 paris cedex 05, France.
- [46] George Liberopoulos und Yves Dallery. On the Optimization of Single-Stage Kanban-Type Control Systems in Manufacturing. Forschungsbericht, Institute Blaise Pascal, Dezember 1995. Laboratoire MASI (URA 818 CNRS), Institute Blaise Pascal , Université Pierre et Marie Curie, 4, place Jussieu , 75252 paris cedex 05, France.
- [47] John D. Little. A Proof for the Queueing Formula: $L=\lambda W$. Forschungsbericht, Case Institute of Technology, Cleveland, Ohio, November 1960.
- [48] W.G. Marchal. An Approximate Formula for Waiting Time in Single Server Queues. *AIIE Transactions*, 8:473 ff, 1976.
- [49] Raymond Marie. An Approximate Analytical Method for General Queueing Networks. *IEEE Transactions on Software Engineering*, 2(4):530–538, September 1979.
- [50] Raymond Marie. Calculating Equilibrium Probabilities for $\lambda(n)|C_k|l|N$ Queues. *ACM*, 1980.
- [51] Raymond Marie und William J. Steward. A Hybrid Iterative-Numerical Method for The Solution of a General Queueing Network. In H.Beilner und E.Gelenbe, Herausgeber, *Measuring, Modelling and Evaluating Computer Systems*, Seiten 173–187. North-Holland Publishing Company, 1977. IRISA, Universit de Rennes, 35031 Rennes France.
- [52] D. Neuse und K. Chandy. SCAT: A Heuristic Algorithm for Queueing Network Models of Computing Systems. *ACM*, Seiten 59–79, 1981.
- [53] M. Reiser und S.S. Laverberg. Mean-Value Analysis of Closed Multichain Queueing Networks. *Journal of the Association for Computing Machinery*, 27(2):313–322, April 1980.
- [54] Charles H. Sauer und K. Mani Chandy. *Computer Systems Performance Modeling*. Prentice-Hall, Englewood Cliffs, 1981.
- [55] K.C. Sevcik und I. Mitrani. The Distribution of Queueing Network States at Input and Output Instants. *Journal of the Association of Computing Machinery*, 28(2):358–371, April 1981.
- [56] Rajan Suri. Robustness of Queueing Network Formulas. *Journal of the Association for Computing Machinery*, 30(3):564–594, Juli 1983.

- [57] Horst Tempelmeier und Heinrich Kuhn. *Flexible Manufacturing Systems, Decision Support for Design and Operation*. Wiley Series in Systems Engineering. John Wiley, New York, 1993.
- [58] Phuoc Tran-Gia. *Analytische Leistungsbewertung verteilter Systeme*. Springer, Berlin, 1996.
- [59] Günter Bamberg und Franz Baur. *Statistik*. Oldenbourg, München, 1982.
- [60] Ward Whitt. Approximating a Point Process by a Renewal Process, I: Two Basic Methods. *Operations Research*, 30(1):125–147, Januar–February 1982.
- [61] Ward Whitt. Performance of the Queueing Network Analyzer. *Bell Systems Technical Journal*, 62:2817–2843, 1983.
- [62] Ward Whitt. The Queueing Network Analyzer. *Bell Systems Technical Journal*, 62:2779–2815, 1983.
- [63] Ward Whitt. Approximations for Departure Processes and Queues in Series. *Naval Research Logistics Quarterly*, 31:499–521, 1984.
- [64] Ward Whitt. Open and Closed Models for Networks of Queues. *AT&T Bell Laboratories Technical Journal*, 63(9):1911–1979, 1984.
- [65] Hans-Peter Wiendahl und Peter Nyhuis. Durchlauforientierte Losbildung. *VDI-Z*, 132(5):39–48, Mai 1990.
- [66] Ronald W. Wolff. *Stochastic Modeling and the Theory of Queues*. Prentice-Hall, Englewood Cliffs, 1989.