

Verkehrstheoretische Modelle und Stochastische Simulationstechniken zur Leistungsanalyse von Kommunikationsnetzen

Der Fakultät für Elektrotechnik
der Rheinisch-Westfälischen Technischen Hochschule Aachen
vorgelegte Habilitationsschrift
zur Erlangung der Lehrbefugnis für
"Kommunikationsnetze"

vorgelegt von
Dr. rer. nat. Carmelita Görg
aus Göttingen

Referent: Univ.-Prof. Dr.-Ing. B. Walke
Korreferent: Univ.-Prof. Dr. F. Hoßfeld
Korreferent: Univ.-Prof. Dr.-Ing. P. Kühn
Tag des Habilitationskolloquiums: 24. Juni 1997

AACHENER BEITRÄGE ZUR MOBIL- UND TELEKOMMUNIKATION

Herausgeber:

Universitätsprofessor Dr.-Ing. Bernhard Walke

Görg, Carmelita:
Verkehrstheoretische Modelle und stochastische Simulationstechniken zur
Leistungsanalyse von Kommunikationsnetzen
1. Auflage Aachen:
Verlag der Augustinus Buchhandlung, 1997
(Aachener Beiträge zur Mobil- und Telekommunikation, Band 13)

ISBN 3-86073-622-1

© 1997 Carmelita Görg

Verlag der Augustinus Buchhandlung
Pontstraße 66/68
52062 Aachen

Druck: Druckerei Paletti, Aachen
Gedruckt auf chlorfrei gebleichtem Papier

Zusammenfassung

Ausgehend von aktuellen Entwicklungen im Bereich Kommunikationsnetze wird die Notwendigkeit der Modellierung und Leistungsbewertung bei der Entwicklung neuer Netze und Dienste aufgezeigt.

Zur Einführung wird im ersten Kapitel eine Übersicht über die Methoden der Leistungsbewertung, die zugehörige Modellbildung sowie über Kommunikationsnetze gegeben. Im zweiten Kapitel werden Simulationssysteme allgemein vorgestellt und am Beispiel einer objektorientierten Simulationsbibliothek vertieft. Verkehrsmodelle und deren Beschreibung durch Zufallsprozesse sind Gegenstand des dritten Kapitels. Im vierten und fünften Kapitel werden Zufallsgeneratoren und statistische Auswertemethoden vorgestellt, wie sie für die stochastische Simulation benötigt werden.

Die Anwendung der beschriebenen Techniken wird im sechsten Kapitel an ausgewählten Beispielen verschiedener Netztypen vorgestellt. Dazu wurden das Hochgeschwindigkeitsnetz FDDI, ein Telekommunikationsnetz auf der Basis des Intelligenten Netzes (IN) zur Untersuchung neuer Dienste und Beispiele für Mobilfunknetze ausgewählt.

Im anschließenden siebten Kapitel werden allgemein Verfahren zur Simulationsbeschleunigung beschrieben, um dann im achten Kapitel eine neue Methode, das RESTART/LRE-Verfahren, zur Behandlung seltener Ereignisse zu vertiefen. Der RESTART/LRE-Algorithmus wurde in dieser Arbeit als Mehrstufenverfahren weiterentwickelt und als Teil eines Simulationssystems implementiert. Die Leistungsfähigkeit des Verfahrens wurde durch ausführliche Simulationen verschiedener Modelle nachgewiesen. Das System M/M/1/N sowie ein System mit korreliertem Eingabestrom SSMP(2)/D/1/N ermöglichen die Validierung der Methode durch den Vergleich mit analytisch berechneten Werten. Als Zielgröße wird dabei die komplementäre Verteilungsfunktion $G(x)$ der Ankunftsbelegung untersucht, die als Endwert die Verlustwahrscheinlichkeit P_B liefert. Hierbei konnte durch Simulation die Verlustwahrscheinlichkeit von Zellen in ATM-Netzen in der Größenordnung $P_B = 10^{-9}$ mit guter Genauigkeit in einigen Minuten bzw. wenigen Stunden durchgeführt werden. Eine Simulation dieser Beispiele ohne Benutzung des hier vorgestellten Verfahrens hätte mehrere Jahre gedauert. Mit diesem Verfahren können auch noch sehr viel seltenere Ereignisse simulativ nachgewiesen werden, dies wird an weiteren Beispielen gezeigt, z. B. mit $P_B = 10^{-20}$. Am Ende des achten Kapitels wird die Erweiterung der Methode für Warternetze vorgestellt und an Beispielen für Modelle von ATM-Referenzverbindungen validiert.

Der Wert der Arbeit liegt einerseits im Überblick über die Modellierung und Leistungsanalyse von Kommunikationsnetzen und andererseits in der vertieften Behandlung der Beschleunigungsmöglichkeiten für die Simulation seltener Ereignisse.

Abstract

In view of current developments in the field of communication networks the necessity of modelling and performance analysis of new network and service developments is shown.

As an introduction an overview of performance evaluation methods, modelling techniques, and communication networks is given in the first chapter. The second chapter introduces simulation systems and describes an example of an object-oriented simulation library. Traffic models and more specifically their description by random processes are the main subject of the third chapter. In the fourth and fifth chapter random generators and statistical evaluation methods are presented for the use in stochastic simulation systems.

The application of the presented methods is shown in the sixth chapter for examples of different network types. Here the high-speed network FDDI, a telecommunication network based on the intelligent network (IN) for new services, and examples of mobile networks were chosen.

In the subsequent seventh chapter methods for investigating rare events by simulation are surveyed and in chapter eight a new method, the RESTART/LRE-method, is presented. The RESTART/LRE-algorithm has been extended to multiple stages and implemented as part of a simulation system. The performance of this method has been verified by extensive simulations of several models. The system $M/M/1/N$ and a system with correlated arrivals, the system $SSMP(2)/D/1/N$, allow the validation of the method through a comparison with analytical values. The goal of these investigations is the evaluation of the complementary distribution function $G(x)$ of the arrival occupancy, which includes as the last value the loss probability P_B . With this approach it was possible to investigate very low cell loss probabilities in ATM networks in the order of $P_B = 10^{-9}$ with good accuracy in several minutes up to a few hours. A straightforward simulation of these examples without the RESTART/LRE-method would have taken several years. With this approach even lower values for loss probabilities can be evaluated, e. g. $P_B = 10^{-20}$. At the end of chapter eight the extension of the method for queueing networks is presented and validated for examples of ATM reference connections.

The main contribution of the presented work is on one hand the survey of modelling techniques and performance analysis of communication networks and on the other hand the in-depth study of the speed-up possibilities for the simulation of rare events.

VORWORT

Seit meiner ersten Beschäftigung mit der *stochastischen, diskreten Simulationstechnik* in einem Praktikum während des Studiums in Karlsruhe haben mich die Möglichkeiten dieser Technik fasziniert. Während des Studiums habe ich das Praktikum, in welchem verschiedene Aufgaben aus dem Bereich Betriebssysteme in Simula67 auf einem Großrechner behandelt wurden, als studentische Hilfskraft betreut.

Kommunikationsnetze wurden zu meinem thematischen Schwerpunkt als Assistentin am Lehrstuhl Kommunikationsnetze (früher Datenfernverarbeitung). Die Wahl dieses Schwerpunkts hat sich durch die rasante Entwicklung in diesem Bereich als ein spannendes Gebiet mit ständig neuen Herausforderungen erwiesen.

Die *Verkehrstheorie* als analytisches Werkzeug konnte ich als Übungsleiterin der entsprechenden Vorlesung erarbeiten und in meiner Dissertation [Gör83] vertiefen, in der die Warteschlangendisziplin SRPT (*Shortest Remaining Processing Time first*) mit Wartezeit analytisch betrachtet und die Ergebnisse simulativ validiert wurden. Diese Warteschlangendisziplin ist insbesondere für Kommunikationsnetze interessant, da bei der Übertragung von Daten die Verarbeitungszeit proportional zur Datenmenge ist und damit im voraus bekannt ist. Die Untersuchung spezieller Protokolle war das Thema einiger von mir betreuter Arbeiten, in denen Paketvermittlungsnetze [Sug81, Lie82], Lokale Netze [Hau90, Bes91, Meh92], Lastmodelle für multimediale Anwendungen [Bey93] sowie neue Dienste im Intelligenten Netz [Rüß95, Meh96] analytisch und simulativ untersucht wurden. Meine Veröffentlichungen habe ich in einem getrennten Literaturverzeichnis am Ende dieser Arbeit aufgeführt. Sie reichen von ersten Veröffentlichungen im Jahre 1981 bis zu mehreren Veröffentlichungen im Jahre 1997 und decken thematisch das Spektrum: Verkehrstheorie, Simulationswerkzeuge und den Anwendungsbereich Kommunikationsnetze ab.

Nach meiner Promotion 1983 habe ich einige Jahre selbständig gearbeitet und u. a. ein Netz zusammen mit einer Anwendung für einen Kunden entwickelt sowie beim Forschungszentrum Jülich (KFA) an der Entwicklung des KFA-Netzes mitgearbeitet [Gör87, Gör88], wodurch ich diesen Anwendungsbereich vertiefen konnte.

Während meiner Zeit als Assistentin und auch später als Oberingenieur am Lehrstuhl Kommunikationsnetze habe ich die Weiterentwicklung der Simulationstechnik als Werkzeug betrieben bzw. betreut. Die Simulationswerkzeuge [Koc82, GKS91, JG93] und insbesondere die zur Zeit am Lehrstuhl Kommunikationsnetze bestehende Simulationsbibliothek CNCL repräsentieren den jeweiligen Stand der Technik und sind die Basis für alle Simulationsprojekte des Lehrstuhls.

Um den Studenten diese Techniken auch praktisch näherzubringen, habe ich zu Beginn des Wintersemesters 91/92 am Lehrstuhl ein Simulationspraktikum aufgebaut und angeboten, welches jährlich stattfindet. In diesem Praktikum werden die Studenten mit der Simulationstechnik am Beispiel der Simulationsbibliothek CNCL vertraut gemacht mit dem Ziel, nach einigen einführenden Versuchen, die Modellierung und simulative Bewertung des FDDI-Zugriffsprotokolls zu lernen und durchzuführen, siehe Abschnitt 6.1. Die theoreti-

sche Abrundung zu diesem Praktikum wird mit der seit dem Sommersemester 1992 vom Lehrstuhl angebotenen Vorlesung *Stochastische Simulationstechnik mit Anwendungen auf Kommunikationsnetze* [SG96a] geboten, zu der ich wesentliche Teile erarbeitet habe.

Lehraufträge an der Gesamthochschule Essen (Wintersemester 94/95, Sommersemester 95, Wintersemester 95/96) gaben mir die Gelegenheit zu einer aktuellen Aufarbeitung des Stoffes durch Vorlesungen zum Thema Kommunikationsnetze (*Betriebliche Netze und Verbundsysteme, Kommunikationssysteme II*).

Möglichkeit zur Diskussion mit Experten aus den entsprechenden Bereichen hatte ich bei Tagungen und Workshops, die ich zum Teil mitorganisiert habe, wie z. B. die Tagung *Messung, Modellierung und Bewertung von Rechen- und Kommunikationssystemen* 1993 [WS93] und *Intelligence in Services and Networks* 1994 [KMN94] sowie einen Workshop zum Thema Simulation 1996 [GE96]. Die Mitarbeit in der ITG-Fachgruppe 5.2.1 (früher 1.2.1) *Verkehrstheorie* seit 1983 brachte insbesondere die Möglichkeit, mit Experten aus der Industrie zusammenzuarbeiten, um so die praktisch relevanten Fragestellungen aufzugreifen. In dieser Fachgruppe wurde die neue ITG-Empfehlung zu den *Begriffen der Nachrichtenverkehrstheorie* erarbeitet, woran ich als Editor intensiv mitgearbeitet habe.

Als Gruppenleiterin für das europäische RACE II-Projekt Mobilise von 1992 – 1995 habe ich die Arbeiten für das Demonstrationssystem und die Leistungsbewertung – auch projektweit als *Core Task Leader* – gesteuert. Die Leistungsbewertung neuer Telekommunikationsdienste wurde einerseits analytisch [FKG95, Gun96] und andererseits simulativ durchgeführt [KG95, Kle96, GGK97]. Die Arbeit in diesem Projekt ermöglichte mir die enge Zusammenarbeit mit europäischen Kollegen. Als *Evaluator* und *Auditor* im Rahmen des aktuellen europäischen ACTS-Forschungsprogramms konnte ich 1996 und 1997 die Zielsetzung der neuen Projekte bewerten.

In den letzten Jahren habe ich mich in der Forschung mit dem Thema *Simulation seltener Ereignisse* befaßt. Erste Ergebnisse dazu wurden auf dem *International Teletraffic Congress 1994* [SG94] vorgetragen. Dieses Verfahren wird in der vorliegenden Arbeit als mehrstufiges Verfahren wesentlich erweitert, siehe Kapitel 8. Die optimale Stufenzahl wird analytisch betrachtet und es können nunmehr mit diesem Verfahren z. B. Systeme mit seltenen Ereignissen in der Größenordnung 10^{-20} in wenigen Stunden durchgeführt werden, was bei einer direkten Simulation bei gleicher Genauigkeit ungefähr 10^{12} Jahre benötigt hätte. Diese Ergebnisse [GS96] sowie eine Studie [Gör96] wurden auf der *Winter Simulation Conference 1996* präsentiert. Ende August 1997 habe ich zusammen mit Dr. C. Kelling in Aachen einen zweitägigen, internationalen Workshop [GE97] zum Thema *Rare Event Simulation* organisiert, bei dem 25 Experten aus 8 Ländern teilgenommen haben.

In der vorliegenden Arbeit wird eine Übersicht über die Leistungsanalyse von Kommunikationsnetzen gegeben, indem einerseits die grundsätzlichen Techniken vorgestellt werden und andererseits diese Techniken an einer Vielzahl von Modellen demonstriert werden.

Aachen, im November 1997

Carmelita Görg

INHALTSVERZEICHNIS

1	Einführung	1
1.1	Methoden zur Leistungsbewertung	2
1.2	Modellbildung	5
1.3	Kommunikationsnetze	6
2	Simulative Leistungsbewertung	9
2.1	Simulationssysteme	11
2.2	Ereignis- und prozeßorientierte Simulation	13
2.3	Anforderungen an höhere Programmiersprachen	17
2.4	Eine objektorientierte Simulationsbibliothek	18
3	Verkehrsmodelle für Kommunikationsnetze	27
3.1	Zufallsgrößen und Zufallsprozesse	28
3.1.1	Kennwerte und Verteilungen	28
3.1.2	Stochastische Prozesse	33
3.2	Markoff-Ketten	35
3.2.1	Kenngößen	35
3.2.2	Die 2-gliedrige Markoff-Kette	37
3.3	Korrelation	39
3.3.1	Der lokale Korrelationskoeffizient	39
3.3.2	Korrelierte Prozesse	41
3.4	Elementare Wartesysteme	43
3.4.1	Klassifikation von Wartesystemen	43
3.4.2	Wichtige Ergebnisse	46
3.4.3	Analysemethoden	50

3.4.4	Das elementare Referenzmodell $M/M/1/N$	51
3.5	Warteschlangennetze	54
4	Zufallszahlengeneratoren	59
4.1	Basisgeneratoren	60
4.2	Generatoren für beliebige Verteilungen	63
4.3	Empfehlungen	64
5	Statistische Auswertung	65
5.1	Initialisierung und transiente Phase	65
5.2	Terminierende Simulationen	67
5.3	Stationäre Auswertemethoden	67
5.4	Das Batch-Means-Verfahren	71
5.5	Der LRE-Algorithmus	73
6	Modelle für Kommunikationsnetze und zugehörige Ergebnisse	85
6.1	Hochgeschwindigkeitsnetz FDDI	85
6.2	Intelligente Netze (IN)	94
6.2.1	Erweiterte UPT-Dienste	95
6.2.2	Bewertungsaspekte und Einflußgrößen	97
6.2.3	Verkehrsmodell	99
6.2.4	Bewertungsergebnisse	105
6.2.5	Zusammenfassende Bewertung	111
6.3	Mobilfunknetze	113
6.3.1	Allgemeine Anforderungen	113
6.3.2	DECT	120
6.4	Simulationsprojekte in der Praxis	130

7	Verfahren zur Simulationsbeschleunigung	132
7.1	Verteilte und parallele Simulation	133
7.2	Hybride Techniken	134
7.3	Varianzreduktion	135
7.3.1	Gemeinsame Zufallszahlen (<i>common random numbers</i>)	135
7.3.2	Antithetische Stichproben (<i>antithetic sampling</i>)	137
7.3.3	Steuervariablen (<i>control variables</i>)	137
7.3.4	Indirekte Schätzung (<i>indirect estimation</i>)	138
7.4	Simulation seltener Ereignisse	139
7.4.1	Geschichtete Stichproben (<i>stratified sampling</i>)	140
7.4.2	Wesentliche Stichproben (<i>importance sampling</i>)	141
7.4.3	Wesentliche Partitionierung (<i>importance splitting</i>)	142
8	Simulation seltener Ereignisse mit RESTART/LRE	145
8.1	Das RESTART/LRE-Prinzip	145
8.2	Die minimale Versuchszahl für m Stufen	147
8.3	Die näherungsweise optimale Stufenzahl	151
8.3.1	Ergebnisformeln	153
8.3.2	Beispiel M/M/1/N	154
8.4	Einstufiges RESTART/LRE-Verfahren	160
8.5	Beschreibung des RESTART/LRE-Algorithmus	160
8.6	RESTART/LRE für Wartenetze	162
8.7	Verlustwahrscheinlichkeiten in ATM-Netzen	164
8.7.1	Modelle für einzelne ATM-Knoten	166
8.7.2	Modelle für sequentielle ATM-Knoten (Netzmodelle)	171
9	Zusammenfassung und Ausblick	175

A Wichtige Verteilungen	177
B Ergänzungen zum RESTART/LRE-Verfahren	188
B.1 Minimale Versuchszahl n_{min}	188
B.2 Optimale Niveaus des RESTART/LRE-Verfahrens	189
B.3 Nachweis des Minimums	189
B.4 Maximaler Gesamtfehler d_{max}	191
C Bezeichnungen	193
D Abkürzungsverzeichnis	197
Persönliches Literaturverzeichnis – Quellenverzeichnis	201
Literaturverzeichnis – Quellenverzeichnis	205

Einführung

Die Simulationstechnik ist inzwischen ein unentbehrliches Hilfsmittel zur Untersuchung komplexer Systeme geworden. Diese Entwicklung wird gefördert durch die Verfügbarkeit von immer leistungsfähigeren und preiswerteren Rechnern. Workstations sind inzwischen an jedem Arbeitsplatz vorhanden und sogar in vielen privaten Haushalten sind leistungsfähige Rechner zu finden. Simulative Untersuchungen, die früher nur auf sehr teuren Großrechnern möglich waren, lassen sich heute auf einem PC (PC – *Personal Computer*) durchführen. Diese Entwicklung ermöglicht andererseits die Untersuchung immer komplexerer Systeme auf heutigen Höchstleistungsrechnern.

Durch die Verfügbarkeit von leistungsfähigen Workstations und kommerziellen Softwareprodukten wird diese Technik inzwischen nicht nur von wenigen Spezialisten, sondern von einer Vielzahl von Anwendern aus sehr verschiedenen Bereichen eingesetzt. Die graphischen Benutzungsoberflächen der kommerziellen Werkzeuge täuschen eine leichte Anwendung der Methoden vor, die nach wie vor mit sehr viel Sorgfalt gehandhabt werden müssen, um sinnvolle Ergebnisse zu erzielen.

In dieser Arbeit wird ein Überblick über die Methoden der Leistungsbewertung gegeben. Den Schwerpunkt der Arbeit bilden die Methoden, die auf verkehrstheoretischen Modellen aufbauen. Dies sind zum einen Methoden, die die analytische Berechnung von Modellgrößen erlauben, und zum anderen die stochastische Simulationstechnik. Der Einsatz der Methoden zur Leistungsbewertung von Kommunikationsnetzen und ihren Protokollen sind in Anbetracht der aktuellen Erweiterungen bestehender Netze und dem Aufbau neuer Netze von außerordentlicher Wichtigkeit. In den folgenden Abschnitten dieser Einführung wird das Zusammenspiel der Methoden beschrieben, in die Modellbildung eingeführt und eine kurze Übersicht zum Anwendungsgebiet Kommunikationsnetze gegeben.

Im zweiten Kapitel dieser Arbeit wird die simulative Leistungsbewertung erläutert und am Beispiel einer objektorientierten Klassenbibliothek vertieft. Eine Übersicht über Verkehrsmodelle und wichtige Ergebnisse analytischer Verfahren sind im dritten Kapitel zusammengestellt. Die für die stochastische Simulation benötigten Zufallsgeneratoren und statistischen Auswerteverfahren werden im anschließenden vierten und fünften Kapitel beschrieben. Ausgewählte Kommunikationsnetze, ihre Modelle und die zugehörigen durch analytische Berechnung (exakt und approximativ) und

stochastische Simulation gewonnenen Ergebnisse werden am Beispiel eines lokalen Netzes (FDDI – *Fiber Distributed Data Interface*), des Intelligenten Netzes (IN – *Intelligent Network*) und eines Mobilfunknetzes (DECT – *Digital Enhanced Cordless Telephone*) im sechsten Kapitel dargestellt.

Ein Überblick über Verfahren zur Simulationsbeschleunigung, eine detaillierte Beschreibung des in dieser Arbeit erweiterten RESTART/LRE-Verfahrens sowie die Anwendung dieses Verfahrens auf Modelle von ATM-Netzen bilden die Schwerpunktkapitel 7 und 8 dieser Arbeit. Eine Bewertung der Arbeit sowie ein Ausblick auf Folgearbeiten, die durch diese Arbeit möglich geworden sind, werden im Abschlußkapitel 9 diskutiert.

1.1 Methoden zur Leistungsbewertung

Insbesondere zur Leistungsbewertung neuer Kommunikationsnetze und Dienste werden durch die große Komplexität der Netze und die extremen Anforderungen an die Dienstgüte neue Anforderungen an die Methoden zur Leistungsbewertung gestellt. Der flexibelste Ansatz ist die rechnergestützte Simulationstechnik. Als Alternativen stehen, wie Abbildung 1.1 zeigt, analytische Methoden und Messungen zur Verfügung.

Analytische Methoden sind meist sehr effizient, erfordern aber einen hohen Abstraktionsgrad und sind in der Durchführung sehr anspruchsvoll. Kleine Veränderungen im Modell können sehr große Änderungen in der Methode erforderlich machen bzw. die Anwendbarkeit der Methode insgesamt in Frage stellen. Bei komplexen Modellen stoßen auch die analytischen Methoden an ihre Grenzen mit hohen Forderungen an Speicher- und Zeitaufwand sowie Genauigkeitsfragen bei den erzielten numerischen Ergebnissen.

Messungen und Experimente am realen System sind nur dann möglich, wenn das reale System oder ein Prototyp bereits existiert. In diesem Fall geben sie die präziseste Bewertung des Systems. Die Entwicklung von Prototypen ist jedoch mit hohen Kosten verbunden und limitiert die Möglichkeiten der Untersuchung großer Systeme.

Abhilfe schaffen hier die Methoden, die auf Modellen der realen und geplanten Systeme aufbauen. Dies gilt für die simulativen und analytischen Methoden, die zur Untersuchung von Erweiterungen bestehender Systeme und neuer Systementwürfe herangezogen werden. Vorteil der simulativen Methode ist die große Flexibilität und ein prinzipiell beliebiger Detaillierungsgrad, nachteilig ist der meist hohe Rechenzeitaufwand, z. B. bei sehr hohen Dienstgüteanforderungen in Hochgeschwindigkeitsnetzen. Hier ist das Problem der *seltenen Ereignisse* zu nennen, das bei einfacher direkter Simulation nicht mit genügender Genauigkeit der Ergebnisse in angemessener Zeit durchgeführt werden kann. Die Problematik und damit die Notwendigkeit

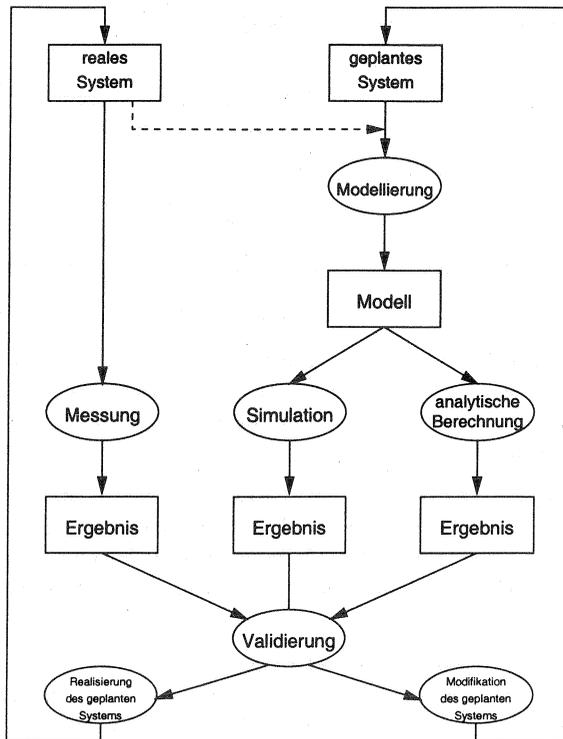


Abbildung 1.1: Zyklus der Leistungsbewertung

zur Bestimmung der Wahrscheinlichkeiten seltener Ereignisse findet sich in den verschiedensten Anwendungen, z. B. Verlustwahrscheinlichkeiten im Breitband-ISDN oder Fehlerwahrscheinlichkeiten bei Kommunikationsprotokollen wie z. B. bei der Fahrzeug-Baken-Kommunikation. Hier sind neue effiziente Methoden notwendig, die auf eine möglichst große Problemklasse anwendbar sind.

Im folgenden werden am Beispiel der Simulation die Schritte einer Leistungsanalyse beschrieben:

Modellierung	(z. B. Festlegung der Systemstruktur und Bedienstrategien in einem Warteschlangennetz)
Definition der Eingabeparameter	(z. B. stochastische Prozesse)
Implementierung	(z. B. Simulationssprache)
Auswertung der Ergebnisse	(z. B. statistische Auswerteverfahren)
Validierung	(z. B. Vergleich mit analytischen Ergebnissen)

Die *Modellierung* bildet ein reales oder geplantes System auf ein Modell ab, siehe Abschnitt 1.2. Bei jeder Modellbildung muß vereinfacht werden, und die Kunst besteht darin, die wesentlichen Eigenschaften in das Modell einzubringen und unwesentliche Details zu vernachlässigen.

Zu den *Eingabeparametern* gehören alle Größen des Modells, die variiert werden können, z. B. stochastische Prozesse und deren Parameter, die Anzahl von Bedienstationen oder die Größe von Puffern. Teilweise können die Verteilungen aus Messungen stammen, oder es werden Annahmen getroffen, die ein Spektrum der realen Parameter abdecken.

Für die *Implementierung* ist ein geeignetes Simulationssystem zu wählen. Werkzeuge mit graphischen Oberflächen ohne Programmierung sind meist nur für eine eingeschränkte Problemklasse sinnvoll anwendbar. Für neue Problemstellungen benötigt man weiterhin flexible, allgemein anwendbare Werkzeuge, wie z. B. eine höhere Programmiersprache mit Programmbibliotheken zur Simulationsunterstützung.

Die Laufzeit einer simulativen Untersuchung und die Zuverlässigkeit der so gewonnenen Ergebnisse hängen von der *statistischen Auswertemethode* ab, die zur Fehlerabschätzung bzw. zur Steuerung des Simulationslaufes eingesetzt wird, siehe Kapitel 5. Zugleich mit der Laufzeit einer Simulation ist auch die Genauigkeit begrenzt, mit der ein seltenes Ereignis in einer zu untersuchenden Sequenz charakterisiert werden kann. Ein Schwerpunkt dieser Arbeit ist die Untersuchung eines neuen Verfahrens zur statistisch gesicherten Bestimmung der Häufigkeit seltener Ereignisse durch Beschleunigung der Simulationsläufe, siehe Kapitel 8.

Die *Validierung*¹ [Jai91] bezeichnet die Überprüfung der Annahmen und die Übereinstimmung des realen Systems mit dem Modell bzgl. der erzielten Ergebnisse. Die Validierung erfolgt im allgemeinen für eine spezielle Wahl der Eingabeparameter durch den Vergleich mit Ergebnissen, die durch eine andere Methode gewonnen

¹Im Unterschied dazu bedeutet *Verifikation* [Jai91] die korrekte Implementierung des Modells, dazu sind alle Techniken anwendbar, die es zur Verifikation von Programmen gibt: wie z. B. modulares Programmieren, top-down design, Plausibilitäts- und Stresstests.

wurden, z. B. durch Messungen oder analytische Berechnungen.

Sämtliche Schritte einer Simulationsstudie werden anhand von Beispielen an verschiedenen Stellen in dieser Arbeit beschrieben, siehe Kapitel 6 und Abschnitt 8.7.

1.2 Modellbildung

Grundlage sowohl der analytischen als auch der simulativen Leistungsbewertung eines Systems ist die Modellbildung. Die Abgrenzung des Systems ist dabei abhängig von der durchzuführenden Untersuchung. Ziel ist die Nachbildung des Systems als *Modell* mit möglichst wenigen Parametern, welches trotzdem die wesentlichen Eigenschaften für die geforderte Untersuchung enthält.

Der Begriff *System* [LK91, Bei90] ist hier sehr allgemein zu verstehen. Ein *System* kann definiert werden als eine Menge von Komponenten mit ihren spezifischen Eigenschaften, z. B. Personen, Maschinen oder Rechner, die einzeln agieren und miteinander interagieren, um ein vorgegebenes Ziel zu erreichen. Bei der Beschreibung des Systems müssen die Systemgrenzen klar gezogen werden, d. h. daß das System gegen die System-Umwelt abgegrenzt werden muß. Die Grenzen des Systems hängen von der Aufgabenstellung der Leistungsuntersuchung ab. Es kann z. B. ein Rechner oder ein gesamtes Rechnernetz untersucht werden. Der Systembegriff kann in diesem Sinne hierarchisch verstanden werden, d. h. ein System kann wiederum aus Teilsystemen bestehen.

Der *Zustand* eines Systems bzw. des zugehörigen Modells wird definiert als die Menge der Variablen, die ein System bzw. Modell zu einem bestimmten Zeitpunkt beschreiben. Bei einem lokalen Netz gehören dazu z. B. die Pakete, die auf Übertragung warten, oder die Restübertragungszeit des Pakets, welches zu dem Zeitpunkt übertragen wird.

Systeme, deren Zustandsvariablen sich unverzüglich, d. h. ohne Zeitdauer für die Änderung, nur zu klar unterscheidbaren Zeitpunkten ändern, nennt man (*zeit-*)*diskret*. Bei einem Rechnernetz sind dies z. B. die Zeitpunkte von Paketankünften oder das Ende der Bearbeitung eines Pakets. Ändern sich die Zustandsvariablen kontinuierlich bezüglich der Zeit, so spricht man von einem *kontinuierlichen* System. Der Aufenthaltsort eines Teilnehmers in einem Mobilfunknetz setzt sich z. B. als Koordinatenpaar (x, y) aus kontinuierlichen Größen zusammen. Die meisten Systeme sind nicht rein diskret oder rein kontinuierlich, so daß die Unterscheidung meist nur den vorherrschenden Typ der Zustandsvariablen beschreibt.

Um ein System mit Hilfe eines Modells zu untersuchen, kann man entweder *physikalische* Modelle oder *abstrakte* Modelle benutzen. Ein physikalisches Modell nutzt physikalische Eigenschaften aus, z. B. elektromagnetische Eigenschaften in einem

Versuchsaufbau. Im folgenden werden ausschließlich *abstrakte* Modelle betrachtet, die die logischen und quantitativen Zusammenhänge beschreiben und die verändert werden können, um zu sehen wie sich das Verhalten des Modells — und bei übereinstimmender Modellierung entsprechend das reale System — ändern würde.

Man kann bzgl. des Abstraktionsgrades zwei Modelltypen [TG96] unterscheiden:

$$\text{reales System} \rightarrow \text{Systemmodell} \rightarrow \text{Verkehrsmodell}$$

Ein *Systemmodell* ist ein systemnahes, detailliertes Modell des Systems, welches eine detaillierte Modellsimulation erlaubt. Neuere Ansätze dazu beziehen Teile des realen Systems mit in die Simulation ein. Bei der Bewertung von Protokollen [Böh96] kann man z. B. direkt von der formalen Protokollspezifikation ausgehen, wobei Teile des realen Systems unmittelbar in eine Modellumgebung eingebettet werden. Dies vermeidet Modellierungsfehler und erlaubt z. B. die Emulation des Protokolls auf bestimmten Prozessortypen, um die notwendige Leistungsfähigkeit der Zielhardware zu bestimmen. Für umfangreiche Leistungsbewertungen in größeren Systemen sind diese systemnahen Modelle jedoch zu detailliert.

Ein *Verkehrsmodell* (*traffic model*) beschreibt nur die Teile des Systems, die für die Verkehrsabläufe und die Leistungsbewertung wichtig sind, d. h. man kann z. B. einen Bedienvorgang auf die Zeit abbilden, die für die Bedienung benötigt wird und diese durch eine Bediendauerverteilung modellieren. Beispiele für Verkehrsmodelle sind Wartesysteme (Abschnitt 3.4) und Warteschlangennetze (Abschnitt 3.5).

Ideal wäre eine automatisierte oder teil-automatisierte Abstraktion des Systemmodells in ein Verkehrsmodell, welches eine effizientere Simulation oder eine analytische Lösung erlaubt. Dies ist bislang praktisch nicht realisierbar.

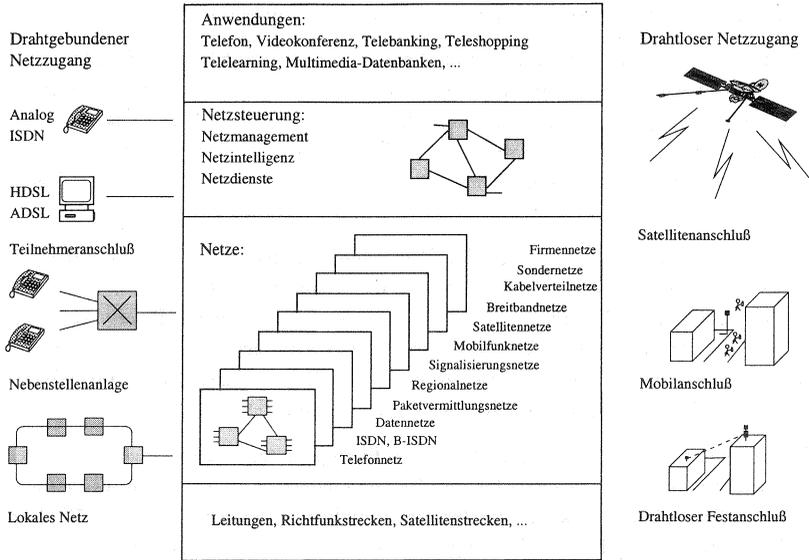
1.3 Kommunikationsnetze

Der Begriff Kommunikationsnetz ist ein umfassender Begriff, der die unterschiedlichsten Arten von Netzen zusammenfaßt, siehe Abbildung 1.2 [vA95]. Dazu gehören sowohl lokale Netze als auch Weitverkehrsnetze, öffentliche sowie nicht-öffentliche Netze, Festnetze sowie Funknetze. Die traditionelle Trennung von Datennetzen und Telefonnetzen ist durch Integration der entsprechenden Dienste in den neuen Netzen aufgehoben. Sie ermöglichen durch dieses Zusammenwachsen neue multimediale Anwendungen.

Die Anwendungsgebiete von Kommunikationsnetzen reichen von Telefonie über Videokonferenzen zu Multimedia-Datenbanken sowie Telebanking, Teleshopping und Telelearning. In Firmennetzen laufen die üblichen Firmenanwendungen zur Unter-

stützung der Entwicklung von Produkten, der Produktion, der Verwaltung, und des Vertriebs.

Im Teilnehmer-Anschlußbereich findet man den direkten Anschluß von Geräten (Telefon, Terminal, Einzelrechner) oder ihren indirekten Anschluß über ein lokales Netz oder eine Nebenstellenanlage. Öffentliche Netze sind das analoge und digitale (ISDN) Telefonnetz, bei dem die Teilnehmer fast ausschließlich über Kupferleitungen angeschlossen sind. Als weiteres öffentliches Netz gibt es das Breitbandverteilstrecknetz, bei dem der Teilnehmer über Koaxialkabel für die Übertragung von Hörfunk- und Fernsehprogrammen angeschlossen ist. Erweiterungen des Netzes sehen eine Nutzung für bidirektionale Kommunikation vor. Dazu wurden verschiedene Verfahren entwickelt, die eine asymmetrische (ADSL – *Asymmetric Digital Subscriber Line*) bzw. symmetrische (HDSL – *High Speed Digital Subscriber Line*) Aufteilung der Übertragungsbandbreite vorsehen.



HDSL: High Speed Digital Subscriber Line ISDN: Integrated Digital Services Network
 ADSL: Asymmetric Digital Subscriber Line B-ISDN: Breitband-ISDN

Abbildung 1.2: Übersicht Kommunikationsnetze

Der Netzzugang kann auch drahtlos erfolgen. Dies ist insbesondere im Hinblick auf die Liberalisierung im Telekommunikationsmarkt für neue Netzbetreiber eine interessante Alternative für den Aufbau eines neuen Zugangsnetzes, um schnell viele ortsfeste Teilnehmer anzuschließen. Auch Mobilfunknetze sind drahtlose Zugangsnetze,

wobei eine terrestrische Basisstation oder ein Satellit die Verbindung zum Festnetz herstellen.

Öffentliche Netze sind hybrid durch Kabel-Verbindungen (Kupferkabel und Glasfaserkabel) und Funkverbindungen (Richtfunkstrecken und Satellitenstrecken) realisiert. Teilweise werden sie durch überlagerte Signalisierernetze ergänzt. Diese ermöglichen die Realisierung der sog. Intelligenten Netze, die die Funktionen, Schnittstellen und dienstunabhängigen Basisbausteine als Grundlage für die Implementierung von Mehrwertdiensten in modernen digitalen Telekommunikationsnetzen spezifizieren. Ein Beispiel ist UPT (*Universal Personal Telecommunication*), ein Dienst für die Unterstützung der Teilnehmer-Mobilität in Festnetzen.

Mobilfunknetze umfassen lokale drahtlose Netze (z. B. HIPERLAN – *High Performance Radio LAN*), Bündelfunknetze (z. B. TETRA – *TransEuropean Trunked Radio*), zellulare Netze (z. B. GSM – *Global System for Mobile Communication*) und weltweit geplante Satellitenetze (z. B. Iridium, Globalstar).

Firmennetze bestehen aus den lokalen Netzen an den verschiedenen Standorten einer Firma sowie aus dem Verbindungsnetz dieser Teilnetze, das aus Standleitungen, Wählverbindungen sowie ggf. Satellitenverbindungen bestehen kann. Als ein Dienst der neuen Intelligenten Netze werden sog. *Virtual Private Networks* angeboten, die ein auf Firmen zugeschnittenes Dienstpaket anbieten, z. B. die geschlossene Benutzergruppe.

Die z. Zt. wichtigste Entwicklung für Breitbandnetze ist das Breitband-ISDN (B-ISDN), in welchem die ATM-Technik (*ATM – Asynchronous Transfer Mode*) eingesetzt wird, siehe Abschnitt 8.7. ATM stellt eine sehr schnelle Paketvermittlung mit kleinen Paketen fester Länge zur Verfügung. Wegen der geringen Bitfehlerwahrscheinlichkeit der für die Übertragung benutzten Glasfaserkabel wird auf Fehlersicherung in Teilabschnitten des Netzes weitgehend verzichtet.

Im lokalen Bereich werden hauptsächlich Hochgeschwindigkeitsnetze als Verbindungsnetz (*backbone*) zwischen lokalen Netzen eingesetzt. Zunehmend werden auch Arbeitsplatzrechner direkt an Hochgeschwindigkeitsnetze angeschlossen. Wichtige Beispiele für lokale Hochgeschwindigkeitsnetze sind FDDI (*FDDI – Fiber Distributed Data Interface*), siehe Abschnitt 6.1, und DQDB (*DQDB – Distributed Queue Dual Bus*). ATM LANs verschiedener Konfigurationen erlauben den nahtlosen Übergang in ATM-Weitverkehrsnetze.

Photonische Netze [vA95], d. h. Netze mit optischer Übertragung und Vermittlung, versprechen für die etwas fernere Zukunft eine weitere Steigerung der Übertragungsgeschwindigkeit und damit auch eine Weiterentwicklung der Anwendungen.

Simulative Leistungsbewertung

Unter Simulation versteht man allgemein das Durchspielen eines Modells mit vorgegebenen Eingaben zur Beobachtung der resultierenden Ausgaben. Die Technik der Simulation wird einerseits in vielen Bereichen der Ingenieurwissenschaften und Physik, andererseits aber auch im Bereich der Wirtschafts- und Sozialwissenschaften angewandt.

Ein Modell wird beschrieben durch eine Menge von Zuständen und zugehörige erlaubte Zustandsübergänge, siehe Abschnitt 1.2. Abhängig von der Art der Zustandsübergänge unterscheidet man verschiedene Arten der Simulation: *zeitdiskrete* bzw. *zeitkontinuierliche* Simulation. Abkürzend spricht man von *diskreter* bzw. *kontinuierlicher* Simulation. In dieser Definition ist der Zeitbegriff enthalten, so daß diese Art von Simulation auch als *dynamische* Simulation bezeichnet wird, im Unterschied zu einer *statischen* Betrachtung, die einen festen Zeitpunkt beschreibt.

In dieser Arbeit wird die *diskrete, dynamische* Simulation betrachtet, bei der nur zu bestimmten, klar unterscheidbaren, aufzählbaren Zeitpunkten t_i der neue Modellzustand betrachtet bzw. berechnet wird. Kontinuierliche Veränderungen werden näherungsweise nachgebildet (*quasikontinuierlich*). Im Bereich der diskreten Simulation unterscheidet man nochmals zwischen *periodenorientiert* (*time sequencing*) und *ereignisorientiert* (*event sequencing*) in Abhängigkeit davon, ob Zustandsänderungen in festen Intervallen (Periode) oder zu beliebigen diskreten Zeitpunkten (Ereignis) stattfinden können. Darüber hinaus können Ereignisse zu sog. *Prozessen* zusammengefaßt werden, dann spricht man von *prozessorientierter* Simulation. Die Einordnung in die allgemeine Simulationsklassifikation ist in Abbildung 2.1 dargestellt.

Die periodenorientierte Simulation, die häufig auch als Zeitraster-Verfahren bezeichnet wird, wird hier nicht gesondert betrachtet, da sie als Spezialfall der ereignisorientierten Simulation aufgefaßt werden kann.

Ein weiteres Unterscheidungsmerkmal bei Simulationssystemen betrifft die Eingabewerte des Modells. Sind die Werte fest vorgegeben, so spricht man von *deterministischer* Simulation, die zu jeder Eingabe des Systems genau eine Ausgabe liefert. Werden Zufallsgrößen zur Beschreibung des Modells verwendet, so spricht man von *stochastischer* Simulation. Nach Möglichkeit werden Meßdaten zur Erstellung von Modellen und Verteilungen für Eingabedaten benutzt, so daß bei der Simulation Zufallszahlen mit ähnlichen Eigenschaften generiert werden können. Für un-

abhängige Eingabedaten genügt es die Verteilung als Tabelle abzulegen. Eine weitere Möglichkeit ist die Approximation der gemessenen Verteilung durch eine Mischung von formelmäßig vorgegebenen Verteilungen, z. B. Erlangmischverteilungen [Sch92]. Die vorgegebenen Verteilungen beschreiben z. B. die Zwischenankunftszeit oder die Länge von Nachrichten in einem Kommunikationsnetz.

Vorteil der Simulationstechnik ist die Nachbildung beliebiger realer Systeme durch ein Programmsystem, das nur durch die Rechen- und Speicherkapazität begrenzt ist. Nachteil gegenüber analytischen Verfahren ist meist der erhöhte Laufzeitaufwand, der für statistisch abgesicherte Ergebnisse benötigt wird, so daß analytische Verfahren besser für umfangreiche Parameterstudien geeignet sind. Weitere Probleme der Simulationstechnik liegen in der Bestimmung von Start- und Ende-Zeitpunkten der statistischen Auswertung einer stationären Simulation [Paw90]. Der Startzeitpunkt sollte so gewählt werden, daß die Gesamtlaufzeit optimiert wird, d. h. der Startzeitpunkt ist auch vom Endekriterium abhängig. Das Endekriterium beinhaltet eine Fehlerschranke, die es zu unterschreiten gilt und die zur Steuerung der Simulation während des Simulationslaufes bestimmt werden muß. Fehlermaße sind z. B. Konfidenzintervalle oder der LRE-Fehler (LRE – *Limited Relative Error*) des Untersuchungsergebnisses, vgl. Kapitel 5.

Zur Steigerung der Simulationsrate (Ereignisse/Rechenzeit) werden verschiedene Arten der parallelen bzw. verteilten Simulation eingesetzt. Sie unterscheiden sich in der Art der Aufteilung der Funktionen, die z. B. nach simulationstechnischen oder modellmäßigen Gesichtspunkten erfolgen kann. Teilweise wurde spezielle Hardware entwickelt, um Simulationen zu beschleunigen [Leh79, Bar83]. Inzwischen werden meist universelle Multiprozessorsysteme [Klu90] zu diesem Zweck eingesetzt, wie z. B. Transputer.

Im folgenden wird im Sinne der obigen Definitionen die dynamische, diskrete, stochastische und ereignisorientierte Simulationstechnik betrachtet, siehe Abbildung 2.1. Als Einführung werden im folgenden Simulationssysteme und deren Anforderungen beschrieben. Die Umsetzung dieser Konzepte in eine objektorientierte Klassenbibliothek zur Unterstützung der Entwicklung von Simulationssystemen findet man in Abschnitt 2.4.

Die Grundlagen zu *Zufallszahlengeneratoren* und *statistischen Auswerteverfahren* als essentiellen Komponenten eines Simulationssystems werden in den Kapiteln 4 und 5 beschrieben.

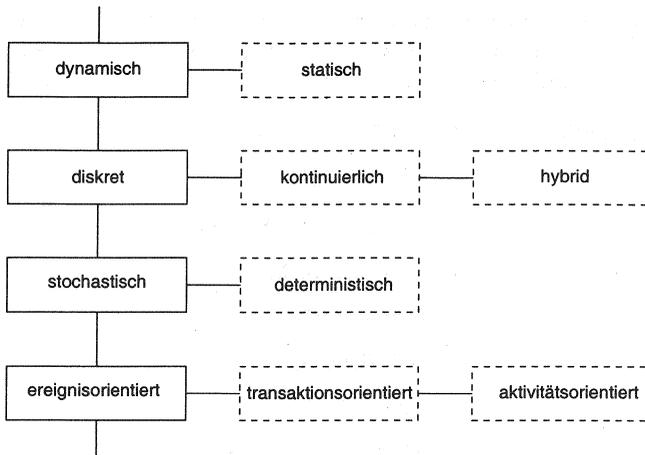


Abbildung 2.1: Einordnung in die Simulationsklassifikation

2.1 Simulationssysteme

Betrachtet man existierende Simulationsprogramme, so stellt man fest, daß häufig für jedes Problem ein komplett neues Programm geschrieben wird. Dies resultiert daraus, daß die zu untersuchenden Modelle meist sehr unterschiedlich sind und nicht sofort auf allgemeine Modelle, wie z. B. Warteschlangennetze abgebildet werden können. In all diesen Programmen gibt es jedoch einige grundlegende Gemeinsamkeiten, z. B. Zufallszahlengeneratoren und statistische Auswertungen. Aber auch in der Modellierung und Ablaufsteuerung selbst gibt es bewährte Prinzipien, die als Basis für die meisten Simulationsprogramme gewählt werden können.

Die Systeme zur diskreten, ereignisorientierten Simulation unterscheiden sich in der Beschreibung der Zustände und Zustandsübergänge. Vom Prinzip her sind die Systeme alle ereignisorientiert. Hauptsächlich werden folgende Typen unterschieden:

- ereignisorientiert:** z. B. SIMSCRIPT [Sie91],
OPNET [MIL97],
CNCL [JG93, JSBP96]
- prozeßorientiert:** z. B. SIMULA [BDMN73],
Demos [Bir79],
SIC [HHK+91],
MODSIM [BDM+89],
SimPlusPlus [Ros92]
- aktivitätsorientiert:** z. B. ECSL [Kre86, Pag91]
- transaktionsorientiert:** z. B. GPSS [BFS87]

Der *aktivitätsorientierte* Ansatz basiert auf der Zusammenfassung einer Menge von Operationen während eines Zeitintervalls, die als Aktivität bezeichnet wird. Zustandsänderungen werden dem Ende des Intervalls zugeordnet (z. B. ECSL, activity scanning) [Kre86, Pag91]. Das Prinzip der Fortschaltung von Ereignis zu Ereignis wird beibehalten, aber es gibt keine universelle Liste von Ereigniszeiten. Es können beliebig komplexe Bedingungen für Beginn und Ende deklariert werden. Diese müssen für alle Aktivitäten zu jedem Simulationszeitpunkt überprüft werden. Daraus resultieren Effizienzprobleme, so daß dieses Konzept nicht sehr verbreitet ist.

Typischer Vertreter für ein *transaktionsorientiertes* System ist GPSS [BFS87]. Die statischen Komponenten sind hier Blöcke, die permanent vorhanden sind. Die dynamischen Komponenten sind die sog. Transaktionen. GPSS-Programme werden durch ein Flußdiagramm von Blöcken beschrieben. Transaktionen bewegen sich durch diese Blöcke. Ein elementares Ereignis ist hier das Aufeinandertreffen von Transaktion und Block. Transaktionen entsprechen in etwa Prozessen. GPSS wird deshalb manchmal auch als prozeßorientiert bezeichnet.

Im folgenden wird der ereignisorientierte und prozeßorientierte Ansatz genauer beschrieben.

2.2 Ereignis- und prozeßorientierte Simulation

Ereignisse kennzeichnen die diskreten Zustandsänderungen eines modellierten Systems. Beispiele für Ereignisse sind: das Eintreffen eines Auftrags, z. B. eines Druckauftrags, der Start oder die Beendigung der Bedienung eines Auftrags oder der Ablauf einer Zeitbedingung, z. B. bei einem Kommunikationsprotokoll.

Ein *Prozeß* ist die logische Zusammenfassung einer wiederkehrenden Folge von zusammengehörenden *Ereignissen* und *Objekten*. Für einen Prozeß ist zu jedem Zeitpunkt immer maximal ein Ereignis geplant. Beispiele für Prozesse sind: ein Generator für Aufträge, eine Bedienstation oder spezielle Überwachungsprozesse.

Ein ereignis-/prozeßorientiertes System besteht aus:

- *statischen Komponenten*: permanente Objekte (z. B. Zufallszahlengenerator)
- *dynamischen Komponenten*:
 - Ereignisse/Ereignisroutinen bzw. Prozesse
 - temporäre Objekte (z. B. Aufträge).

Die *Ablaufsteuerung* besteht aus einer *Ereignis-/Prozeßliste*, die sequentiell abgearbeitet wird, indem von Ereignis zu Ereignis gesprungen wird, d. h. die sog. *tote* Zeit wird übersprungen, siehe Abbildung 2.2.

Ereignisroutinen bzw. Prozesse:

- ändern Attribute von Objekten: z. B. Länge oder Priorität einer Nachricht
- generieren oder löschen temporäre Objekte: z. B. Nachrichten oder Aufträge
- tragen Ereignisse/Prozesse in die Ereignis-/Prozeßliste ein:
 - z. B. Ankunft eines neuen Auftrags
- löschen Ereignisse/Prozesse aus der Ereignis-/Prozeßliste:
 - z. B. Zeitüberwachungen (*timer*), die nicht mehr benötigt werden

Wesentlich zum Verständnis der Simulationstechnik bei ereignisorientierter, diskreter Simulation ist die Unterscheidung von *Modellzeit* und *Rechenzeit*. In Abbildung 2.2 wird der Zusammenhang zwischen Modellzeit und Rechenzeit schematisch dargestellt. Ereignisse im Modell (E_i) sind z. B. die Ankunft eines Auftrags im System oder das Ende der Bearbeitung eines Auftrags. Ein Ereignis hat bezogen auf die Modellzeit *keine Zeitdauer* sondern nur einen *Ereignis-Zeitpunkt* (t_i). Die Bearbeitung eines solchen Ereignisses beansprucht jedoch *Rechenzeit* (T_S, T_{E_i}), z. B. zur Generierung von Zufallszahlen oder für die statistische Auswertung. Die Modellzeit wird auch *virtuelle Zeit* (*virtual time*) genannt im Unterschied zur Rechenzeit, die dann als *reale Zeit* (*real calendar time*) oder in dieser Arbeit auch als *Laufzeit* bezeichnet wird.

Das Verhältnis von Modellzeit und Rechenzeit ist abhängig vom Anwendungsfall. Die Rechenzeit kann um Größenordnungen kleiner als die Modellzeit sein, z. B. kann eine Modellzeit von einigen Stunden in wenigen Minuten Rechenzeit simuliert werden. Andererseits kann für komplexe Modelle, z. B. von Kommunikationsnetzen, die Rechenzeit auch ein Vielfaches der Modellzeit betragen, d. h. einige Sekunden oder Minuten Modellzeit benötigen eine Rechenzeit von Stunden oder Tagen.

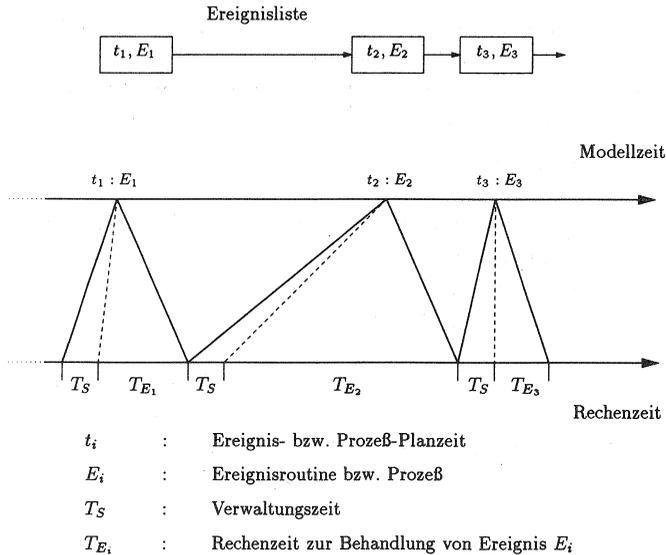


Abbildung 2.2: Modellzeit und Rechenzeit bei der ereignisorientierten Simulation

Im Unterschied zu den Ereignisroutinen eines *rein ereignisorientierten* Systems wird ein Prozeß dort fortgesetzt, wo er zuletzt unterbrochen wurde, d. h. er verhält sich wie eine *Koroutine*. Dieses Sprachkonzept findet man z. B. in der Sprache SIMULA und Modula-2. Eine Koroutine ist ein gleichrangiger, nebenläufiger Prozeß. Der Status des Prozesses mit allen Attributwerten muß bei einer Unterbrechung des Prozesses festgehalten werden.

Die prozeßorientierte Beschreibung liefert eine übersichtliche Struktur, da der Ablauf nicht in Ereignisroutinen aufgelöst werden muß. Die Beschreibung des Simulationsmodells kann sich nahe am realen Modell orientieren und eignet sich somit gut zur Darstellung von parallelen Vorgängen mit komplexen Interaktionen.

Ein Prozeß wird aktiviert (Prozeßzustand: aktiv) und führt (ohne Zeitverzug in der Modellzeit) Zustandsänderungen durch, d. h. die Simulationsuhr steht still. Dies entspricht der Ereignisroutine. Danach kann der Prozeß seine nächste Aktivierung planen oder passiv werden.

Abbildung 2.3 zeigt ein Wartesystem mit allgemeinem Ankunftsprozeß (G), allgemeinem Bedienprozeß (G), einer Bedienstation (1) und einer unbegrenzt großen Warteschlange, das abkürzend als G/G/1-Modell bezeichnet wird, siehe Abschnitt 3.4. Abbildung 2.4 zeigt eine mögliche prozeßorientierte Darstellung des Modells durch zwei Prozesse: *Generator* und *Server*. Der Ablauf der Prozesse ist zyklisch. Im Generator werden fortlaufend Aufträge generiert, die in die Warteschlange eingereicht werden. Nachdem ein Auftrag generiert wurde, reißt der Generator sich mit einer neuen Aktivierungs- bzw. Ankunftszeit in die Prozeßliste ein. Der Server bearbeitet die Aufträge zyklisch. Bei Warteschlangendisziplinen, die unabhängig von der Bedienzeit sind, kann die Bearbeitungszeit im Server bestimmt werden. Der Server wird deaktiviert und nach der Bearbeitung des Auftrags wieder aktiviert.

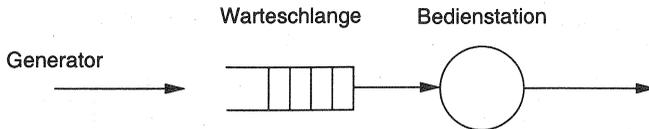
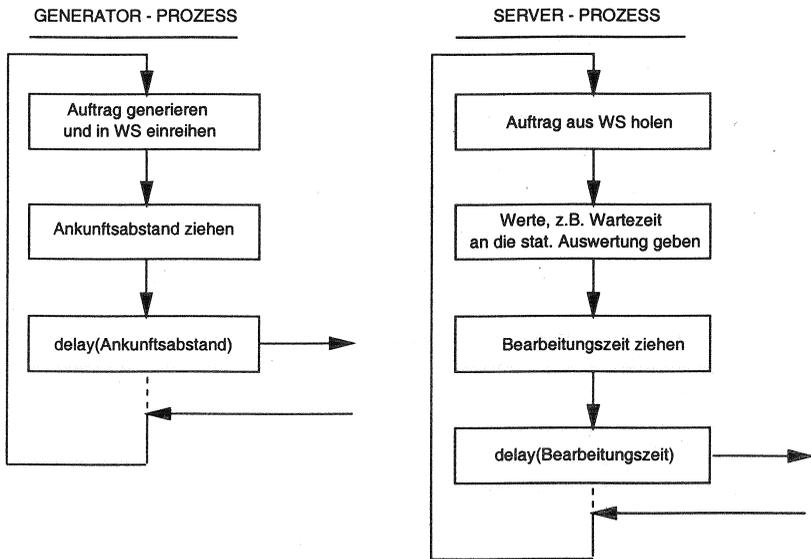


Abbildung 2.3: G/G/1-Modell mit Warteschlange

Auf elementarer Ebene liegen all diesen Konzepten immer Ereignisfolgen zugrunde, die einer modellinternen Simulationsuhr zugeordnet werden. Die Zeit zwischen zwei Ereignissen ist variabel. Treten zwei Ereignisse gleichzeitig auf, so muß eine Aussage über die Reihenfolge gemacht werden. Diese kann festliegen, z. B. in Reihenfolge der Generierung der Ereignisse, oder sie kann durch den Benutzer direkt oder in Form von Prioritäten spezifiziert werden.

Bei einfachen Modellen ist die ereignisorientierte und die prozeßorientierte Darstellung äquivalent. Beide Methoden sind übersichtlich und problemnah. Bei komplexen Modellen wird meist die prozeßorientierte Beschreibung vorgezogen. Eine geschickte Abbildung des rein ereignisorientierten Konzepts auf eine objektorientierte Sprache kann die Vorteile des prozeßorientierten Ansatzes jedoch weitgehend kompensieren und ist sogar etwas flexibler, siehe Abschnitt 2.4.



WS=Warteschlange

Mit delay(...) wird der Ablauf unterbrochen und der Prozeß wird in die Prozeßliste eingereiht.

Abbildung 2.4: Beispiel: G/G/1 als Prozeß-Modell

2.3 Anforderungen an höhere Programmiersprachen

Im folgenden werden die allgemeinen Anforderungen eines ereignis- bzw. prozeßorientierten Simulationssystems zusammengestellt.

Man benötigt folgende allgemeine Komponenten:

- Zustandsvariablen
- Simulationsuhr
- Ereignisroutinen/Prozesse
- Initialisierungsroutine/Initialisierungsprozeß
- Ereignis-/Prozeßliste
- Ereignis-/Prozeß-Scheduler
- Warteschlangen
- Statistische Auswertungsroutinen
- Zufallsgeneratoren

Allgemeine Anforderungen an Programmiersprachen, wie z. B. Strukturiertheit, Lesbarkeit und Testhilfen, gelten in besonderem Maße auch für Simulationssprachen. Zur Unterstützung der Programmierung von Simulationssystemen in einer allgemeinen, höheren Programmiersprache sind jedoch einige spezielle Forderungen zu stellen.

Objektorientierte Sprachen: Vorteilhaft für die Implementierung von Simulationssystemen sind modulare, insbesondere *objektorientierte Sprachen*. *SIMULA* ist als *erste objektorientierte Sprache* der Vorläufer aller objektorientierten Sprachen und wurde, wie der Name sagt, als Simulationssprache entworfen, siehe [BDMN73, SIM87, Sch90]. Das Koroutinenkonzept ist sehr komplex und leider waren die ersten Compiler nicht sehr effizient, aber *SIMULA* wird nach wie vor für Simulationsprogramme verwendet und ist auch Basis von Weiterentwicklungen, wie z. B. *HIT* [Bei91]. Auch die inzwischen sehr verbreitete objektorientierte Sprache *C++* [Str89] wurde als Simulationshilfsmittel entwickelt.

Effizienz: Die Laufzeit-Effizienz ist für Simulationssysteme wichtiger als bei vielen anderen Anwendungen. Simulationsexperimente sind aufwendig, insbesondere wenn gesicherte Ergebnisse gefordert werden. Dies ist ein wesentlicher Grund für die Wahl von *C++* für viele neuere Simulationssysteme, darunter *SIC* [GKS91] und *CNCL* [JG93].

Prozeßkonzept: Bei der Implementierung eines Simulationssystems mit einer höheren Programmiersprache ist die *Hauptschwierigkeit* die Realisierung des *Prozeßkonzeptes*. Eine direkte Abbildung ist nur bei Programmiersprachen mit Koroutinen möglich (z. B. SIMULA und Modula-2). Bei Sprachen wie C++ gibt es mehrere Möglichkeiten ein koroutinenartiges Verhalten von Prozeduren zu erzeugen:

- Abbildung der Koroutine auf eine Prozedur und Erzeugung von unterschiedlichen Einsprungstellen zur Kennzeichnung der Fortsetzungsstellen im Prozeß. Diese Struktur kann direkt vom Programmierer erzeugt werden (z. B. SIMPAS [TG89] oder AMADEUS [Wen91]) oder man wählt eine Vorübersetzung, um den Programmierer zu unterstützen (z. B. Lex-Programme für SIC [HHK⁺91]). Nachteil dieser Methode ist, daß die Prozesse einigen Einschränkungen unterliegen.
- Abbildung der Koroutinen auf Prozeduren und Manipulation der Speicherverwaltung (*Stack*), so daß sich die Prozedur wie eine Koroutine verhält (z. B. AT&T C++ [ATT90], SimPlusPlus [Ros92]). Diese Lösungen sind maschinenabhängig.

Zur Verwaltung von temporären Objekten wird eine *dynamische Speicherverwaltung* benötigt, die in objektorientierten Sprachen im allgemeinen vorhanden ist. Zur Erzeugung von Objekten stehen meist Operationen wie *new* oder *allocate* zur Verfügung. Zur Freigabe gibt es entsprechende Operationen, z. B. *delete*, *dispose* oder *deallocate*.

Sowohl zur Verwaltung von Ereignis- bzw. Prozeßlisten als auch zur Verwaltung von Warteschlangen werden geeignete Sprachkonzepte zur Verwaltung von Listen sowie Funktionen zum Einfügen, Entfernen und Verschieben von Ereignissen bzw. Prozessen benötigt.

Als ein Beispiel für ein Simulationssystem wird im folgenden Abschnitt eine auf C++ aufbauende Simulations-Klassenbibliothek beschrieben.

2.4 Eine objektorientierte Simulationsbibliothek

Als Beispiel für den prinzipiellen Aufbau einer Simulationsbibliothek und als Basis für die Struktur der Beispiele in späteren Kapiteln wird im folgenden eine Übersicht über die ereignisorientierte CNCL-Simulationsbibliothek¹ [JG93, JSBP96] gegeben. Allgemein herrscht die Meinung vor, daß prozeßorientierte Systeme den ereignisori-

¹CNCL=Communication Networks Class Library, Klassenbibliothek des Lehrstuhls Kommunikationsnetze (ftp.comnets.rwth-aachen.de) Diese Bibliothek, die als Basis fast aller Simulationsprojekte am Lehrstuhl Kommunikationsnetze benutzt wird, wurde von M. Junius und der Autorin entworfen und implementiert und von vielen Mitarbeitern des Lehrstuhls konsistent weiterentwickelt.

entierten Systemen, insbesondere zur Untersuchung großer Systeme, überlegen sind [LK91]. Die objektorientierten Techniken ermöglichen in der CNCL die Definition sog. *Event Handler*, die vom Abstraktionsgrad in etwa den Prozessen entsprechen. Der Nachteil, daß der Ablauf nicht sequentiell wie bei den Prozessen beschrieben werden kann, wird durch die Möglichkeit wieder ausgeglichen, daß für einen Event Handler mehr als ein Ereignis in der Ereignisliste stehen kann. Eine Station in einem Kommunikationsnetz, die sowohl Ankunftsereignisse generiert als auch bedient, kann z. B. durch einen Event Handler dargestellt werden, müßte aber in einem prozeßorientierten Modell auf mehrere Prozesse abgebildet werden, nämlich entsprechend der Anzahl möglicher paralleler Abläufe in einer Station, also mindestens auf einen Generator- und einen Server-Prozeß. Auch zur Abbildung von Zeitschranken (*timer*) sind Events und Event Handler gut geeignet.

Die CNCL wurde als allgemeine C++-Bibliothek mit folgenden Eigenschaften konzipiert: baumartige Klassenstruktur, Laufzeittypinformation und -überprüfung sowie allgemeine Containerklassen. Zur Unterstützung bei der Erstellung von Simulationsprogrammen wurden Klassen für die effiziente und flexible Generierung von Zufallszahlen, für die statistische Auswertung, zur Unterstützung einer rein ereignisorientierten Simulation und spezielle Konstrukte zur Implementierung von Strukturen der Spezifikationsprache SDL (*SDL – Specification and Description Language*) vorgesehen.

Aufbauend auf der CNCL werden Projektbibliotheken realisiert, z. B. die Bibliothek SIMCO3++ [RBGW94] mit den Komponenten für die spezifischen Aufgabenstellungen bei der Modellierung von Diensten und Protokollen aus dem Bereich mobiler Funknetze, z. B. die Modellierung von Funkausbreitung, Kanälen und Mobilität von Stationen, siehe Abschnitt 6.3.

Besonderes Augenmerk bei der Implementierung lag darauf, daß die Strukturen der formalen Spezifikationsprache SDL möglichst direkt auf die Konzepte der Simulationsbibliothek abgebildet werden können. Die Abbildung wird zusätzlich unterstützt durch die automatische Umsetzung der sprachlichen Form von SDL in ein CNCL-C++-Simulationsprogramm. Zur graphischen Darstellung der Ergebnisse als Animationssequenz dient die Schnittstelle zu GIST (*GIST – Graphic Interactive Simulation result Tool*) [PH⁺96], das sowohl einen direkten (*on-line*) als auch abgesetzten (*off-line*) Betrieb erlaubt, siehe Abbildung 6.26.

Im folgenden werden die zentralen Aspekte der Realisierung beschrieben, zunächst das allgemeine Konzept der Simulations- und Klassenbibliothek CNCL, dann die für die Simulation wichtigen Komponenten.

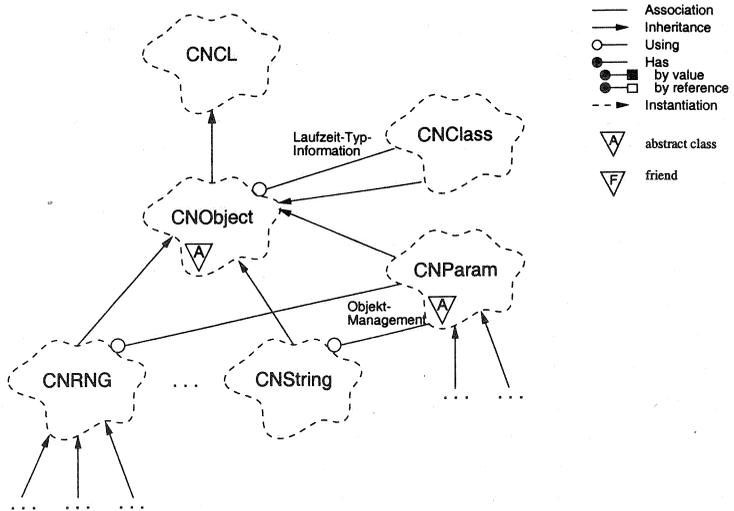


Abbildung 2.5: CNCL Basisklassen

Klassenbibliothek CNCL: Die CNCL-Klassenbibliothek besteht im wesentlichen aus den Basisklassen, die in Abbildung 2.5² zu sehen sind:

CNCL: Dies ist die oberste Klasse in der Hierarchie. Sie enthält im wesentlichen statische Elementfunktionen/Methoden (*member functions*) zur Einstellung von Parametern und zur Fehlerbehandlung.

CNObject: CNObject ist die eigentliche Wurzel der Hierarchie. Alle Klassen sind direkt oder indirekt von CNObject abgeleitet und daher mit CNObject typverträglich, was die Implementierung von generischen Containern ermöglicht.

CNClass: Zu jeder Klasse der CNCL-Hierarchie gibt es ein zugehöriges Beschreibungsobjekt vom Typ CNClass. Dies ermöglicht Laufzeittypinformation.

CNParam: Diese Klasse ist als generisches Parameterobjekt für eine zukünftige Objektverwaltung der CNCL-Klassenbibliothek vorgesehen.

Zufallszahlen: Die Zufallszahlenerzeugung basiert, wie in Kapitel 4 dargestellt wird, auf zwei verschiedenen Komponenten: den Basisgeneratoren und den Klassen zur Erzeugung von Zufallswerten mit vorgegebenen Verteilungen. Die Basisgeneratoren und Verteilungsgeneratoren der CNCL bauen auf den Konzepten der GNU-Bibliothek [GNU97] und der SIC-Bibliothek [GKS91] auf.

²Die Abbildungen verwenden die in [Boo94] vorgestellte Notation.

Basisgeneratoren erzeugen Basiszufallszahlen als gleichverteilte Werte im Bereich $(0, 1)$. Verteilungsklassen generieren aus den Basiszufallszahlen Werte mit einer bestimmten Verteilungsfunktion, z. B. normalverteilt oder negativ-exponentiell verteilt.

Basisgeneratoren und Verteilungen können beliebig miteinander kombiniert werden. Dies wird durch zwei getrennte Klassenhierarchien implementiert:

- CN RNG:** mit den abgeleiteten Klassen für die Basisgeneratoren
- CN Random:** mit den abgeleiteten Klassen für die Verteilungen

Implementiert sind derzeit 6 verschiedene Basisgeneratoren, u. a. auch der physikalische Generator PURAN 2 [Ric93], der auf einer Datei (CD-ROM) mit aus physikalischem Rauschen generierten Zufallszahlen basiert, und 20 verschiedene Verteilungen. Neue Generatoren und Verteilungen können problemlos in das Konzept eingebunden werden.

Das Konzept erlaubt auch die Verwendung *eines* Basisgeneratorexemplars für *verschiedene* Verteilungen. Damit erübrigt sich die Auswahl unterschiedlicher Startwerte für die verschiedenen Verteilungen, die bei ungeeigneter Wahl zu Korrelationen zwischen den einzelnen Generatoren und damit zur einer Verfälschung des Simulationsergebnisses führen können.

Statistische Auswertung: Im System CNCL sind zur Zeit drei Auswertelgorithmen vorgesehen, die als eigenständige Klassen implementiert sind. Alle können eine Folge von eindimensionalen Zufallszahlen auswerten:

CN Moments: Diese Klasse ist zur einfachen Überwachung von statistischen Größen gedacht und berechnet die wichtigsten Momente ohne Fehlermaß.

CN LRE: Diese Klasse bestimmt unter Benutzung des LRE-Algorithmus (LRE – *Limited Relative Error*) [GS90, Sch88a] die Verteilungsfunktion des beobachteten Zufallsprozesses und zugleich den lokalen Korrelationskoeffizienten als Maß der Abhängigkeit aufeinanderfolgender Werte. Ein weiterer Vorteil des Verfahrens ist die Einteilung des Wertebereichs der Verteilungsfunktion in äquidistante Abstände. Als relatives Fehlermaß dient ein spezieller lokaler Variationskoeffizient, siehe Abschnitt 5.5.

CN Batches: Diese Klasse liefert unter Benutzung der Batch-Means-Methode den Mittelwert und die Varianz mit zugehörigem Konfidenzintervall, siehe [Fis78, LK91, Kre73] und Abschnitt 5.4. Zusätzlich wird die Verteilungsfunktion gemäß einer vorgegebenen Klasseneinteilung ermittelt, wobei für jeden Punkt der Funktion ein Konfidenzintervall angegeben wird.

Für jede auszuwertende Folge von Meßwerten muß ein Objekt der gewünschten Klasse angelegt werden. Um auf alle für die statistische Auswertung wichtigen Klassen einheitlich zugreifen zu können, wurde die Basisklasse `CNStatistics` eingeführt.

Ereignisgesteuerte Simulation: Im Rahmen der CNCL wurde eine ereignisorientierte Simulationssteuerung implementiert, die nicht nur den Vorteil der konzeptuellen Klarheit hat, sondern auch relativ einfach und portabel in C++ zu implementieren ist. Ein Simulationsprogramm besteht aus mehreren sogenannten *Event Handlern*, die über den Austausch von als *Events* bezeichneten Ereignissen miteinander kommunizieren. Events werden über einen *Scheduler* verwaltet, der Events entgegennimmt und sie zur gegebenen (Modell-)Zeit weiterleitet.

Event Handler: Ein Event Handler ist ein Objekt, das Events empfängt und verarbeitet. Er ist ein Zustandsautomat, der abhängig von ankommenden Events den Zustand wechseln und neue Events generieren kann. Die Generierung von Events bildet die einzige Schnittstelle eines Event Handlers zum Scheduler und zu anderen Event Handlern. Andere Funktionen aus prozeßorientierten Systemen, die den Ablauf der Prozesse steuern, müssen über die explizite Generierung von Events realisiert werden. Für das eigentliche Simulationsprogramm müssen von der abstrakten Basisklasse `CNEventHandler` die problemspezifischen Event Handler abgeleitet werden. Der Zustand des Event Handlers wird lokal verwaltet. Ferner ist dem Event Handler die aktuelle Modellzeit und der aufrufende Scheduler bekannt.

Events: Events bilden die zentralen Objekte zur Kommunikation zwischen den Event Handlern. Events in CNCL verfügen über eine eindeutige Kennung, eine Modellzeit, zu der sie ausgeführt werden, einen vom Anwender festgelegten Typ und einen Zeiger auf ein beliebiges Objekt. Events sind jeweils von einem Event Handler (Absender) an einen zweiten Event Handler (Empfänger) adressiert. Um ein Event zu versenden, muß der Absender ein neues Event generieren, die entsprechenden Informationen, zu denen auch die geplante Ausführungszeit gehört, und die Zieladresse eintragen und das Event schließlich an den Scheduler übergeben, der die Sendezeit automatisch hinzufügt.

Scheduler: Der Scheduler (`CNEventScheduler`) realisiert die zentrale Steuerung der Simulation in CNCL. Er verwaltet die ankommenden Events in einer Liste und übergibt sie dann zur Ausführungszeit an den entsprechenden Event Handler. In der Liste werden die Events jeweils anhand der Modellzeit, zu der sie ausgeführt werden sollen, und anhand ihrer Priorität einsortiert. Die eigentliche Tätigkeit des Schedulers besteht darin, jeweils das nächste Event aus der Liste zu nehmen, die Modellzeit fortzuschreiben und den angesprochenen Event Handler aufzurufen. Der Scheduler verfügt außerdem über die Möglichkeit, Events aus der Liste zu löschen. Dies ist z. B. für Zeitbedingungen wichtig, wenn eine gesetzte Zeitbedingung durch ein vorher eintreffendes Event gestoppt werden soll.

SDL Spezifikationen: Die CNCL-Klassenbibliothek bietet als besonderen Vorteil, daß sich Spezifikationen der formalen Spezifikationsprache SDL direkt in die CNCL-Event-Strukturen umsetzen lassen. Da im Rahmen der Spezifikation von Protokollen in zunehmendem Maß SDL benutzt wird, bietet sich die unmittelbare Übersetzung der SDL-Spezifikation in CNCL-Klassen an, um Leistungsbewertungen solcher Protokolle vorzunehmen. Die grundlegenden SDL-Konstrukte können automatisch in einen CNCL Event Handler umgesetzt werden.

Implementierung in C++: Die ereignisgesteuerte Simulation wurde in den C++-Klassen `CNEEvent`, `CNEEventHandler` und `CNEEventScheduler` implementiert. Die interne Liste der Events im Scheduler, sortiert nach Ausführungszeit und Priorität, wird mit `CNEEventList` realisiert. Bild 2.6 zeigt diese Klassen und ihre Beziehungen im Überblick. `CNDLList` ist eine CNCL-Klasse zur Verwaltung doppelt verketteter Liste, die intern in `CNEEventList` benutzt wird. Ein Simulationsmodell wird im Regelfall auf einen Scheduler und mehrere Event Handler abgebildet, z. B. als Umsetzung mehrerer SDL-Prozesse. Diese Event Handler sind jeweils als eine von `CNEEventHandler` abgeleitete Klasse zu deklarieren, die eine virtuelle Funktion `event_handler()` mit der entsprechenden Teilfunktionalität des Simulationsmodells enthalten.

Beispiel G/G/1-Wartezimmer: Als Beispiel für die Anwendung der ereignisorientierten Simulation in CNCL wird im folgenden die Simulation des G/G/1-Warterraums vorgestellt. Zur Realisierung von Warterraummodellen stellt die CNCL eine Auftragsklasse (Job) und Warteschlangen mit verschiedenen Warteschlangendisziplinen zur Verfügung (z. B. FIFO – First In First Out, RANDOM – gleichverteilte zufällige Auswahl, SPT – Shortest Processing Time first, PRIO – Prioritätenverfahren).

In Anlehnung an die prozeßorientierte Lösung werden 2 Event Handler definiert:

Generator: Ein Event Handler dieses Typs erzeugt Aufträge (Jobs) und übergibt sie an den Server. Aufträge werden mit einer zufälligen Zwischenankunftszeit generiert. Der Generator hat nur einen Zustand.

Server: Ein Event Handler dieses Typs verwaltet ankommende Aufträge in einer Warteschlange und bearbeitet sie mit einer zufälligen Bearbeitungszeit. Ein Server hat zwei Zustände:

Waiting: Der Server ist frei und wartet auf einen Auftrag.

Serving: Der Server ist belegt und bearbeitet einen Auftrag.

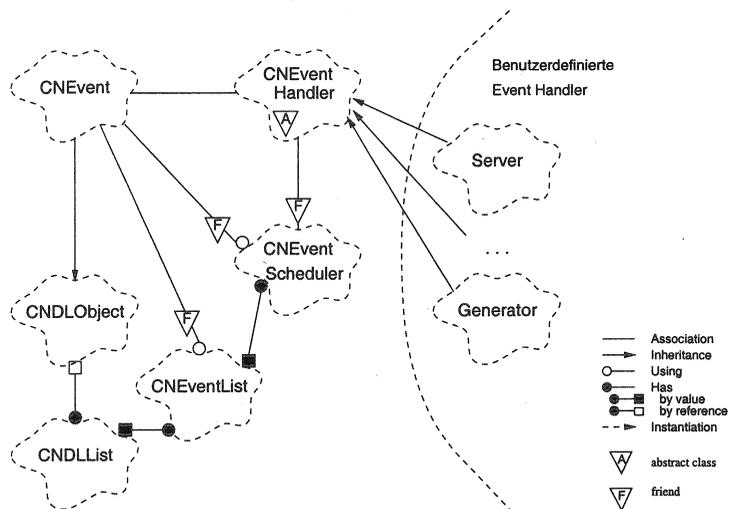


Abbildung 2.6: CNCL-Klassen für die ereignisgesteuerte Simulation

Für die Kommunikation der Event Handler werden folgende Event-Typen benötigt:

- TIMER_G:** Event zur Planung der nächsten Zwischenankunftszeit (Generator)
- TIMER_S:** Event zur Planung des Bearbeitungsendes eines Auftrags (Server)
- JOB:** Absenden eines Auftrags vom Generator zum Server

Im Simulationsprogramm werden jeweils ein Objekt `generator` und ein Objekt `server`, sowie ein Object `scheduler` erzeugt. Abbildung 2.7 zeigt das Zusammenspiel der Objekte im Überblick und die Abbildungen 2.8 und 2.9 zeigen die Struktur der Event Handler in einer SDL-ähnlichen Darstellung.

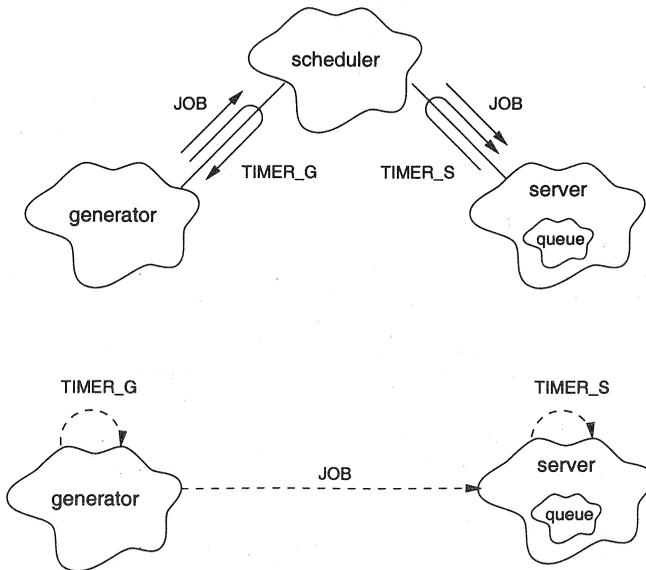


Abbildung 2.7: G/G/1-Simulation: Ereignisse

Event Handler Generator

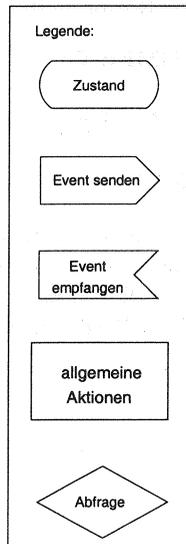


Abbildung 2.8: G/G/1-Simulation: CNCL Generator

Event Handler Server

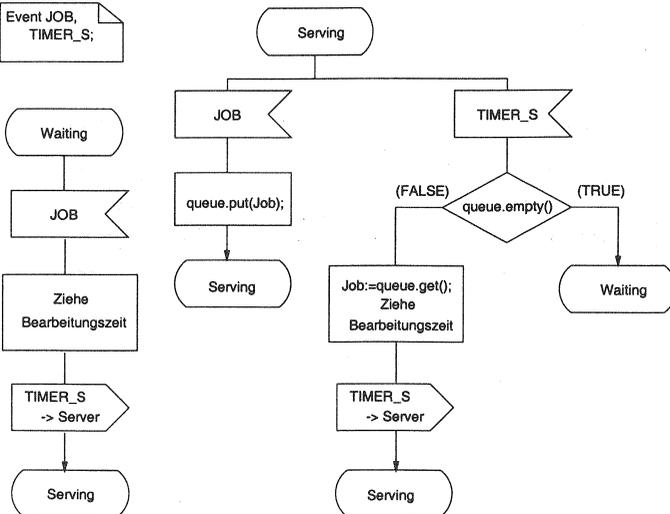
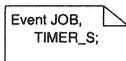


Abbildung 2.9: G/G/1-Simulation: CNCL Server

Verkehrsmodelle für Kommunikationsnetze

Verkehrsmodelle bilden die Grundlage der verschiedenen Verfahren zur Leistungsbewertung von Kommunikationsnetzen. Das Grundmodell für einen einzelnen Knoten eines Systems besteht aus einer Anzahl von Quellen, die konkurrierende Betriebsmittel-Anforderungen stellen. Die Zuteilung der limitierten Betriebsmittel oder auch das Warten auf Betriebsmittel wird von einer Betriebsmittelverwaltung durchgeführt, siehe Abbildung 3.1. Im allgemeinen ist die Zahl der Betriebsmittel kleiner als die mögliche Zahl gleichzeitiger Anforderungen der Verkehrsquellen, so daß es zu Engpässen kommt. Ziel der Verkehrstheorie ist es, gute – nach Möglichkeit optimale – Strategien zu finden, die einerseits die Betriebsmittel gut auslasten und andererseits akzeptable Leistungsgrößen aus der Sicht der einzelnen Anforderungen bieten und diese zu berechnen. Modelle für mehrstufige Systeme können aus diesen Grundmodellen zusammengesetzt werden, siehe Abschnitt 3.5.

Beispiele, die auf diesem Grundmodell aufbauen, sind:

- Telekommunikationssysteme mit den Fernsprechteilnehmern als Verkehrsquellen, den Leitungen und Vermittlungen als Betriebsmitteln und anderen Teilnehmern als Verkehrssenken
- Rechnersysteme mit Terminals als Verkehrsquellen/-senken und Prozessoren als Betriebsmitteln

Sowohl bei analytischen Berechnungen als auch bei Untersuchungen mit Hilfe stochastischer Simulationstechniken werden zur Beschreibung des Modells Zufallsgrößen und Zufallsprozesse eingesetzt, siehe Abschnitt 3.1. Ein wichtiges Hilfsmittel bei der Beschreibung und Berechnung von Verkehrsmodellen ist die Markoff-Kette, die in Abschnitt 3.2 beschrieben wird. Abhängigkeiten bzw. Korrelationen von Ereignissen innerhalb einer Zufallssequenz, wie sie in realen Systemen auftreten und deshalb modelliert werden müssen, werden in Abschnitt 3.3 betrachtet. Elementare Modelle für Systeme mit Wartevorgängen werden in Abschnitt 3.4 beschrieben. Den Abschluß dieses Kapitels bildet eine Übersicht über Wartennetzmodelle in Abschnitt 3.5.

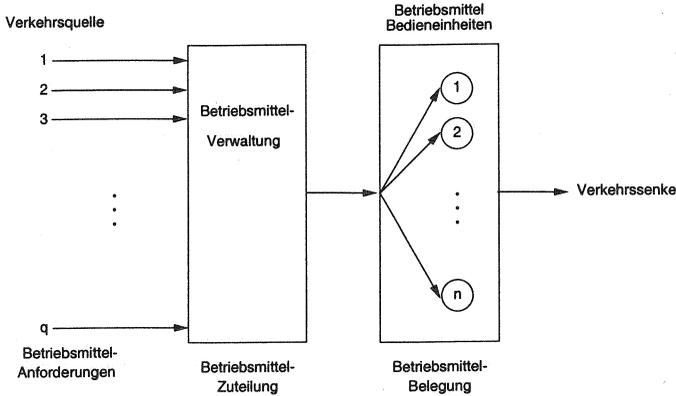


Abbildung 3.1: Grundmodell der Verkehrstheorie

3.1 Zufallsgrößen und Zufallsprozesse

Zufallsgrößen werden mit Hilfe von Verteilungen und Kennwerten beschrieben. Im folgenden werden die wichtigsten Definitionen zusammen mit ihren Formelsymbolen eingeführt, siehe [Sac92, Kle75, TG96, Wal96a].

3.1.1 Kennwerte und Verteilungen

Die *Verteilungsfunktion* (VF) $F_X(x)$ einer diskreten oder kontinuierlichen Zufallsvariablen X ist als die Wahrscheinlichkeit $P\{X \leq x\}$ für das Ereignis $\{X \leq x\}$ definiert:

$$F_X(x) = F(x) = P\{X \leq x\} \quad \text{für} \quad -\infty < x < \infty \quad (3.1)$$

Analog wird die *komplementäre Verteilungsfunktion* (KVF) $G_X(x)$ als Wahrscheinlichkeit für das Ereignis $\{X > x\}$ definiert:

$$G_X(x) = G(x) = 1 - F_X(x) = P\{X > x\} \quad \text{für} \quad -\infty < x < \infty \quad (3.2)$$

Beispielsweise wird die Bediendauer eines Auftrags als Zufallsvariable beschrieben. Die Zufallsvariable als Index kann entfallen, wenn der Kontext eindeutig ist.

Diskrete Zufallsvariable: Die *Wahrscheinlichkeitsfunktion* (WF) $P(i)$ einer diskreten Zufallsvariablen X , d. h. einer Zufallsvariablen, deren Wertebereich höchstens abzählbar viele Werte enthält, entspricht den Wahrscheinlichkeiten p_i der einzelnen Ereignisse. Ist der Wertebereich endlich, so läuft der Index i bis zu einem Maximal-

wert. Über die Summe der Wahrscheinlichkeiten p_i erhält man die Verteilungsfunktion $F_X(x)$:

$$\left. \begin{aligned} p_i &= P(i) = P\{X = i\}, \quad i = 0, 1, \dots, \infty \\ \text{Vollständigkeitsbedingung: } \sum_{i=0}^{\infty} p_i &= 1 \\ F_X(x) &= \sum_{i \leq x} p_i \quad \text{für alle } -\infty < x < \infty \end{aligned} \right\} \quad (3.3)$$

Kontinuierliche Zufallsvariable: Kontinuierliche Zufallsvariable werden mit Hilfe einer sog. *Dichtefunktion* beschrieben, wobei die Dichtefunktion selbst keine Wahrscheinlichkeitswerte definiert. Die Integration über die Dichtefunktion führt zu Wahrscheinlichkeitswerten für die entsprechenden Intervalle. Die *Verteilungsdichtefunktion* oder *Dichtefunktion* (DF) $f_X(x)$ einer kontinuierlichen Zufallsvariablen X kann als erste Ableitung der Verteilungsfunktion $F_X(x)$ angegeben werden:

$$\left. \begin{aligned} f_X(x) &= \frac{d}{dx} F_X(x) \\ \text{Vollständigkeitsbedingung: } \int_{-\infty}^{\infty} f_X(t) dt &= 1 \\ \text{und } F_X(x) &= \int_{-\infty}^x f_X(t) dt \end{aligned} \right\} \quad (3.4)$$

Für kontinuierliche Zufallsvariablen ist keine Wahrscheinlichkeitsfunktion definiert. Eine diskrete Zufallsvariable kann als kontinuierliche Zufallsvariable dargestellt werden, indem die Verteilungsdichtefunktion als Summe von gewichteten Dirac-Stößen [Pap91] beschrieben wird:

$$f_X(x) = \sum_{i=0}^{\infty} p_i \delta(x - i) \quad (3.5)$$

Dies erlaubt im folgenden die einheitliche Verwendung von Verteilungsdichtefunktionen zur Beschreibung sowohl kontinuierlicher als auch diskreter Zufallsvariablen.

Da im Rahmen der Verkehrstheorie fast ausschließlich positive Zufallsvariable betrachtet werden, wird der Wertebereich der Zufallsvariablen im folgenden auf $[0, \infty)$ eingeschränkt.

Kennwerte von Verteilungen: Der *Mittelwert* μ bzw. *Erwartungswert* $E[X]$ einer Zufallsvariablen X ist der Schwerpunkt der Verteilung:

$$\mu = \mu_X = E[X] = \sum_{i=0}^{\infty} i \cdot p_i \quad \text{bzw.} \quad \int_0^{\infty} x f_X(x) dx \quad (3.6)$$

Die *Varianz* wird mit σ_X^2 oder $\text{Var}[X]$ bezeichnet und ist ein Maß für die Variabilität einer Zufallsvariablen um ihren Mittelwert:

$$\sigma_X^2 = \text{Var}[X] = \sum_{i=0}^{\infty} (i - \mu_X)^2 p_i \quad \text{bzw.} \quad \int_0^{\infty} (x - \mu_X)^2 f_X(x) dx \quad (3.7)$$

σ_X wird als *Streuung* oder *Standardabweichung* bezeichnet.

Eine dimensionslose Größe zur Beschreibung der Variabilität ist der sog. quadratische *Variationskoeffizient* c_X :

$$c_X = \frac{\sigma_X}{\mu_X} \quad \text{für} \quad \mu_X > 0 \quad (3.8)$$

Momente bzw. *Nullmomente* $X^{(k)}$ für $k = 0, 1, \dots$ sind Erwartungswerte von X^k :

$$X^{(k)} = \text{E}[X^k] = \sum_{i=0}^{\infty} i^k \cdot p_i \quad \text{bzw.} \quad \int_0^{\infty} x^k f_X(x) dx \quad (3.9)$$

Zentralmomente beschreiben die Abweichungen k -ter Ordnung für $k = 0, 1, \dots$ vom Mittelwert μ_X :

$$\text{E}[(X - \mu_X)^k] = \sum_{i=0}^{\infty} (i - \mu_X)^k \cdot p_i \quad \text{bzw.} \quad \int_0^{\infty} (x - \mu_X)^k f_X(x) dx \quad (3.10)$$

Zur Umrechnung zwischen σ_X , $X^{(2)}$ und c_X können folgende Beziehungen benutzt werden:

$$\left. \begin{aligned} \sigma_X^2 &= X^{(2)} - \mu_X^2 = \mu_X^2 c_X^2 \\ X^{(2)} &= \sigma_X^2 + \mu_X^2 = \mu_X^2 (c_X^2 + 1) \\ c_X^2 &= \frac{\sigma_X^2}{\mu_X^2} = \frac{X^{(2)}}{\mu_X^2} - 1 \end{aligned} \right\} \quad (3.11)$$

Wichtige Verteilungen: Im Anhang A werden wichtige Verteilungen bei der Bewertung von Verkehrsmodellen mit ihren Kennwerten zusammengestellt, siehe [HP75], [LK91].

In der Verkehrstheorie sind für die Beschreibung von allgemeineren Verteilungen insbesondere die sog. *Phasenmodelle* von Interesse, die sich aus parallelen oder seriellen exponentiellen Phasen zusammensetzen. Den allgemeinen Fall zeigt Abbildung 3.2, wobei die Verzweigungswahrscheinlichkeiten durch p_i bezeichnet werden. Im allgemeinen gilt $p_0 = 1$. Mit $p_0 \neq 1$ können auch sog. *entartete* Verteilungen dargestellt werden. Die Phase wird durch die Rate μ_i eindeutig festgelegt. Zwei Spezialfälle sind die Erlang- k -Verteilung, siehe Abbildung 3.3, und die Hyperexponentielle Verteilung, siehe Abbildung 3.4.

Die Bedeutung dieser Modelle liegt darin, daß man beliebige Verteilungen durch eine geeignete Kombination von Phasen beschreiben kann [Mor58] und Wartemodelle in dieser Form einfacher zu behandeln sind.

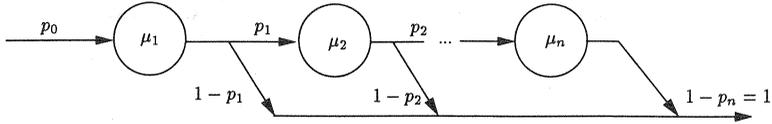


Abbildung 3.2: Allgemeine Phasenverteilung

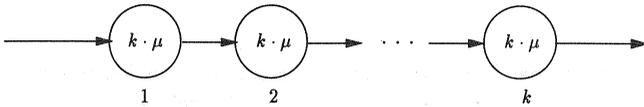


Abbildung 3.3: Erlang-k-Verteilung

Transformationen von Verteilungen: Der Umgang mit Zufallsvariablen läßt sich teilweise durch Transformationen wesentlich vereinfachen. Transformationen beschreiben eine Verteilung vollständig. Im folgenden werden die bei der Behandlung verkehrstheoretischer Modelle häufig benutzten Transformationen der Wahrscheinlichkeitsfunktion für diskrete, positive Zufallsvariablen bzw. Dichtefunktion für kontinuierliche, positive Zufallsvariablen angegeben.

Im diskreten Fall wird häufig die *erzeugende Funktion* $EF_X(z)$ benutzt:

$$EF_X(z) = \sum_{i=0}^{\infty} p_i \cdot z^i \tag{3.12}$$

Im kontinuierlichen Fall benutzt man für positive Zufallsvariablen $X \geq 0$ die *Laplace-Transformation* oder *Laplace-Stieltjes-Transformation (LST)*:

$$L_X(s) = \int_0^{\infty} f_X(x) \cdot e^{-sx} dx \tag{3.13}$$

Statistische Unabhängigkeit von Zufallsvariablen: Zwei Zufallsvariablen X und Y heißen *statistisch unabhängig*, wenn die *Verbunddichtefunktion* $f_{XY}(x, y)$ gleich dem Produkt der Rand-Verteilungsdichtefunktionen f_X und f_Y ist:

$$f_{XY}(x, y) = f_X(x) \cdot f_Y(y) \tag{3.14}$$

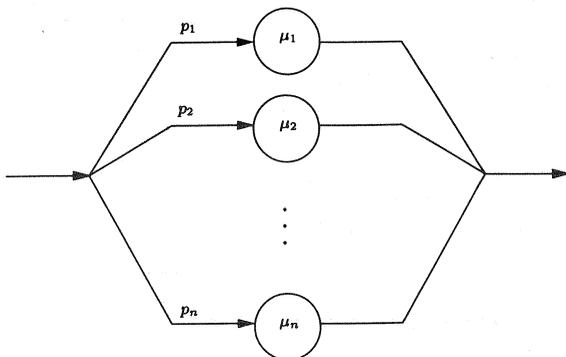


Abbildung 3.4: Hyperexponentielle Verteilung

Abhängigkeitsmaße: Zur Beschreibung der statistischen Abhängigkeit zweier Zufallsvariablen X und Y wird die *Kovarianz* $\text{Cov}(X, Y)$ definiert:

$$\left. \begin{aligned} \text{Cov}(X, Y) &= E[(X - \mu)(Y - \nu)] = E[XY] - \mu\nu \\ \text{mit } E[X] &= \mu \text{ und } E[Y] = \nu \\ \text{Cov}(X, Y) &= \text{Cov}(Y, X) \\ \text{für } X = Y &\text{ gilt } \text{Cov}(X, X) = \sigma_X^2 \end{aligned} \right\} \quad (3.15)$$

Die Zufallsvariablen X und Y heißen *unkorreliert*, falls $\text{Cov}(X, Y) = 0$ gilt. Ist $\text{Cov}(X, Y) > 0$ so spricht man von *positiver Korrelation* und entsprechend bei $\text{Cov}(X, Y) < 0$ von *negativer Korrelation*. Aus der statistischen Unabhängigkeit von X und Y folgt die Unkorreliertheit, aber nicht notwendig umgekehrt.

Da die Kovarianz $\text{Cov}(X, Y)$ nicht dimensionslos ist, wird der sog. *Korrelationskoeffizient* ϱ_{XY} definiert:

$$\left. \begin{aligned} \varrho_{XY} = \text{Corr}(X, Y) &= \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y} \\ \text{mit } -1 &\leq \varrho_{XY} \leq 1 \end{aligned} \right\} \quad (3.16)$$

Summen von Zufallsvariablen: Die Verteilungsdichtefunktion $f_Z(z)$ der Summenzufallsvariable $Z = X + Y$ als Summe zweier statistisch unabhängiger Zufallsvariablen X und Y ergibt sich als *Faltungintegral*, welches sich im transformierten

Bereich zu einem Produkt vereinfacht:

$$\begin{aligned} f_Z(z) &= \int_{-\infty}^{+\infty} f_X(z-y)f_Y(y) dy \\ L_Z(s) &= L_X(s) \cdot L_Y(s) \end{aligned} \quad (3.17)$$

Für den Mittelwert der Summe gilt allgemein, auch wenn X und Y statistisch abhängig sind:

$$E[Z] = E[X] + E[Y] \quad (3.18)$$

Für die Varianz der Summe gilt allgemein:

$$\text{Var}[Z] = \text{Var}[X] + \text{Var}[Y] + 2 \cdot \text{Cov}(X, Y) \quad (3.19)$$

Für statistisch unabhängige bzw. unkorrelierte Zufallsvariablen X, Y gilt:

$$\text{Var}[Z] = \text{Var}[X] + \text{Var}[Y] \quad (3.20)$$

3.1.2 Stochastische Prozesse

Ein *stochastischer Prozeß* ist eine Familie $\{X(t), t \in T\}$ von Zufallsvariablen $X(t)$ über einem Parameterbereich T . Der Parameterbereich T kann eine diskrete (endliche oder unendliche) Menge oder eine Teilmenge der reellen Zahlen sein. In verkehrstheoretischen Anwendungen ist der Parameterbereich meist die *Zeit*. Wertebereiche sind z. B. die natürlichen Zahlen, die reellen Zahlen oder Vektoren. Die Elemente des Wertebereichs von $X(t)$ nennt man auch *Zustände*.

Ein stochastischer Prozeß $X(t)$ heißt *stationär* (im strengen Sinne), wenn er statistisch invariant gegenüber Zeitverschiebungen ist, d. h. daß die Prozesse

$$X(t) \text{ und } X(t + \varepsilon)$$

für jedes ε statistisch gleich sind. Für die Dichtefunktion n -ter Ordnung eines stationären Prozesses gilt dann:

$$f(x_1, \dots, x_n; t_1, \dots, t_n) = f(x_1, \dots, x_n; t_1 + \varepsilon, \dots, t_n + \varepsilon) \text{ für jedes } \varepsilon$$

Die *Autokorrelation* $R(t_1, t_2)$ eines stochastischen Prozesses wird als Erwartungswert des Produkts $X(t_1) \cdot X(t_2)$ definiert:

$$R(t_1, t_2) = E[X(t_1) \cdot X(t_2)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x_1 x_2 f(x_1, x_2; t_1, t_2) dx_1 dx_2 \quad (3.21)$$

Entsprechend zu Gleichung (3.15) wird die *Autokovarianz* als Kovarianz der Zufallsvariablen $X(t_1)$ und $X(t_2)$ definiert:

$$C(t_1, t_2) = \text{Cov}(X(t_1), X(t_2)) = R(t_1, t_2) - E[X(t_1)] \cdot E[X(t_2)] \quad (3.22)$$

Ein stochastischer Prozeß heißt *schwach stationär*, wenn der Erwartungswert konstant und die Autokorrelation nur von der Zeitdifferenz τ abhängt:

$$\begin{aligned} E[X(t)] &= \mu = \text{const} \\ E[X(t + \tau)X(t)] &= R(\tau) \end{aligned} \quad (3.23)$$

3.2 Markoff-Ketten

Markoff-Ketten bilden eine wichtige Grundlage bei der Beschreibung und Untersuchung von Wartesystemen. Vereinfacht läßt sich eine Markoff-Kette als Zustandsraum beschreiben, wobei der zeitlich nachfolgende Zustand *nur* vom augenblicklichen Zustand nicht aber von weiter zurückliegenden Zuständen abhängig ist.

Eine präzisere Definition verlangt die Erklärung der Begriffe: *Markoff-Eigenschaft* und *Markoff-Prozeß*.

Ein stochastischer Prozeß besitzt die sog. *Markoff-Eigenschaft*, wenn zu einem Zeitpunkt $t_0 \in T$ das Verhalten des Prozesses für $t > t_0$ bei bekanntem Zustand zum Zeitpunkt t_0 nicht vom Verhalten des Prozesses vor dem Zeitpunkt t_0 abhängt. Als Synonym für die Markoff-Eigenschaft wird häufig der Begriff *Gedächtnislosigkeit* verwendet.

Ein *Markoff-Prozeß* ist ein stochastischer Prozeß, der zu jedem Zeitpunkt die Markoff-Eigenschaft besitzt.

Eine *Markoff-Kette* (*Markov chain*) in diskreter bzw. stetiger Zeit ist ein Markoff-Prozeß mit abzählbarem Zustandsraum, siehe Abbildung 3.5. Ist der Zeitbereich diskret, so heißt der Prozeß *Markoff-Kette in diskreter Zeit*, und ist er ein Teilbereich der reellen Zahlen, so heißt er *Markoff-Kette in stetiger Zeit*. Eine Markoff-Kette heißt *homogen*, falls die Zustandsübergangswahrscheinlichkeiten nicht vom Zeitparameter abhängen.

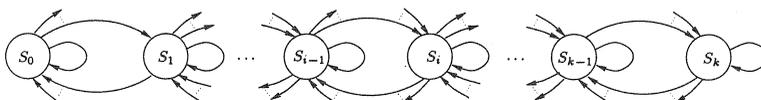


Abbildung 3.5: Markoff-Kette

3.2.1 Kenngrößen

Eine Markoff-Kette in diskreter Zeit wird beschrieben durch die *Übergangswahrscheinlichkeiten* erster Ordnung, die den Übergang "nach Zustand j von Zustand i " in einem Schritt der Dauer Δt beschreiben:

$$p_{ji} = P\{j|i\} \quad (3.24)$$

mit der Vollständigkeitsbedingung $\sum_{j=0}^k p_{ji} = 1$

Dies entspricht der Reihenfolge der Indizes als bedingte Wahrscheinlichkeit und vereinfacht die Formeln in Matrixschreibweise. In der Literatur werden die Indizes häufig in umgekehrter Reihenfolge definiert.

$P_j(m)$ bezeichnet die Wahrscheinlichkeit für das Erreichen des Zustands j in m Zeitschritten. Die Bestimmungsgleichungen für $P_j(m)$ für $j = 1, 2, \dots, k$ und $m = 1, 2, \dots$

$$P_j(m) = \sum_{i=0}^k p_{ji} P_i(m-1) = \sum_{i=0}^k p_{ji}(m) P_i(0) \quad (3.25)$$

können in Matrixschreibweise folgendermaßen dargestellt werden:

$$[P_j(m)] = [p_{ji}] [P_i(m-1)] = [p_{ji}]^m [P_i(0)] \quad (3.26)$$

wobei $[p_{ji}]$ als *Übergangsmatrix erster Ordnung* definiert ist.

$$[p_{ji}] = \begin{pmatrix} p_{00} & p_{01} & \dots & p_{0k} \\ p_{10} & p_{11} & \dots & p_{1k} \\ \vdots & \vdots & \ddots & \vdots \\ p_{k0} & p_{k1} & \dots & p_{kk} \end{pmatrix} \quad \text{mit} \quad \sum_{j=0}^k p_{ji} = 1 \quad (3.27)$$

Die Potenz $[p_{ji}]^m$ in Gleichung (3.26) stellt die *Übergangsmatrix m -ter Ordnung* dar:

$$[p_{ji}]^m = \begin{pmatrix} p_{00}(m) & p_{01}(m) & \dots & p_{0k}(m) \\ p_{10}(m) & p_{11}(m) & \dots & p_{1k}(m) \\ \vdots & \vdots & \ddots & \vdots \\ p_{k0}(m) & p_{k1}(m) & \dots & p_{kk}(m) \end{pmatrix} \quad (3.28)$$

$$\text{mit} \quad \sum_{j=0}^k p_{ji}(m) = 1 \quad \text{und} \quad p_{ji}(1) \equiv p_{ji} \quad (3.29)$$

die die *Übergangswahrscheinlichkeiten m -ter Ordnung* $p_{ji}(m) = P(j|i; m)$ enthält für den Übergang "nach Zustand j von Zustand i in m Zeitschritten Δt ". Diese Wahrscheinlichkeiten $p_{ji}(m)$ erhält man durch $(m-1)$ -fache Anwendung der Gleichung

$$p_{ji}(r) = \sum_{h=0}^k p_{jh} p_{hi}(r-1) \quad \text{mit} \quad r = 2, 3, \dots, m \quad \text{und} \quad p_{hi}(1) \equiv p_{hi} \quad (3.30)$$

Die *stationären Zustandswahrscheinlichkeitsvektoren*

$$\lim_{m \rightarrow \infty} [P_j(m)] = [P_j] = (P_0, P_1, \dots, P_k) \quad (3.31)$$

können aus folgendem Gleichungssystem gewonnen werden:

$$0 = \sum_{i=0}^k p_{ji} P_i - P_j \quad \text{für } j = 1, 2, \dots, k \quad \text{und} \quad P_0 = 1 - \sum_{j=1}^k P_j \quad (3.32)$$

Der Markoff-Prozeß, bei dem in der zugehörigen Markoff-Kette nur Übergänge zu Nachbarzuständen erfolgen, heißt *Geburt- und Sterbeprozess* (G&S-Prozeß). Von einem Zustand S_j sind dabei nur Übergänge zu S_{j-1} , S_j , und S_{j+1} erlaubt. Dieser Prozeß liegt der Analyse von Exponential-Warteräumen zugrunde.

3.2.2 Die 2-gliedrige Markoff-Kette

In diesem Abschnitt werden die wichtigsten Eigenschaften der 2-gliedrigen Markoff-Kette zusammengestellt, da dieser Spezialfall eine zentrale Rolle bei der Bestimmung des Korrelationsmaßes für die statistische Auswertung mit dem LRE-Algorithmus, siehe Abschnitt 5.5, und bei der Erzeugung von korrelierten Zufallszahlen, siehe Abschnitt 3.3.2, spielt.

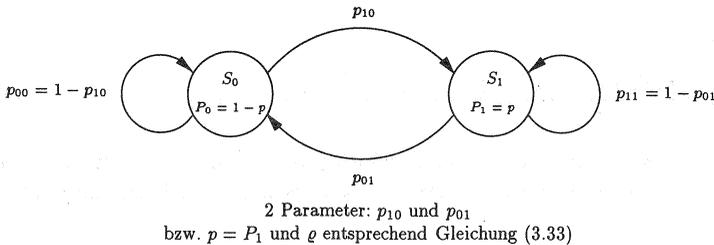


Abbildung 3.6: 2-gliedrige Markoff-Kette

Die Zustandsfolge der 2-gliedrigen Markoff-Kette kann entsprechend der Bezeichnungen der Zustände S_0 und S_1 auf eine 0/1-Sequenz abgebildet werden. Die Parameter der 2-gliedrigen Markoff-Kette können durch $p_{10} = P(1|0)$ und $p_{01} = P(0|1)$ oder äquivalent durch die beiden Parameter ϱ and p beschrieben werden:

$$\left. \begin{aligned} \varrho &= 1 - (p_{10} + p_{01}) = p_{00} + p_{11} - 1 \\ \varrho_{\min}(p) &\leq \varrho \leq 1 \quad \text{bzw.} \quad \varrho_{\min}(p) \leq 0 \\ p &= p_{10}/(p_{10} + p_{01}) = p_{10}/(1 - \varrho) \end{aligned} \right\} \quad (3.33)$$

wobei ϱ der (Auto-)Korrelationskoeffizient erster Ordnung der 0/1-Sequenz und $p = P_1$ bzw. $1 - p = P_0 = p_{01}/(1 - \varrho)$ die stationäre Wahrscheinlichkeit für den Zustand

S_0 bzw. S_1 ist. Die Wahrscheinlichkeitsfunktion der Bernoulli- oder Binär-Verteilung nach Gleichung (A.13):

$$P(x) := P_x = p^x(1-p)^{1-x} \quad \text{für } x = 0, 1 \quad (3.34)$$

beschreibt die stationären Wahrscheinlichkeiten der korrelierten 0/1-Sequenz der 2-gliedrigen Markoff-Kette.

Die Gleichung für $\rho = p_{10}/(p_{10}+p_{01})$ folgt direkt aus Gleichung (3.32). Die Herleitung von ρ ergibt sich nach Gleichung (3.16) aus:

$$\rho = \text{Corr}(X_n, X_{n+1}) = \frac{\text{Cov}(X_n, X_{n+1})}{\sigma_{X_n} \sigma_{X_{n+1}}}$$

mit

$$E[X_n] = E[X_{n+1}] = \mu = p \quad (3.35)$$

$$\sigma_{X_n}^2 = \sigma_{X_{n+1}}^2 = \sigma^2 = p(1-p)$$

$$\text{Cov}(X_n, X_{n+1}) = E[X_n X_{n+1}] - \mu^2$$

$$E[X_n X_{n+1}] = p \cdot p_{11} = p(1-p_{01})$$

Dies ergibt

$$\rho = \frac{E[X_n X_{n+1}] - \mu^2}{\sigma^2} = \frac{p(1-p_{01}) - p^2}{p(1-p)} = 1 - (p_{10} + p_{01}) \quad (3.36)$$

Positive Werte von ρ sind im ganzen Bereich $0 \leq \rho \leq 1$ möglich, wohingegen negative Werte von ρ auf einen Bereich $\rho_{\min}(p) \leq \rho \leq 0$ beschränkt sind, der von p abhängig ist:

$$\rho_{\min}(p) = \begin{cases} -p/(1-p) & \text{für } 0 \leq p \leq 1/2 \\ -(1-p)/p & \text{für } 1/2 \leq p \leq 1 \end{cases} \quad (3.37)$$

Die folgenden Spezialfälle sind von Interesse:

- Der Fall $\rho = 0$ führt zu $p_{10} + p_{01} = 1$, $p = p_{10} = p_{11}$ und $1-p = p_{01} = p_{00}$, d. h. der Folgezustand ist vom momentanen Zustand unabhängig, so daß man die *unabhängige* 0/1-Sequenz der binären Verteilung erhält.
- Der Fall $p_{10} = p_{01}$ führt zu einer *gleichverteilten* korrelierten 0/1-Sequenz mit $p = 1/2$ und $\rho = 1 - 2p_{10}$.

Dies enthält zwei deterministische Grenzfälle. $p_{10} = 0$ führt zu $\rho = +1$ und damit abhängig vom Startzustand "0" oder "1" zu einer Sequenz von konstanten Werten (0, 0, 0, ... oder 1, 1, 1, ...). $p_{10} = 1$ führt zu $\rho = -1$ und damit zu einer alternierenden Sequenz (... , 0, 1, 0, 1, 0, 1, ...).

3.3 Korrelation

Ein Hauptproblem der Statistik und damit insbesondere auch von Simulationsuntersuchungen ist die adäquate Beschreibung von Korrelationsphänomenen. Dies gilt sowohl für die Beschreibung von Zufallsprozessen als Eingabegröße eines Modells als auch für die Auswertung von Simulationsdaten [BFS87]. Im folgenden wird ein neues Maß – der sog. *lokale Korrelationskoeffizient* $\rho(x)$ [Sch87a, DS90, PS93] – zur Beschreibung der Autokorrelation¹ definiert, welches im LRE-Algorithmus Abschnitt 5.5 zur Steuerung von Simulationsläufen benutzt wird.

3.3.1 Der lokale Korrelationskoeffizient

Das Attribut *lokal* wurde in [DS90] hinzugefügt, um zu betonen, daß sich dieser Koeffizient als Funktion des Zufallswertes x der zugehörigen 2-gliedrigen Markoff-Kette vom bekannten *globalen* Korrelationskoeffizienten ρ unterscheidet, der eine Art Mittelwert bezogen auf alle Werte der Zufallsvariablen darstellt.

$\rho(x)$ ist ein relatives Maß mit dem Wertebereich $-1 \leq \rho(x) \leq 1$, welches als Korrelationsmaß erster Ordnung eines Markoff-Systems definiert ist [DS90, Sch94, Her94].

Dieser lokale Korrelationskoeffizient $\rho(x)$ stellt eine wohldefinierte Korrelationsfunktion erster Ordnung dar, die sowohl für die Belegung bzw. Ankunftsbelegung des elementaren Wartesystems M/M/1/N, siehe Abschnitt 3.4.4, als auch eines komplexen Semi-Markoff-Systems [Her94] abgeleitet wurde. Wie in [DS90] gezeigt, kann die Funktion $\rho(x)$ auch für diskrete Markoff-Ketten mit einer unendlichen Zahl von Zuständen $k \rightarrow \infty$ und für kontinuierliche Zufallsvariablen X abgeleitet werden, wie z. B. die Wartezeit in einer Warteschlange.

Üblicherweise wird als Maß für die Korrelation, d. h. die Nicht-Unabhängigkeit, eines (im strengen Sinne) stationären stochastischen Prozesses der folgende (globale) Korrelationskoeffizient h -ter Ordnung für die Zustandsverweilzeiten $A_n := T_{n+1} - T_n$ verwendet, wobei T_n die Folge der Zeitpunkte des stochastischen Prozesses bezeichnet:

$$\kappa_h = \text{Corr}(A_n, A_{n+h}) = \frac{\text{Cov}(A_n, A_{n+h})}{\sqrt{\text{Var}[A_n]\text{Var}[A_{n+h}]}} \quad (3.38)$$

mit

$$\text{Cov}(A_n, A_{n+h}) = E[A_n \cdot A_{n+h}] - E[A_n]^2 \quad (3.39)$$

¹In dieser Arbeit wird abkürzend der Begriff *Korrelation* häufig gleichbedeutend mit *Autokorrelation* verwendet.

nach Gleichung (3.15) und

$$\text{Var}[A_n] = E[(A_n - E[A_n])^2] = E[A_n^2] - E[A_n]^2 = \text{Var}[A_{n+h}] \quad (3.40)$$

Aus der Unabhängigkeit von A_n und A_{n+h} folgt $\kappa_h = 0$, die Umkehrung gilt jedoch i. allg. nicht.

Im Gegensatz zu κ_h ist der lokale Korrelationskoeffizient $\varrho_h(x)$ eine Funktion der möglichen Werte x , die die Zufallsvariable X annehmen kann. Die Definition erfolgt mit Hilfe einer binären Zufallsvariablen, die der Zustandsfolge A_n zugeordnet ist:

$$\zeta_n^{[A]}(x) = \begin{cases} 1 & \text{für } A_n > x \\ 0 & \text{für } A_n \leq x \end{cases} \quad (3.41)$$

unter Verwendung des globalen Korrelationskoeffizienten durch

$$\varrho_h(x) = \text{Corr}[\zeta_n^{[A]}(x), \zeta_{n+h}^{[A]}(x)] \quad (3.42)$$

Nach [Din91a, Din91b] läßt sich Gleichung (3.42) auch folgendermaßen schreiben:

$$\varrho_h(x) = 1 - \frac{\text{P}\{A_n > x, A_{n+h} \leq x\}}{\text{P}\{A_n \leq x\}(1 - \text{P}\{A_n \leq x\})} \quad (3.43)$$

Dies entspricht dem in [Sch94] angegebenen Ausdruck. Im folgenden wird im wesentlichen der lokale Korrelationskoeffizient erster Ordnung ($h = 1$) verwendet:

$$\varrho(x) = \varrho_1(x) = 1 - \frac{\text{P}\{A_n > x, A_{n+1} \leq x\}}{\text{P}\{A_n \leq x\}(1 - \text{P}\{A_n \leq x\})} \quad (3.44)$$

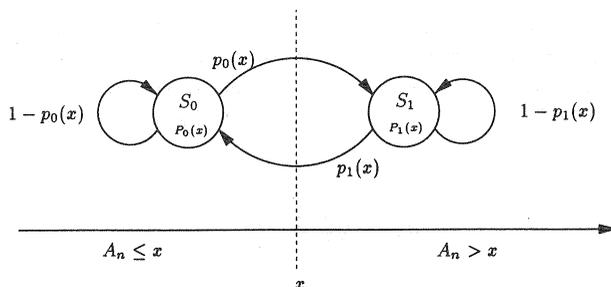


Abbildung 3.7: Zuordnung der 2-gliedrigen Markoff-Kette zum Wert x

Mit diesen Definitionen entspricht der lokale Korrelationskoeffizient $\varrho(x)$ für jeden Wert von X dem Korrelationskoeffizienten einer 2-gliedrigen Markoff-Kette, die den

Wertebereich von X an der Stelle x in die zwei Bereiche $A_n > x$ und $A_n \leq x$ aufteilt, siehe Abbildung 3.7:

$$\varrho(x) = \varrho_1(x) = 1 - [p_0(x) + p_1(x)] \quad (3.45)$$

Für $p_0(x) + p_1(x) < 1$ erhält man eine positive Korrelation. Die Bezeichnungen wurden im Vergleich zu Abschnitt 3.2.2 etwas vereinfacht: $p_{10}(x) = p_0(x)$ und $p_{01}(x) = p_1(x)$.

3.3.2 Korrelierte Prozesse

Zur Modellierung korrelierter Prozesse gibt es eine Reihe von Prozeßtypen, die sich mit Hilfe zusammengesetzter Markoffscher Phasen beschreiben lassen. Ein wichtiger Prozeßtyp ist der sog. Markoffsche Ankunftsprozeß (MAP – *Markovian Arrival Process*), der durch eine zeitkontinuierliche Markoff-Kette gesteuert wird. Ankunftsereignisse können nur bei Zustandsübergängen der Markoff-Kette generiert werden, wobei die Wahrscheinlichkeitsverteilung der Anzahl vom Zustandsübergang abhängig ist.

Auch die Klasse der Markoff-modulierten Poisson-Prozesse (MMPP – *Markov Modulated Poisson Process*) [Mei84, FMH93] wird häufig zur Modellierung korrelierter Ankunftsprozesse benutzt. Hier handelt es sich um einen speziellen doppelt stochastischen Prozeß, dessen momentane Ankunftsrate durch eine zeitkontinuierliche Markoff-Kette gesteuert wird. Ein MMPP kann auf einen MAP abgebildet werden [TG96].

In [Din91a, Her95] wird ein sog. *Spezieller-Semi-Markoff-Prozeß* (SSMP – *Special Semi-Markov Process*) definiert und untersucht, bei dem die Verteilung der Zustandsdauer $F_{ji}(t) = F_i(t)$ nur vom aktuellen Zustand i abhängt.

Bei der Untersuchung getakteter Kommunikationssysteme sind insbesondere diskrete Prozesse von Interesse, so daß es für die diskrete Form dieser Prozesse eine Vielzahl von Untersuchungen gibt.

Im folgenden wird der SSMP(2) mit zwei Zuständen näher beschrieben. Wie in Bild 3.8 angedeutet kann jedem der Zustände eine beliebige Verteilung zugeordnet werden. Für die Beispiele in dieser Arbeit wird die Verteilung in jedem der Zustände als geometrische Verteilung gewählt, so daß für diesen Prozeß auch die genauere Bezeichnung SSMP(Geo,Geo) verwendet wird. Dieser Prozeß wird in Abschnitt 8.7 als Eingabeprozeß benutzt.

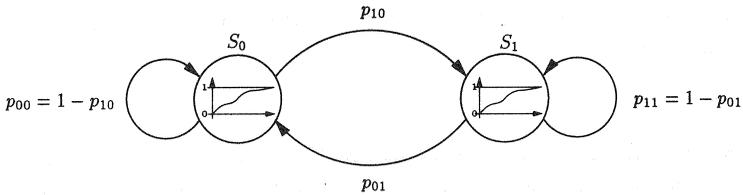


Abbildung 3.8: Strukturbild des SSMP(2)

Die geometrische Verteilung im Zustand $s = 0, 1$ hat die Wahrscheinlichkeitsfunktion, siehe Gleichung (A.18):

$$\left. \begin{aligned} P\{X_s = i\} &= p_s \cdot q_s^{i-1} \quad \text{für } i = 1, 2, \dots \quad \text{und } q_s = 1 - p_s \\ E[X_s] &= \frac{1}{p_s} \end{aligned} \right\} \quad (3.46)$$

Der Korrelationskoeffizient dieser 2-gliedrigen Markoff-Kette wird mit κ bezeichnet im Unterschied zum lokalen Korrelationskoeffizienten ρ des Prozesses:

$$\kappa = 1 - p_{10} - p_{01} \quad (3.47)$$

Die stationären Zustandswahrscheinlichkeiten sind gegeben durch:

$$\left. \begin{aligned} P_0 &= \frac{p_{01}}{p_{10} + p_{01}} \\ P_1 &= 1 - P_0 = \frac{p_{10}}{p_{10} + p_{01}} \end{aligned} \right\} \quad (3.48)$$

Häufig wird dieser Prozesstyp als Eingabeprozeß mit der Eingaberate λ verwendet. Sind λ, κ, q_0 und q_1 gegeben, so können die restlichen Parameter P_0, P_1, p_{10}, p_{01} eindeutig bestimmt werden:

$$\left. \begin{aligned} P_0 &= \frac{\frac{1}{\lambda} - \frac{1}{1 - q_1}}{1 - q_0 - \frac{1}{1 - q_1}} \\ p_{10} &= (1 - \kappa)(1 - P_0) \\ p_{01} &= (1 - \kappa)P_0 \end{aligned} \right\} \quad (3.49)$$

Der Gesamtmittelwert berechnet sich als gewichtete Summe:

$$E[X] = P_0 E[X_0] + P_1 E[X_1] = \frac{1}{\lambda} \quad (3.50)$$

Das Beispiel des SSMP(Geo,Geo) in Abschnitt 8.7 verwendet zwei geometrische Verteilungen mit $\lambda = 4/30, \kappa = 0.6, q_0 = 0.9$ und $q_1 = 0.3$. Aus diesen Parametern

folgt $P_0 = 17/24 = 0.708333$, $p_{10} = 7/60 = 0.11666$, $p_{01} = 17/60 = 0.28333$. Der Mittelwert der geometrischen Verteilung im Zustand S_0 beträgt $E[X_0] = 10$ und im Zustand S_1 beträgt er $E[X_1] = 10/7 = 1.428$. Der Gesamtmittelwert beträgt $E[X] = 7.5$. In diesem Beispiel gibt es einen häufigeren Zustand S_0 , in dem im Mittel größere Abstände erzeugt werden und einen zweiten Zustand S_1 , in dem im Mittel kleinere Abstände erzeugt werden, wie dies z. B. für die Zellabstände einer komprimierten Videoübertragung typisch sein kann.

3.4 Elementare Wartesysteme

Wartesysteme bilden die Basis der Verkehrsmodelle zur Untersuchung von Kommunikationsnetzen, siehe Abbildung 3.9. Im folgenden Abschnitt wird eine Klassifikation von Wartesystemen beschrieben. Anschließend werden wichtige Ergebnisse und Analysemethoden zusammengestellt. Die benutzte Notation lehnt sich an die *Begriffe der Nachrichtenverkehrstheorie* [K⁺97] an. Als Beispiel wird das elementare Referenzmodell $M/M/1/N$ beschrieben, da es ein wichtiges Modell mit begrenzter Wartekapazität ist und in Kapitel 8 zur Validierung von Simulationsergebnissen benötigt wird.

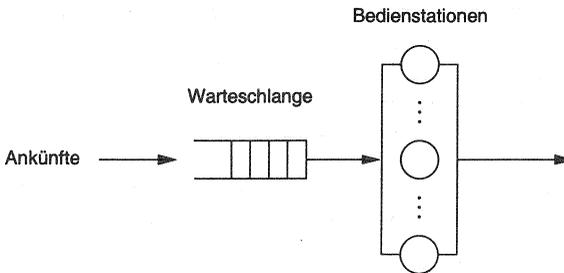


Abbildung 3.9: Struktur eines Wartesystems

3.4.1 Klassifikation von Wartesystemen

Ein *Bediensystem* (*service system*) ist ein Verkehrsmodell, das durch *Ankunftsprozesse* von Anforderungen, *Bedienprozesse*, *Bedienstrategien* und die topologische Struktur gekennzeichnet ist. Im allgemeinen wird der Begriff nur für Systeme mit einer Bedienstufe verwendet. Die Begriffe *Wartesystem*, *Warteraum* und *Warteschlangensystem* (*queueing system*) werden als Synonyme benutzt.

Der *Ankunftsprozeß* (*arrival process*) beschreibt die zeitliche Abfolge von Ankunftseignissen. Der Ankunftsprozeß wird häufig durch eine Wahrscheinlichkeitsverteilung

lung der Ankunftsabstände τ_a beschrieben. Ankunftsereignisse sind z. B. das Eintreffen von Rufen, Zellen, Paketen oder Gruppen von Anforderungen. In bestimmten Zusammenhängen werden auch die Begriffe *Anrufprozeß* oder *Einfallprozeß* verwendet. Anstatt von Anforderungen spricht man auch häufig von Aufträgen.

Der *Bedienprozeß* (service process) beschreibt die Eigenschaften des Bedienvorganges. Der Prozeß wird häufig durch die Wahrscheinlichkeitsverteilung der Bediendauern τ_b dargestellt.

Zur *Bedienstrategie* gehört die *Warteschlangendisziplin* und die *Abfertigungsdisziplin*.

Die *Warteschlangendisziplin* (*queueing discipline*) bestimmt die Reihenfolge, in der Anforderungen zum Zweck der Bedienung aus einer Warteschlange entnommen werden. Beispiele für Warteschlangendisziplinen sind:

- In der Reihenfolge des Eintreffens
(FIFO: first-in, first-out; oder FCFS: first-come, first-served)
- Invers zur Ankunftsreihenfolge
(LIFO: last-in, first-out; oder LCFS: last-come, first-served)
- In zufälliger Reihenfolge (RANDOM, SIRO – service in random order)
- Nach kürzester Bediendauer
(SPT: shortest processing time first; oder SJF: shortest job first)
- Nach längster Bediendauer
(LPT: longest processing time first; oder LJF: longest job first)
- Nach kürzester Restbediendauer
(SRPT: shortest remaining processing time first)

Die *Abfertigungsdisziplin* (*service discipline*) ist die Strategie, nach der den Anforderungen Bedienzeit zugeteilt wird. Beispiele für Abfertigungsdisziplinen sind:

- Einzelabfertigung, Gruppenabfertigung
- Getaktete Abfertigung
- Zeitscheibengesteuerte Abfertigung (RR – Round-Robin)
Als Grenzfall für infinitesimal kleine Zeitscheiben ergibt sich das sog. *Processor Sharing* (PS).
- Erschöpfende Abfertigung
- Torgesteuerte Abfertigung
- Begrenzte Abfertigung

Im theoretisch interessanten Fall von unendlich vielen Bedienstationen spricht man von der IS-Strategie (IS – *Infinite Server*).

Zur formalen Beschreibung von Bediensystemen wird allgemein die sog. *Kendallsche Klassifikation* benutzt. In der erweiterten Form "A/B/S/N" bedeuten:

- A : Typ des Ankunftsprozesses
- B : Typ des Bedienprozesses
- S : Anzahl der Bedieneinheiten
- N : Anzahl der Warteplätze (optional)
 Fehlt diese Angabe, so wird im allgemeinen eine unendliche Wartekapazität angenommen. Ein reines Verlustsystem wird mit A/B/S/0 oder explizit als *Verlustsystem* A/B/S angegeben. An dieser Stelle wird manchmal auch die Gesamtkapazität $N + S$ einschließlich der Bedienstationen angegeben, so daß ein reines Verlustsystem damit als A/B/S/S angegeben wird.

Für A (Abstandsverteilungen der ankommenden Anforderungen) und B (Bedienzeitverteilungen) können z. B. stehen:

- D : Deterministisch
- M : Negativ-exponentiell (Markoff)
- E_k : Erlang k-ter Ordnung
- H_k : Hyperexponentiell k-ter Ordnung
- G : Beliebig (General)
- GI : Beliebig und voneinander unabhängig (General Independent)
- $G^{[X]}$: Beliebig mit Wahrscheinlichkeitsverteilung X der Gruppengröße

Weitere Strukturmerkmale zur genaueren Beschreibung können hinzugefügt werden, z. B. Quellenanzahl oder Warteschlangendisziplin. Ihre Verwendung ist jedoch nicht einheitlich. Beispiele für elementare Wartesysteme sind:

- M/M/S: Poisson-Ankunftsprozeß, negativ-exponentiell verteilte Bediendauer, S Bedieneinheiten;
- M/M/1/N: Poisson-Ankunftsprozeß, negativ-exponentiell verteilte Bediendauer, 1 Bedieneinheit, N Warteplätze;
- $MAP^{[Geo]}/GI/1/N$: Markoffscher Ankunftsprozeß (MAP – *Markovian Arrival Process*) mit geometrisch verteilter Gruppengröße, beliebig verteilter Bediendauer, einer Bedieneinheit und N Warteplätzen. Ein Markoffscher Ankunftsprozeß (MAP) wird durch eine zeitkontinuierliche Markoff-Kette gesteuert. Ankunftsereignisse werden nur bei Zustandsübergängen der Markoff-Kette generiert, wobei die Wahrscheinlichkeitsverteilung der Anzahl vom Zustandsübergang abhängig ist [TG96], siehe Abschnitt 3.3.2.

SSMP(2)/D/1/N: *Special Semi-Markov Process*, korrelierter Ankunftsprozeß, siehe Abschnitt 3.3.2, deterministische Bediendauer, 1 Bedieneinheit, N Wartplätze.

3.4.2 Wichtige Ergebnisse

Das Ziel bei der Bewertung von Wartesystemen ist die Beschreibung der resultierenden Systemprozesse, wie z. B. die Entwicklung der Belegung β , d. h. der Anzahl von Anforderungen im System, oder der Wartezeit τ_w . Zur Beschreibung wird häufig die Wahrscheinlichkeitsverteilung der interessierenden Größen bestimmt. Ferner interessieren die Korrelationseigenschaften dieser Prozesse. Weitere Zielgrößen sind zusammengesetzte Größen wie z. B. die Durchlaufzeit τ_d als Summe aus Wartezeit τ_w und Bediendauer τ_b .

In Tabelle 3.2 und 3.3 sind beispielhaft wichtige Ergebnisse zusammengestellt. Die in dieser Tabelle verwendeten Bezeichnungen sind zur Übersicht in Tabelle 3.1 und in Anhang C zusammengestellt.

Eine grundlegende Beziehung zur Umrechnung von Mittelwerten bzgl. der *Anzahl* und der *Zeit in einem System* unter Verwendung der Eingaberate λ stellt die *Formel von Little* zur Verfügung:

$$E[\text{Anzahl im System}] = \lambda E[\text{Zeit im System}] \quad (3.51)$$

wobei das System hier sehr allgemein verstanden werden kann, d. h. man kann die Formel von Little auf das Gesamtsystem mit Warteschlangen und Bedieneinheiten beziehen oder z. B. nur auf eine Warteschlange. Es genügt also einen der beiden Mittelwerte zu bestimmen, um mit Hilfe der Eingaberate λ den jeweils zugeordneten zu bestimmen. Eine Erweiterung dieser Formel, die mit Hilfe von faktoriellen Momenten definiert wird, gilt auch für höhere Momente [Bru72].

Ein weiterer Prozeß, der *Abgangsprozeß* oder *Ausgangsprozeß* beschreibt die zeitliche Abfolge von Abgangsereignissen bzw. Ausgangsereignissen. Meist wird der Abgangsprozeß durch die statistischen Eigenschaften (Wahrscheinlichkeitsverteilung, Korrelation) der Abstände der Abgangsereignisse beschrieben.

Wie in Tabelle 3.2 und 3.3 beispielhaft zu sehen ist, sind in Abhängigkeit vom Warteraumtyp eine Vielzahl von analytischen Ergebnissen bekannt.

Tabelle 3.1: Elementare Wartesysteme: Bezeichnungen

Bezeichnung	Bedeutung
N	Kapazität des Wartespeichers
τ_a	Zeitzufallsvariable: Ankunftsabstände bzw. Zwischenankunftszeiten
τ_b	Zeitzufallsvariable: Bediendauer
$\mu = 1/E[\tau_b]$	Bedienrate
S	Anzahl Bedienstationen
$A = \frac{E[\tau_b]}{E[\tau_a]}$	Verkehrsangebot, entspricht der Verkehrsbelastung des Systems, wenn keine Aufträge abgewiesen werden
$\eta = \frac{E[\tau_b]}{S \cdot E[\tau_a]}$	relatives Verkehrsangebot bezogen auf die Anzahl der Bedienstationen S
β	Zufallsvariable: Belegung, d. h. Anzahl Aufträge im System entsprechende Definitionen für die Ankunftsbelegung α
B	maximale Belegung $B = N + S$
$P_\beta(j) = P\{\beta = j\}$	Belegungswahrscheinlichkeit: Wahrscheinlichkeit für $\beta = j$ Aufträge im System
$E[\beta], \text{Var}[\beta]$	Mittelwert und Varianz der Belegung
P_0	Wahrscheinlichkeit für das leere System
ρ_β	lokaler Korrelationskoeffizient der Belegung
P_B	Verlustwahrscheinlichkeit
L	Zufallsvariable: Anzahl Aufträge in der Warteschlange
$P_L(j) = P\{L = j\}$	Wahrscheinlichkeit für $L = j$ Aufträge in der Warteschlange
$E[L], \text{Var}[L]$	Mittelwert und Varianz der Auftragsanzahl in der Warteschlange
τ_w	Zufallsvariable: Wartezeit
$F_w(\tau)$	Verteilungsfunktion der Wartezeit
$E[\tau_w], \text{Var}[\tau_w]$	Mittelwert und Varianz der Wartezeit
τ_d	Zufallsvariable: Durchlaufzeit
$F_d(\tau)$	Verteilungsfunktion der Durchlaufzeit
$E[\tau_d], \text{Var}[\tau_d]$	Mittelwert und Varianz der Durchlaufzeit
τ_g	Zufallsvariable: <i>busy period</i>
$F_g(\tau)$	Verteilungsfunktion der <i>busy period</i>
τ_I	Zufallsvariable: Leerzeit (<i>idle time</i>)
$F_I(\tau), L_I(s)$	Verteilungsfunktion und Laplace-Transformierte der Leerzeit

Eine vollständige Liste aller Bezeichnungen befindet sich in Anhang C.

Tabelle 3.2: Elementare Wartesysteme: Übersicht

Kennzeichen	Warteraum	Methode & Wichtige analytische Ergebnisse
Exponential-Warteraum	M/M/1-FIFO	Stabilitätsbedingung: $\eta < 1$ $P_\beta, F_w(\tau), F_d(\tau)$ G&S-Prozeß
Exponential-Warteraum	M/M/S-FIFO	Stabilitätsbedingung: $\eta < 1$ $P_\beta, F_w(\tau), F_d(\tau)$ G&S-Prozeß
endliche Wartekapazität	M/M/1/N-FIFO	System immer stabil $\eta < \infty$ P_β , lokaler Korrelationskoeffizient $\rho_\beta, F_w(\tau)$ G&S-Prozeß
endliche Wartekapazität	M/M/S/N-FIFO	System immer stabil $\eta < \infty$ $P_\beta, F_w(\tau)$ für $N = 0$: Erlang-B-Formel $P_B = P_\beta(S)$, siehe Gl. (6.13) G&S-Prozeß
Pollaczek-Khintchine Warteraum	M/G/1-FIFO	Stabilitätsbedingung: $\eta < 1$ eingebettete Markoff-Kette $P_0, E[\beta]$ (Pollaczek-Khintchine Formel), $\text{Var}[\beta], E[L], \text{Var}[L], E[\tau_w]$ (gilt auch für andere Strategien LIFO, SIRO), $\text{Var}[\tau_w], E[\tau_d]$ (gilt auch für andere Strategien LIFO, SIRO), $\text{Var}[\tau_d], F_I(\tau), L_w(s) = f(s, \eta, \lambda, L_b(s))$
Konstante Bedienzeit	M/D/1-FIFO	Stabilitätsbedingung: $\eta < 1$ $P_\beta, E[L], \text{Var}[L], F_d(\tau)$ Spezialfall M/G/1
Erlang-k Bedienzeit	M/E _k /1-FIFO	Stabilitätsbedingung: $\eta < 1$ $P_\beta, F_w(\tau)$ Spezialfall M/G/1 Phasenmethode
Hyperexponentielle Bedienzeit	M/H _k /1-FIFO	Stabilitätsbedingung: $\eta < 1$ $P_\beta, F_w(\tau)$ Spezialfall M/G/1 Phasenmethode
Processor Sharing	M/G/1-PS	Stabilitätsbedingung: $\eta < 1$ Grenzwertbetrachtung $P_\beta, E[\tau_d]$ (wie bei M/M/1, gilt nur für den Mittelwert)
unbegrenzte Anzahl Bedienstationen	M/G/ ∞	System immer stabil: $\eta < \infty$ P_β , keine Warteschlange: $L = 0, \tau_w = 0,$ $\tau_d = \tau_b, F_d(\tau) = F_b(\tau)$

Tabelle 3.3: Elementare Wartesysteme: Übersicht (Fortsetzung)

Kennzeichen	Warteraum	Methode & Wichtige analytische Ergebnisse
unbegrenzte Anzahl Bedienstationen	G/M/1	Stabilitätsbedingung: $\eta < 1$ Ergebnisse abhängig von der Lösung der Gleichung $\omega = L_a(\mu - \mu \cdot \omega)$ ($\omega = \eta$ für M/M/1), $P_0 = 1 - \eta$, P_β , $F_w(\tau, \omega)$ und $F_d(\tau, \omega)$
Allgemeiner Warteraum	G/G/1-FIFO	Stabilitätsbedingung: $\eta < 1$ Restarbeit $L_w(s) = f(L_a(s), L_b(s), L_I(s))$
Allgemeiner Warteraum	G/G/S-FIFO	Stabilitätsbedingung: $\eta < 1$ mittlere Anzahl belegter Bedienstationen: $S \cdot \eta$
Bevorzugung der neuesten Anforderung	M/M/1-LIFO	Stabilitätsbedingung: $\eta < 1$ G&S-Prozess $F_w(\tau)$, siehe M/G/1
nicht-unterbrechende Prioritäten	M/G/1-NON-PRE-PRIO	Stabilitätsbedingung: $\eta < 1$ Fertigstellungszeit $E[\tau_w]$, $E[\tau_w]$
unterbrechende Prioritäten	M/G/1-PRE-PRIO	Stabilitätsbedingung: $\eta < 1$ Fertigstellungszeit $E[\tau_w]$, $E[\tau_w]$ höchste Priorität M/G/1
Bevorzugung kurzer Ausgabezeiten	M/G/1-SPT	Stabilitätsbedingung: $\eta < 1$ Ansatz: NON-PRE-PRIO $E[\tau_w]$, $\text{Var}[\tau_w]$, optimale nicht-unterbrechende Strategie bzgl. der mittleren Wartezeit
Bevorzugung kurzer Ausgabezeiten mit Unterbrechungen	M/G/1-SRPT	Stabilitätsbedingung: $\eta < 1$ Ansatz: PRE-PRIO $E[\tau_w]$, $\text{Var}[\tau_w]$, Bestimmung der Momente über die Laplace-Transformierten der Anfangswartezeit und Fertigstellungszeit, optimale unterbrechende Strategie bzgl. der mittleren Wartezeit
Zeitscheibenverfahren	M/G/1-RR	Stabilitätsbedingung: $\eta < 1$ $E[\tau_d \tau_a]$, $E[\tau_d]$
korrelierte Eingaben	SMP ^[X] /D/1/N-FIFO	Stabilitätsbedingung: $\eta < 1$ Restarbeit P_α , $F_w(\tau \text{kein Verlust})$
korrelierte Eingaben	DBMAP/G/1/N-FIFO	Stabilitätsbedingung: $\eta < 1$ M/G/1-Typ P_β , P_α , $F_w(\tau)$
korrelierte Eingaben	SSMP(2)/D/1/N-FIFO	Stabilitätsbedingung: $\eta < 1$ Restarbeit P_α , lokaler Korrelationskoeffizient ρ_α

3.4.3 Analysemethoden

In diesem Abschnitt werden einige der wichtigsten Analysemethoden vorgestellt. Grundsätzlich ist es vorteilhaft, bei der Analyse von Wartesystemen im transformierten Bereich der Zufallsvariablen zu arbeiten. Die dabei erzielten Ergebnisse sind jedoch nicht immer rücktransformierbar, so daß man häufig nicht die geschlossene Verteilungsfunktion erhält, aber zumindest die prinzipielle Möglichkeit Momente beliebiger Ordnung zu bestimmen.

Ein weiterer Grundsatz bei der Analyse von Wartesystemen ist der, daß man soweit möglich versucht, die Methoden zur Behandlung von Markoff-Prozessen anzuwenden, d. h. man versucht Markoff-Prozesse in einem System zu identifizieren, in dem auch Nicht-Markoff-Prozesse auftreten.

Man spricht von einer *eingebetteten Markoff-Kette* (*imbedded Markov chain*) [Kle75], die in einem Verkehrsmodell zu bestimmten (eingebetteten) Zeitpunkten, an denen seine zukünftige Entwicklung nur vom gegenwärtigen Zustand abhängt, betrachtet wird. Die Folge aller Zeitpunkte bildet einen Semi-Markoff-Prozeß, während die Folge der eingebetteten Zeitpunkte einen Markoff-Prozeß bildet. Das Verhalten des Verkehrsmodells an den eingebetteten Zeitpunkten kann durch eine Zustands-Übergangsmatrix eindeutig beschrieben werden.

Bei der *Phasenmethode* (*phase method*) [TG96] werden beliebig (G) verteilte Phasendauern durch ein Netzmodell aus gedächtnislosen Teilphasen ersetzt. Die momentan erreichte Teilphase stellt dabei das Gedächtnis dar. Grundlage der Phasenmethode ist die Tatsache, daß sich jede Verteilungsfunktion, die eine rationale Laplace-Transformierte hat, beliebig genau durch ein Netzmodell aus negativ-exponentiell verteilten Teilphasen darstellen läßt.

Die *matrixanalytische Methode* (*matrix analytic method, matrix geometric method*) [Neu81, TG96] ist ein numerisches Analyseverfahren für Verkehrsmodelle, auf die die Methode der eingebetteten Markoff-Kette angewandt werden kann, und deren Zustands-Übergangsmatrix eine bestimmte Blockstruktur aufweist. Mit dieser Methode kann z. B. das MAP/GI/1-System analysiert werden.

Eine Analyse mit Hilfe der sog. *Restarbeit* $U(t)$ (*unfinished work*) [Kle75, Her94] nutzt die Tatsache aus, daß $U(t)$ ein Markoff-Prozeß mit kontinuierlichen Zuständen und Sprungstellen zu Ankunftszeitpunkten ist. Dieser Ansatz wird speziell zur Analyse endlicher Wartesysteme eingesetzt.

Allgemein kann man mit der Methode der *Zusatzvariablen* (*supplementary variable*) [Kle75] durch Hinzufügen von Variablen versuchen einen stochastischen Prozeß in einen Markoff-Prozeß zu überführen. Der Zustandsprozeß $L(t)$, der die Warteschlangenlänge eines M/G/1-Systems beschreibt, wird durch Hinzufügen der Variablen z , die die abgelaufene Bediendauer der Anforderung in der Bedienstation beschreibt, zu einem Markoff-Prozeß $\{L(t), z\}$.

Bei Prioritätensystemen wird zur Reduktion der Komplexität die *Fertigstellungszeit* (*completion time*) [Jai68] betrachtet, die die Zeit vom Beginn der Bearbeitung einer Anforderung bis zum Abschluß inklusive aller zusätzlicher Wartezeiten aufgrund höherer Prioritäten beschreibt. Durch den Beginn einer Fertigstellungszeit wird ein Erneuerungspunkt definiert.

Die *Mittelwertanalyse* (*mean value analysis*) [RL80] ist eine Methode, die sich nur auf die Bestimmung von Mittelwerten konzentriert. Dadurch ist das Verfahren auch auf allgemeinere Verkehrsmodelle wie Prioritätenmodelle und Warteschlangennetze mit Produktlösungsform anwendbar, siehe Abschnitt 3.5.

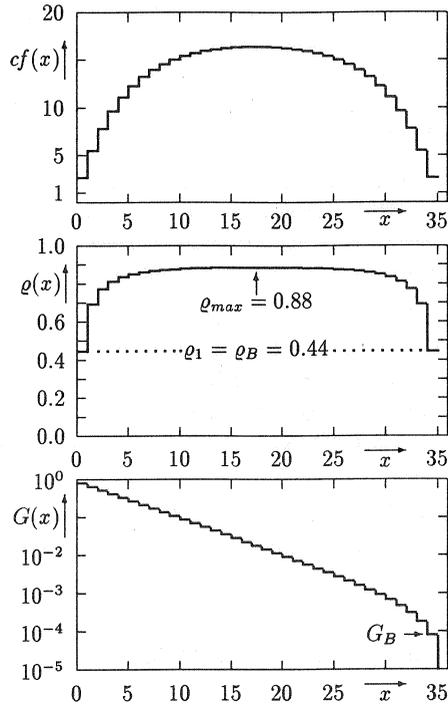
Ein *Flußmodell* (*fluid flow model*) [Kle76, Krö95] ist ein Ersatzmodell, in dem der Verkehr durch einen kontinuierlichen Flüssigkeitsstrom mit zufallsabhängiger Intensität nachgebildet wird. Die Änderung der Intensität entspricht einer Änderung der Rate des Ankunftsprozesses. Als Synonym wird auch der Begriff Flüssigkeitsflußmodell verwendet. Im Englischen wird es auch als *Uniform Arrival and Service Model* (*UAS Model*) bezeichnet.

Da im folgenden insbesondere der Warteraum M/M/1/N als Referenzsystem benutzt wird, werden im anschließenden Abschnitt einige wichtige zugehörige Ergebnisse vorgestellt.

3.4.4 Das elementare Referenzmodell M/M/1/N

Das System M/M/1/N-FIFO mit begrenztem Wartespeicher N und maximaler Belegung $B = N + 1$ ist ein grundlegendes Referenzmodell. Mit Hilfe von Geburt- und Sterbeprozessen können die wesentlichen Ergebnisse zu diesem Warteraum abgeleitet werden, siehe [Kle75]. Die lokalen Korrelationseigenschaften der Belegung wurden in [Sch94] nachgewiesen.

In Kapitel 8 wird dieses Modell für die Validierung einer neuen Methode für die Simulation seltener Ereignisse benutzt, da es damit möglich ist, für dieses Modell die Simulationsergebnisse mit exakt berechneten Werten zu vergleichen.



Lokaler Korrelationsfaktor $cf(x)$ Maximale Belegung: $B = N + 1 = 35$
 Lokaler Korrelationskoeffizient $\rho(x)$ Verkehrsbelastung: $\eta = 0.8$
 Komplementäre VF $G(x)$ Verluste: $P_B = G_B = 8.1 \times 10^{-5}$

Abbildung 3.10: Wartesystem M/M/1/N

Mit dem Verkehrsangebot $\eta = E[\tau_b]/E[\tau_d]$ kann die diskrete Zufallsvariable *Belegung* β durch folgende Gleichungen beschrieben werden:

Wahrscheinlichkeit für das leere System P_0 :

$$P_0 = P\{\beta = 0\} = \begin{cases} \frac{1-\eta}{1-\eta^{B+1}} & \text{für } \eta \neq 1 \\ \frac{1}{B+1} & \text{für } \eta = 1 \end{cases} \quad (3.52)$$

Stationäre Wahrscheinlichkeitsfunktion $P_\beta(j)$:

$$\left. \begin{aligned} P_\beta(j) &= P_0 \eta^j \quad \text{für } j = 0, 1, \dots, B \\ \text{Verlustwahrscheinlichkeit: } P_B &= P\{\beta = B\} = P_0 \eta^B \end{aligned} \right\} \quad (3.53)$$

Stationäre komplementäre VF $G(x) = 1 - F(x)$:

$$\left. \begin{aligned} G(x) = G_i &= P\{\beta \geq i\} = \sum_{j=i}^B P_\beta(j) = \frac{\eta^i - \eta^{B+1}}{1 - \eta^{B+1}} \quad \text{für } i-1 \leq x < i \\ i &= 1, 2, \dots, B \quad \text{mit } G_0 = 1 \quad \text{und } G_B = P_B \end{aligned} \right\} \quad (3.54)$$

Mittlere Anzahl Aufträge im System $E[\beta]$:

$$E[\beta] = \frac{1}{1-\eta} - \frac{(B+1)\eta^{B+1}}{1-\eta^{B+1}} \quad (3.55)$$

Effektive Ankunftsrate:

$$\lambda(1 - P_B) \quad (3.56)$$

Mittlere Durchlaufzeit $E[\tau_d]$:

$$E[\tau_d] = \frac{E[\beta]}{\lambda(1 - P_B)} \quad (3.57)$$

Lokaler Korrelationskoeffizient $\rho(x)$:

$$\rho(x) = \rho_i = 1 - \frac{(1-\eta)(1-\eta^{B+1})\eta^i}{(1+\eta)(1-\eta^i)(\eta^i - \eta^{B+1})} \quad (3.58)$$

für $i-1 \leq x < i$ mit $i = 1, 2, \dots, B$

Lokaler Korrelationsfaktor $cf(x)$:

$$cf(x) = \frac{1 + \rho(x)}{1 - \rho(x)} \quad (3.59)$$

Die komplementäre VF $G(x)$ zeigt in Abbildung 3.10 einen approximativ geometrischen Verlauf vom Niveau $G_0 = 1$ bis zum Niveau $G_B = P_B = 8.1 \times 10^{-5}$, wohingegen der lokale Korrelationskoeffizient $\rho(x)$ sich als Maximumkurve mit gleichem *Startniveau* ρ_1 und *Endniveau* ρ_B zeigt:

$$\rho_1 = \rho_B = \frac{\eta(1 - \eta^{B-1})}{(1 + \eta)(1 - \eta^B)} \approx \frac{\eta}{1 + \eta} \quad (3.60)$$

Bei ungeradem B ergibt sich im Interval $(B - 1)/2 \leq x < (B + 1)/2$ das *Mazimalniveau* ρ_{max} zu

$$\rho_{max} = \frac{2\eta(1 - \eta^{(B-1)/2})}{(1 + \eta)(1 - \eta^{(B+1)/2})} \approx \frac{2\eta}{1 + \eta} \quad (3.61)$$

Auch der lokale Korrelationsfaktor $cf(x)$ folgt einer ähnlichen Maximumkurve wie $\rho(x)$, siehe Abbildung 3.10.

Wählt man die Parameter wie in Abbildung 3.10, so ergibt sich $\rho_1 = \rho_B = 0.44$ und $\rho_{max} = 0.88$. Die Approximationen in Gleichung (3.60) und (3.61) sind anwendbar für $\eta^{B-1} \ll 1$ bzw. $\eta^{(B-1)/2} \ll 1$. Sie sind identisch mit den Termen $\rho(0)$ bzw. $\rho(\infty)$ in Gleichung (19) in [DS90] bezüglich der Belegung β des M/M/1-Systems mit unendlichem Wartespeicher.

Analytische Untersuchungen [Her95] und Simulationen von anderen, komplexeren Bediensystemen mit einer FIFO-Bediensstation und endlichem Wartespeicher haben gezeigt, daß der Charakter der flachen Maximumkurve $\rho(x)$ in Abbildung 3.10 des M/M/1/N-FIFO-Systems repräsentativ für eine größere Klasse von Systemen ist.

3.5 Warteschlangennetze

Warteschlangennetze oder Wartenetze bilden die Grundlage zur Leistungsbewertung vernetzter Systeme, insbesondere Rechensysteme und Kommunikationssysteme. Sie werden z. B. bei der Modellierung von Telekommunikationsnetzen oder Lokalen Netzen eingesetzt.

Zu einer Teilmenge solcher Warteschlangennetze sind analytische Lösungen bekannt [Bol89]. Ändert man in diesen Modellen jedoch eine Modellkomponente, z. B. die Warteschlangendisziplin in einem der Knoten, so ist zu diesem neuen Modell meist keine analytische Lösung bekannt. Eine Änderung dieser Art in einem Simulationsmodell ist dagegen einfach. Im folgenden soll die im Bereich der Warteschlangennetze übliche Terminologie erklärt werden.

Ein Warteschlangennetz besteht aus N einzelnen Stationen, die auch als Knoten bezeichnet werden. Ein Knoten ist ein einfacher Warteraum mit einer oder mehreren

Bedieneinheiten. In jedem Knoten werden Aufträge bearbeitet, in einem Kommunikationsnetz sind dies z. B. Nachrichten, Pakete oder Zellen. Grundsätzlich sind Übergänge zwischen allen Knoten möglich, d. h. daß insbesondere auch Rückkopplungen erlaubt sind. Der Weg eines Auftrags durch das Netz wird durch Übergangswahrscheinlichkeiten beschrieben. Innerhalb eines Netzes können verschiedene Auftragsklassen unterschieden werden. Die Bediendauern oder auch die Übergangswahrscheinlichkeiten können unterschiedlich sein für verschiedene Klassen. Auch der Übergang von einer Auftragsklasse in eine andere ist in manchen Modellen zugelassen.

Die *Betriebsanalyse* (*Operationelle Analyse, operational analysis*) [DB78] ist ein Verfahren zur Analyse von Warteschlangennetzen bei dem bestimmte Randbedingungen für das Modell gelten müssen:

- Alle betrachteten Größen müssen meßbar sein. Die Gültigkeit der Ergebnisse darf nur auf Annahmen beruhen, die durch die Beobachtung des Systems für eine endliche Zeitdauer überprüft werden können.
- Jedes Teilsystem im Warteschlangennetz muß im Flußgleichgewicht sein, d. h. die Anzahl der Ankünfte von Aufträgen muß während eines Beobachtungszeitraums für jedes Teilsystem (fast) gleich der Anzahl der Abgänge bedienter Aufträge sein.
- Jedes Teilsystem muß homogen sein (keine zustandsabhängigen Prozesse).

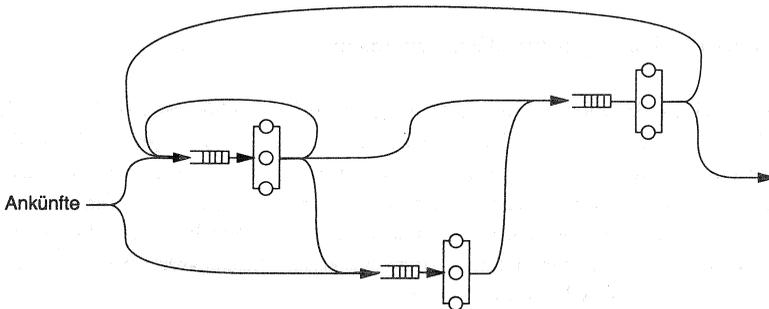
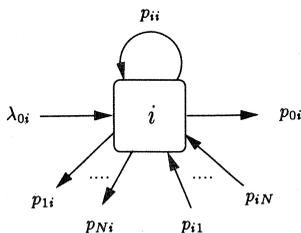


Abbildung 3.11: Beispiel eines offenen Warteschlangennetzes

Ein Warteschlangennetz wird als *offen* bezeichnet, wenn alle Aufträge *von außen* (Quelle) an einem der Netzknoten eintreffen und das Netz nach ggf. mehrfacher Bearbeitung *nach außen* (Senke) wieder verlassen. Ein Beispiel eines offenen Warteschlangennetzes zeigt Abbildung 3.11. Eine allgemeine Darstellung eines Knotens eines offenen Netzes ist in Abbildung 3.12 zu sehen. Ein Warteschlangennetz wird als *geschlossen* bezeichnet, wenn keine externen Auftragsankünfte und -abgänge möglich



- N : Gesamtanzahl Knoten
 λ_{0i} : Eingaberate von außen am Knoten i
 p_{0i} : Wahrscheinlichkeit, daß ein Auftrag das Warteschlangennetz nach seiner Bearbeitung im Knoten i verläßt
 p_{ij} : Verzweigungswahrscheinlichkeit zum i -ten Knoten nach Verlassen des j -ten Knotens.

Abbildung 3.12: Knoten i eines allgemeinen offenen Warteschlangennetzes

sind, d. h. die Auftragszahl ist konstant und wird mit K bezeichnet. Innerhalb eines Warteschlangennetzes können gleichzeitig offene und geschlossene Auftragsklassen definiert werden. Solche Netze bezeichnet man als *gemischte* Warteschlangennetze. Ein Rechnersystem mit interaktiven Terminals und gleichzeitigem Stapelbetrieb von Aufträgen wird beispielsweise als gemischtes Warteschlangennetz modelliert.

Analytische Lösungen für Warteschlangennetze

Für die Klasse der *offenen* Netze gibt es meist einfachere Lösungsverfahren als für die Klasse der *geschlossenen* Netze. Viele Anwendungen erfordern jedoch die Modellierung als geschlossene Systeme, z. B. ein Zentralrechner mit einer *geringen* Anzahl angeschlossener Terminals, da man erst bei einer großen Anzahl von Terminals approximativ wieder mit offenen Modellen arbeiten kann. Detaillierte Beschreibungen von Warteschlangennetzen und zugehörige analytische Lösungsverfahren findet man in [BA82, Bol89, Sch88b, Wal88, Wal92].

Die meisten analytischen Verfahren beruhen auf der Bestimmung der *stationären Zustandswahrscheinlichkeiten* $P(k_1, k_2, \dots, k_N)$ für den Belegungsvektor des Netzes (k_1, k_2, \dots, k_N) , wobei k_i die Anzahl der Aufträge im Knoten i bezeichnet. Aus diesen Zustandswahrscheinlichkeiten lassen sich die wichtigsten Leistungsgrößen berechnen, wie z. B. die mittlere Warteschlangenlänge oder die mittlere Wartezeit.

Geschlossene Netze mit einer nicht zu großen Anzahl von Aufträgen und Knoten sowie Bediendauerverteilungen, die aus exponentiellen Phasen bestehen, lassen sich mit Hilfe von Markoff-Ketten lösen, siehe Abschnitt 3.4.3 und [Sch88b, Wal92].

Viele Lösungsverfahren beruhen auf der sog. *Produktform* der Zustandswahrscheinlichkeiten, die für Teilklassen der Warteschlangennetze gelten, wobei die Faktoren $g_i(k_i)$ nur vom jeweiligen Knotenindex abhängen und $G(K)$ die Normalisierungskonstante darstellt:

$$P(k_1, k_2, \dots, k_N) = \frac{1}{G(K)} g_1(k_1) g_2(k_2) \dots g_N(k_N) \quad (3.62)$$

Die einfachste Form von offenen Netzen sind die sog. *Exponentialnetze* oder auch *Jackson-Netze* [Jac63], die sich durch folgende Eigenschaften auszeichnen:

- offenes Netz mit unbeschränkter Auftragszahl
- eine Auftragsklasse
- unabhängige negativ-exponentiell verteilte Ankünfte
- FIFO-Strategie mit unbegrenzter Wartekapazität in jedem Knoten
- unabhängige negativ-exponentiell verteilte Bediendauern
- s_i identische Bedieneinheiten in jedem Knoten
(ggf. belegungsabhängige Bedienrate)

Das Pendant für geschlossene Netze sind die sog. *Gordon-Newell-Netze* [GN67]. Es müssen alle Voraussetzungen wie bei den Jackson-Netzen erfüllt sein, aber es gibt nur eine konstante Anzahl von Aufträgen im Netz.

Für die sog. *BCMP-Netze* wurde 1975 von Baskett, Chandy, Muntz und Palacios [BCMP75] die Produktform bewiesen. BCMP-Netze sind eine erweiterte Klasse von Warteschlangennetzen, die sowohl die Jackson-Netze als auch die Gordon-Newell-Netze umfassen. Sie lassen mehrere Auftragsklassen (geschlossen und offen), verschiedene Wartestrategien und allgemein verteilte Bedienzeiten zu:

- *Warteschlangendisziplinen bzw. Abfertigungsdisziplinen*: FIFO (*first-in first-out*), PS (*processor sharing*), IS (*infinite servers*), LIFO-PR (*last-in first-out, preemptive resume*)
- *Auftragsklassen*: Aufträge gehören während sie auf Bedienung warten und während sie bedient werden genau einer Auftragsklasse an. Nach der Bedienung dürfen sie aufgrund fester Wahrscheinlichkeiten die Auftragsklasse wechseln.
- *Bediendauerverteilungen*: Bei FIFO-Knoten muß die Bediendauerverteilung für alle Auftragsklassen gleich und negativ-exponentiell sein. An den anderen Knoten müssen die Bediendauerverteilungen eine rationale Laplace-Transformierte haben, verschiedene Auftragsklassen können verschiedene Verteilungen haben.
- *zustandsabhängige Bedienung*: An einem FIFO-Knoten darf die Bediendauer nur von der Gesamtlänge der Warteschlange abhängen. Die Bediendauer einer Auftragsklasse an anderen Knoten (PS, LIFO-PR, IS) darf auch von der Länge der Warteschlange für diese Auftragsklasse abhängen, nicht aber von den Warteschlangenlängen anderer Klassen. Außerdem darf die Gesamtbedienrate eines

- Teilnetzes von der Gesamtanzahl der Aufträge in diesem Teilnetz abhängen.
- *Ankunftsprozesse*: In offenen Netzen ist die Verteilung der Zwischenankunftszeiten negativ-exponentiell. Es sind keine Gruppenankünfte erlaubt. Die Ankunftsraten dürfen zustandsabhängig sein. Es dürfen gleichzeitig offene und geschlossene Auftragsklassen in einem Netz sein.

Viele Anschlußarbeiten befassen sich mit *effizienten* Algorithmen zur Bestimmung der Normalisierungskonstanten $G(K)$ oder zur direkten Bestimmung von Mittelwerten [RL80] der gesuchten Leistungsgrößen. Neben diesen exakten Verfahren sind *Approximationsverfahren* für eine Vielzahl von Modellen definiert worden. Zur Validierung dieser Approximationsverfahren sind *simulative* Nachbildungen der Modelle notwendig. Die Approximationsverfahren sind meist weniger rechenzeit-intensiv und können so zur Untersuchung von Parameterspektren herangezogen werden. Ohne die simulative Validierung wären sie jedoch nicht anwendbar. Einige der Verfahren zur Berechnung von Warteschlangennetzen werden in Werkzeugen angeboten, z. B. [KGZH95].

Dekompositionsverfahren zerlegen netzartige Modelle in einfachere Teilmodelle, die getrennt untersucht werden. Bei der *Aggregation* werden Teilnetze durch einfachere Ersatzsysteme zusammengefaßt, die hinsichtlich des Durchsatzes und der Durchlaufzeiten dem Teilnetz äquivalent sind. Häufig wird das Ersatzsystem zu einem Knoten zusammengefaßt. Unter bestimmten Randbedingungen, wie z. B. bei Produktlösungs-Warteschlangennetzen, kann die Aggregation exakt durchgeführt werden. Die Verfahren werden unter bestimmten Randbedingungen auch approximativ erfolgreich eingesetzt, siehe [Cou77, Küh79].

Zufallszahlengeneratoren

Zufallszahlen sind eine wichtige Grundlage der stochastischen, diskreten Simulation, bei der sehr komplexe bzw. weitgehend unbekannte Prozesse als zufällig beschrieben werden, z. B. die Zwischenankunftszeit von Telefongesprächen oder Nachrichten oder deren Bearbeitungszeit. Zufallszahlen werden auch in anderen Rechneranwendungen benötigt, z. B. Monte-Carlo-Integration oder probabilistische Algorithmen, wie genetische Algorithmen oder Optimierungsverfahren auf Basis des *simulated annealing*.

Meist werden Zufallszahlen in Rechneranwendungen durch einen fest vorgegebenen Algorithmus erzeugt und sind somit *pseudozufällig*. Es gibt auch einige wenige Generatoren, die physikalische Phänomene ausnutzen, wie z. B. radioaktiven Zerfall [Gud85, Gud87] oder elektronisches Rauschen [Ric92].

Ziel dieser Generatoren ist es, einen Basisgenerator für die Gleichverteilung im Intervall $(0, 1)$ ¹ bereitzustellen, der unabhängige Werte liefert. Aufbauend auf diesem Basisgenerator können durch Transformation Zufallszahlen andersartiger Verteilungen erzeugt werden, z. B. die häufig benötigte negativ-exponentielle Verteilung, die durch Inversion gewonnen werden kann. Dieses einfache Generierungsprinzip ist bei allen Verteilungen möglich, bei denen die Inverse der Verteilungsfunktion bekannt ist. Andernfalls kann im allgemeinen eine numerische Inversions-Methode benutzt werden, wenn die Verteilungsfunktion als Formel oder Tabelle vorliegt. Für einige spezielle Verteilungen gibt es aufgrund spezieller Eigenschaften andere Generierungsmethoden.

Basisgeneratoren sollten aufgrund ihrer theoretischen Eigenschaften ausgewählt werden und zusätzlich statistischen Tests unterworfen werden [L'E96]. Ein gutes Abschneiden bei den statistischen Tests ist aber lediglich ein notwendiges und kein hinreichendes Kriterium für einen guten Generator. Weitere Auswahlkriterien für einen Zufallszahlengenerator sind: Geschwindigkeit, Speicherbedarf, Wiederholbarkeit, Portabilität, Komplexität der Implementierung und einfache Aufteilungsmöglichkeiten der Sequenz.

Im folgenden werden die Grundtypen der häufigsten Basisgeneratoren mit ihren Eigenschaften vorgestellt. Nach einer Betrachtung zur Generierung allgemeiner Ver-

¹Dieses Intervall wird auch als $[0, 1]$ oder $[0, 1)$ angegeben. Dies kann bei der Benutzung der inversen Methode ggf. zu Problemen führen, da 0 auf den kleinsten und 1 auf den größtmöglichen Wert abgebildet werden, die in vielen Fällen nicht endlich sind.

teilungen schließt das Kapitel mit einigen Empfehlungen, die bei der Auswahl und Benutzung von Zufallszahlengeneratoren beachtet werden sollten.

4.1 Basisgeneratoren

Pseudozufallsgeneratoren erzeugen meist Zahlen als Funktion einer oder mehrerer Vorgängerzahlen x_1, \dots, x_{n-1} :

$$x_n := f(x_{n-1}, x_{n-2}, \dots, x_1) \quad (4.1)$$

Die so gewonnene Zahl x_n wird auf das Intervall $(0,1)$ abgebildet. Dies erfolgt durch Division mit einem Maximalwert oder durch die Abbildung des Bitmusters auf eine Gleitkommazahl. Je nach Genauigkeit der dargestellten Gleitkommazahlen wird in der Praxis auch mehr als ein Wert der Folge x_n zur Erzeugung einer $(0,1)$ -Gleitkommazahl benutzt.

Aufgrund der funktionalen Abhängigkeit und der Anzahl der Vorgängerwerte, die benutzt werden, lassen sich folgende gebräuchliche Generatortypen unterscheiden:

Lineare Generatoren

- Linearer Kongruenz-Generator
(LCG – *Linear Congruential Generator*):

$$x_n := (ax_{n-1} + b) \bmod m \quad (4.2)$$

- Multiplikativer Kongruenz-Generator (LCG mit $b = 0$)
(MCG – *Multiplicative Congruential Generator*):

$$x_n := (ax_{n-1}) \bmod m \quad (4.3)$$

- Lagged-Fibonacci Generator
(LFG – *Lagged-Fibonacci Generator*):

$$x_n := (x_{n-r} - x_{n-s}) \bmod m \quad (4.4)$$

- Gemischter Lagged-Fibonacci Generator
(MLFG – *Mixed Lagged-Fibonacci Generator*):

$$\left. \begin{aligned} x_n &:= (V_n - Y_n) \bmod m \\ V_n &:= (V_{n-r} - V_{n-s}) \bmod m \\ Y_n &:= (aY_{n-1} + b) \bmod m \end{aligned} \right\} \quad (4.5)$$

Der Modulus m der linearen Generatoren vom Typ LCG und MCG sollte möglichst groß gewählt werden. Die erzeugten Werte liegen zwischen 0 und $m - 1$, so daß die Periodenlänge maximal m sein kann. Aus Effizienzgründen benutzen die meisten

dieser Methoden für m eine Zweierpotenz, da die Modulo-Rechnung sich dann vereinfacht. Empfehlungen für die Wahl der Parameter findet man in [LO89, Knu81, LK91, Jam90, Mar85, PM88]. Lineare Kongruenz-Generatoren werden in den meisten Simulationssystemen bereitgestellt.

Der sog. Tausworthe-Generator [Tau65] ist ein Spezialfall des Lagged-Fibonacci Generators, der mit dem Modulo $m = 2$ als rückgekoppeltes Schieberegister realisiert werden kann.

Der gemischt lineare MLFG-Generator RANMAR bzw. MAV wird in der Physik z. B. auf dem Siemens-Vektorrechner im Rechenzentrum der RWTH Aachen für umfangreiche Monte-Carlo-Simulationen mit 10^9 bis 10^{11} Zufallszahlen je Simulationslauf eingesetzt, da er eine lange Periode und eine relativ hohe Generationsrate hat und zudem bei den statistischen Untersuchungen in [Jam90, Ric92] sehr gut abgeschnitten hat.

Nicht-lineare Generatoren:

- Inversive² Kongruenzgeneratoren [Hel95]
(ICG – *Inversive Congruential Generator*): $x_n := f(x_{n-1}^{-1})$
- Polynom-Generatoren

Polynom-Generatoren und inverse Generatoren sind zwei Beispiele für nicht-lineare Generatoren, die zum Teil sehr gute Eigenschaften haben.

Kryptologische Generatoren: Ebenfalls sehr gute, wahrscheinlich sogar quasi-ideale Eigenschaften hat der aus der Kryptologie abgeleitete sog. RSA-Generator [Ric92], der jedoch wegen des hohen Rechenaufwandes³ eine niedrige Generationsrate hat und daher — wie auch der nachstehend beschriebene *physikalische Generator* — nur als Tabellengenerator effizient zur Verfügung gestellt werden kann.

Physikalische Generatoren:

- radioaktiver Zerfall: [Gud87]
- elektronisches Rauschen: [Ric92]

Auch bei physikalischen Generatoren hat man die Möglichkeit Experimente zu wiederholen, wenn diese Zahlen auf Vorrat generiert werden und somit als Tabelle, z. B. als CD-ROM [Ric92], für beliebige Wiederholungen zur Verfügung stehen. Als Nachteil dieses Generatortyps ist die aufwendige Generierung der Tabellen zu nennen. Wenn die Versuchsanordnung einmal vorhanden ist, können beliebig viele Zahlen auf Vorrat generiert werden. Die Vorteile des unabhängigen Generierungsprinzips

²Das inverse Element x^{-1} bezieht sich hier auf die Modulorechnung, d. h. $(x * x^{-1}) \bmod m = 1$.

³Die Realisierung eines RSA-Generators mit einem aus 40 Transputern T800 bestehenden Multiprozessor-Netzwerk wurde am *Lehrstuhl für Angewandte Mathematik insbesondere Informatik* der RWTH von M. PORTZ [Por93] durchgeführt.

sowie das Fehlen einer Periode wiegen diesen Nachteil wieder auf, so daß dieser Generatortyp als eine Alternative in einem Simulationssystem zur Verfügung stehen sollte.

Generatoren in Simulationssystemen

In den meisten Simulationssystemen werden lineare Kongruenz-Generatoren angeboten. Teilweise sind Zufallsgeneratoren auch Bestandteil von Programmiersprachen und Betriebssystemen. Das Betriebssystem UNIX bietet einen linearen Kongruenz-Generator (`rand48`) sowie einen nicht linearen Generator (`random`) an. Das Programmsystem Maple zur Unterstützung von symbolischen und numerischen mathematischen Berechnungen enthält einen linearen Kongruenz-Generator [KG91] vom Typ MCG. In der CNCL-Simulationsbibliothek [JSBP96] werden ein linearer Kongruenz-Generator (CNLCG), ein gemischt linearer Generator (CNMLCG), ein Lagged-Fibonacci Generator (CNFiboG), ein additiver linearer Generator (CNACG), ein Tausworthe-Generator (CNTausG) sowie ein physikalischer Generator PURAN 2 (CNFileG) angeboten.

Generatortests

Basisgeneratoren müssen insbesondere bzgl. ihrer Grundeigenschaften – Gleichverteilung und Unabhängigkeit bzw. Korrelation – getestet werden. Ferner wird auf Eigenschaften getestet, die aus diesen Grundeigenschaften resultieren, z. B. durch den sog. *Permutationstest*, der die Häufigkeit der Permutationen einer bestimmten Länge testet oder Tests, die die Unabhängigkeit und Gleichverteilung von aufeinanderfolgenden Tupeln testen.

Beispiele für Tests findet man in [Knu81] und [Ric92]. Auch jedes Modell, zu dem analytische Ergebnisse bekannt sind, kann als Test der Zufallsgeneratoren verstanden werden. Um den Zufallsgenerator als Fehlerquelle möglichst ausschließen zu können, sollten in Zweifelsfällen mehrere Basisgeneratoren zur Verfügung stehen, die auf unterschiedlichen Prinzipien aufbauen. In [Knu81] wird der Grundsatz "*innocent, until proven guilty*" für Generatoren aufgestellt, d. h. mit jedem bestandenen Test wächst das Vertrauen in einen Generator, aber ein Test kann immer nur eine bestimmte Eigenschaft testen, so daß der Generator bei einem anderen Test versagen kann.

4.2 Generatoren für beliebige Verteilungen

Simulationsläufe benötigen Wertefolgen x_1, x_2, \dots verschiedener Zufallsvariablen X mit gewünschten Eigenschaften. Für unkorrelierte Zufallsvariablen genügt die Vorgabe der Verteilungsfunktion (VF) $F(x)$. Eine umfassende Darstellung der Verfahren findet sich in [Dev86].

Die sog. *inverse Methode* beruht darauf, daß eine mit der inversen Verteilungsfunktion $F^{-1}(u)$ transformierte, unabhängige, (0,1)-gleichverteilte Zufallsvariable zu einer unabhängigen Zufallsvariable führt, die der Verteilung $F(x)$ genügt. Für die Generierung von Zufallszahlen bedeutet dies, daß die Zahlen des Basisgenerators u_1, u_2, \dots direkt mit $F^{-1}(u_1), F^{-1}(u_2), \dots$ transformiert werden können, wenn die inverse Verteilungsfunktion als geschlossene Formel vorliegt. Für die Erzeugung von negativ-exponentiell verteilten Zufallszahlen benutzt man $F^{-1}(u) = -\ln(1-u)/\lambda$, siehe Gleichung (A.4).

Aufgrund spezieller Eigenschaften gibt es weitere Generierungsvorschriften für Verteilungen, die sich nicht einfach invertieren lassen. Zum Beispiel für die Normalverteilung wird das Verfahren nach Box und Muller [Dev86] eingesetzt, welches aufbauend auf einem Basisgenerator unter Verwendung transzendenter Funktionen (\sin, \cos, \ln) normalverteilte Zufallszahlen generiert.

Liegt die Verteilungsfunktion als Formel oder als Tabelle vor, so kann die inverse Methode approximativ numerisch durchgeführt werden. Dies gilt auch für Verteilungen, die als Meßergebnisse vorliegen.

Korrelierte Zufallssequenzen stellen einen zunehmend wichtigen Typ von Generatoren dar, z. B. zur Modellierung von Videoquellen. Auch dieser Typ Generator wird auf unkorrelierten Basisgeneratoren aufgebaut, da Korrelationseigenschaften gezielt vorgegeben werden müssen und nicht als Seiteneffekt schlechter Basisgeneratoren in unkontrollierter Art und Weise verwendet werden sollten. In [Din91a] wird ein Zufallsgenerator für korrelierte Zufallszahlen ausgehend von einer Markoff-Kette beschrieben, bei dem der lokale Korrelationskoeffizient nach Gleichung (3.44) vorgegeben werden kann, siehe Abschnitt 3.3.2.

4.3 Empfehlungen

In einem Simulationssystem sollten zwei bis drei verschiedene Basisgeneratoren — basierend auf verschiedenen Generierungsprinzipien — zur Verfügung gestellt werden, damit der Basisgenerator als Fehlerquelle möglichst ausgeschlossen werden kann. Der Benutzer muß die Eigenschaften des Generators — Periodenlänge, Einfluß der Startwerte, usw. — kennen, damit Fehler in der Anwendung, wie z. B. die mehrfache Verwendung desselben Generators in einem Programm mit demselben Startwert, vermieden werden können.

Verschiedene Zufallszahlenströme sollten nach Möglichkeit aus *einem* guten Basiszufallszahlenstrom erzeugt werden. Müssen Teilströme pseudozufällig exakt wiederholt werden, so ist eine sorgfältige Wahl der Startwerte des Generators für die verschiedenen Teilströme notwendig.

Die Verwendung des Zufallsgenerators zusammen mit allen Parametern — dazu gehören insbesondere die Startwerte — muß dokumentiert werden, damit Simulationsexperimente von anderen Wissenschaftlern nachvollzogen werden können, um so das Kriterium der Wiederholbarkeit eines wissenschaftlichen Experiments zu erfüllen.

Statistische Auswertung

Ein Hauptproblem bei der Anwendung der Simulationsmethode zur Leistungsuntersuchung ist die statistische Auswertung der bei der Simulation gemessenen – meist korrelierten – Ergebnisse. Um zu einer objektiven Fehleraussage zu kommen und ggf. die Länge des Simulationslaufes abhängig von der Genauigkeit zu steuern, benötigt man leistungsfähige Verfahren. Die meisten bekannten Verfahren haben gravierende Nachteile, wie z. B. die sehr häufig benutzte Auswertemethode *Batch-Means*, [Fis78, BFS87], welche versucht die Korrelation durch Einteilung in quasi-unabhängige und quasi-normalverteilte Gruppen zu eliminieren.

Die in Kommunikationsnetzen und bei deren Simulation auftretenden Zufallsgrößen, wie z. B. Warte- und Durchlaufzeiten, sind in starkem Maße durch Korrelationen zwischen zeitlich aufeinanderfolgenden Werten gekennzeichnet. Auswertemethoden, in denen neben der Verteilungsfunktion $F(x)$ auch die Korrelationseigenschaften der erzeugten Sequenzen gemessen werden können, sind daher notwendig.

Im folgenden Abschnitt wird die Initialisierung und Eliminierung der transienten Phase einer stationären Simulation betrachtet. Danach folgt eine kurze Betrachtung von terminierenden, nicht-stationären Simulationen. Der Schwerpunkt liegt bei den stationären Auswertemethoden, die in den letzten beiden Abschnitten ausführlich behandelt werden. Als erstes wird das häufig benutzte Batch-Means-Verfahren vorgestellt, um dann als abschließenden Abschnitt ein neueres Verfahren, das sog. LRE-Verfahren vorzustellen, welches zusätzlich ein Korrelationsmaß liefert.

5.1 Initialisierung und transiente Phase

Zu Beginn eines Simulationslaufs muß der Anfangszustand festgelegt werden. Da dieser Anfangszustand nicht notwendig einem typischen Zustand einer stationären Simulation entspricht, z. B. ein leeres lokales Netz oder eine leere Telefonvermittlung, sind statistische Auswertungen häufig erst nach einer sogenannten *transienten Phase* sinnvoll. Ist der Anfangszustand ein zulässiger Zustand, so ist auch die vollständige Auswertung vom Simulationsstart an korrekt. Die Simulationsdauer kann jedoch durch die Auswertung vom Start der Simulation an verlängert werden bzw. die Schätzung der Parameter kann dadurch verfälscht werden.

Ist x_1, x_2, \dots, x_n die auszuwertende Folge von Meßwerten, so ist das Ziel von Verfahren zur Reduzierung des Einflusses des Anfangszustands, einen Index r zu finden, so daß die Auswertung der Folge x_r, x_{r+1}, \dots, x_n den zu bestimmenden Parameter besser abschätzt als die vollständige Folge.

Probleme bei der Bestimmung der transienten Phase werden durch den Einfluß von Korrelationen und die Abhängigkeit des Verlaufs der transienten Phase vom Anfangszustand verursacht.

Häufig wird der Index vom Anwender geschätzt und z. B. beim Batch-Means-Verfahren, siehe Abschnitt 5.4, wird oft einfach die erste Gruppe in der Auswertung ignoriert.

Zur Bestimmung eines solchen Index gibt es verschiedene heuristische Verfahren [Lan92], die von einfachen graphischen Verfahren bis zu mathematisch sehr anspruchsvollen Verfahren reichen.

Ein einfaches graphisches Verfahren nach Welch [Lan92] nimmt eine Gruppeneinteilung vor und schätzt die transiente Phase durch die graphische Betrachtung der Gruppenmittelwerte ab. Dieses Verfahren läßt sich in einfacher Weise mit dem Batch-Means-Verfahren kombinieren. Die Hauptproblematik der Methode liegt in der Wahl der Gruppengröße und Anzahl der Gruppen, die durch Variation der Parameter in kurzen Pilotläufen ermittelt werden sollte. Hinweise dazu findet man in [Jai91].

Die meisten anderen Verfahren beruhen auf statistischen Tests. Bei dem Verfahren von Schruben [Lan92] wird z. B. der Einfluß des Anfangszustands auf den Erwartungswert der stationären Phase getestet, d. h. genauer es wird die Nullhypothese H_0 , daß kein Einfluß des Anfangszustands vorliegt, getestet. Beim Verfahren von Schruben, Singh & Thierney [Lan92] wird die Nullhypothese des Tests von Schruben gegen die Alternativhypothese H_1 , daß die Funktion des transienten Erwartungswertes eine bestimmte Form hat, getestet.

Viele dieser Verfahren benötigen a-priori Kenntnisse des Systems, z. B. das Vorzeichen der Abweichung oder die ungefähre Länge der transienten Phase muß bekannt sein. Die verschiedenen Ansätze benötigen unterschiedliche Parameter und zum Teil muß auch der Ablauf der Simulation geändert werden, bei manchen Verfahren werden z. B. k unabhängige Läufe statt eines Laufs benötigt. Bei anderen Methoden müssen alle Werte gespeichert werden.

Bei allen Verfahren gibt es Versuchsergebnisse, die unbefriedigend sind, so daß bislang keine allgemeingültigen Regeln angegeben werden können. Auch die empfohlenen Parameter der Verfahren bieten keine Garantie. Wenn möglich sollte man lieber einen langen Lauf statt vieler kurzer Läufe machen, da in [Gly96] gezeigt wird, daß wenige lange Läufe vielen kurzen asymptotisch überlegen sind, wenn die transiente Anfangsphase nicht gelöscht wird, d. h. es gehen alle Werte in die Auswertung ein.

Man sollte darauf achten, daß der Anfangszustand einen gültigen Systemzustand beschreibt. Bei der Simulation von offenen Wartesystemen kann man mit dem leeren System beginnen, welches ein einfacher und gültiger Zustand ist. Zusätzlich sollte eine graphische Validierung des zeitlichen Verlaufs der Simulationsergebnisse durchgeführt werden.

5.2 Terminierende Simulationen

Terminierende Simulationen untersuchen Systeme in transienten Phasen, bei denen sich die Parameter während der Simulation ändern, oder Systeme, die nur eine begrenzte Zeit laufen. In diesem Fall entfällt das Entfernen der transienten Anfangsphase.

Eine ähnliche Problematik wie bei Anfangsbedingungen und der transienten Anfangsphase kann auch bzgl. des Endzustands bzw. einer transienten Endphase bestehen. In diesem Fall kommen ähnliche Verfahren wie bei der Bestimmung der transienten Anfangsphase zur Anwendung.

5.3 Stationäre Auswertemethoden

In den meisten Simulationen wird eine *stationäre Leistungsbewertung* durchgeführt, d. h. eine Bewertung des Systems im Gleichgewicht, wobei Kennwerte und Verteilungen von stationären Prozessen gemessen werden.

Während eines Simulationslaufs entsteht für jede zu bestimmende Zufallsgröße des simulierten Systems, z. B. Wartezeit oder Durchlaufzeit, eine endliche Sequenz von Meßwerten x_1, x_2, \dots, x_n einer Zufallsvariablen X , die nach einem statistischen Verfahren zu bewerten ist. Die beim Anwendungsschwerpunkt *Kommunikationsnetze* typischen Annahmen und Gegebenheiten sind:

- Der zu untersuchende Zufallsprozeß ist *stationär*, d. h. der Sequenz von gemessenen x -Werten ist eine stationäre, zeitunabhängige Verteilungsfunktion $F_X(x)$ mit stationärem Erwartungswert $\mu = E[X]$ und anderen stationären Momenten zugeordnet.
- Aufgrund der Gedächtnisfunktion in Warternetzen sind die gemessenen Werte x_1, x_2, \dots, x_n untereinander *korreliert*.
- Vor Auswertung der Meßwerte sind nicht nur einzelne Parameterwerte wie z. B. der Erwartungswert $\mu = E[X]$ unbekannt, sondern es ist der *Typ des Zufallsprozesses* unbekannt. Das bedeutet, daß der gesamte Verlauf der Verteilungsfunktion $F_X(x)$ und die Korrelationen als *a-priori* unbekannt zu betrachten sind.

- In der Simulationstechnik liegt die Anzahl der gemessenen x -Werte typisch in der Größenordnung $n = 10^5 \dots 10^7$. Ein Kollektiv dieses Umfangs wird üblicherweise als *großer Stichprobenumfang* (*large sample*) bezeichnet.

Aufgrund dieser Gegebenheiten soll das statistische Auswerteverfahren unter Beachtung einer vorgegebenen *Fehlerschranke* folgende Aussagen aus den Meßwerten x_1, x_2, \dots, x_n gewinnen:

- die durch Simulation bestimmte Verteilungsfunktion $\tilde{F}_X(x)$ als Approximation der unbekanntenen stationären Verteilungsfunktion $F_X(x)$
- die zugehörigen gemessenen Momente $\tilde{X}^{(h)}$ für $h = 1, 2, \dots$, insbesondere der Schätzwert für den Erwartungswert $\tilde{X}^{(1)} = \tilde{\mu}$ der Zufallsvariablen X
- Aussagen über die Korrelationen der x -Sequenz
- ein *objektives Fehlermaß*, das durch Anforderung weiterer x -Werte solange reduziert werden kann, bis die vorgegebene Fehlerschranke im gesamten Verlauf der gemessenen Verteilungsfunktion $\tilde{F}_X(x)$ unterschritten wird und damit ein Abbruch der Simulation veranlaßt werden kann

Diese Statistikaufgabe ist sehr *anspruchsvoll*, da alle Vereinfachungen der Auswerteprozedur, die z. B. im Fall *unabhängige x -Werte* (keine Korrelationen) oder im Fall *Typ des Zufallsprozesses bekannt* möglich sind, hier nicht unmittelbar genutzt werden können. Andererseits kann ein wesentlicher Vorteil der Rechnersimulation ausgenutzt werden, der darin besteht, daß die Anzahl n der x -Werte steuerbar ist. Im folgenden werden die Meßwerte mit x_1, \dots, x_n bzw. y_1, \dots, y_n bezeichnet. Die Meßwerte werden auf eine entsprechende Folge von Zufallsvariablen X_1, \dots, X_n abgebildet, um die statistischen Eigenschaften der Sequenz zu betrachten.

Summenbildung und Mittelwert

Bei den meisten Statistikverfahren zur Untersuchung einer Zufallssequenz X_1, \dots, X_n wird die Summen-Zufallsvariable Y gebildet:

$$Y = \frac{1}{n} \sum_{i=1}^n X_i \quad \text{mit} \quad E[Y] = \mu = E[X] \quad (5.1)$$

Ein Hauptziel des Statistikverfahrens ist die Bestimmung des Erwartungswerts $\mu = E[X] = E[Y]$. Eine solche Vorgehens- und Betrachtungsweise beruht auf der Tatsache, daß in der Regel jede Verteilung einer Summen-Zufallsvariable Y für wachsendes n in die *Normalverteilung* übergeht.

In der Statistik bezeichnet man den *arithmetischen Mittelwert*

$$\tilde{\mu} = \frac{1}{n} \sum_{i=1}^n x_i \quad (5.2)$$

als einen *Schätzwert* $\tilde{\mu}$ für den Erwartungswert μ , der um so genauer den wahren Wert μ approximiert, je größer n ist.

Bestimmung der Verteilungsfunktion

In vielen Fällen ist es zweckmäßig, die gemessenen x -Werte zur Bestimmung der jeweiligen Verteilungsfunktion $F_X(x)$ auszuwerten. Hierzu ist eine Intervalleinteilung der x -Achse nicht unbedingt notwendig. Man kann den gemessenen *chronologischen* Vektor (x_1, x_2, \dots, x_n) durch Größensortierung in den *geordneten* Vektor

$$(x'_1, x'_2, \dots, x'_r, \dots, x'_n) \text{ mit } x'_r \leq x'_{r+1} \quad (5.3)$$

überführen, um an jeder beliebigen Stelle x der reellen Achse feststellen zu können:

- im Bereich $x'_1 < x < x'_n$ liegt *links* bzw. *rechts* von x eine Anzahl von r bzw. $n - r$ Meßwerten, $r = 1, 2, \dots, n - 1$
- im Bereich $x < x'_1$ liegen *links* von x *keine* Meßwerte: $r = 0$
- im Bereich $x > x'_n$ liegen *links* von x *alle* Meßwerte: $r = n$

Durch eine solche Einteilung der größensortierten Meßwerte ist eine Summen-Zufallsvariable definiert worden, deren zufällige Größe von der gewählten Stelle x und dem zugeordneten Wert der Verteilungsfunktion $F(x) = F_X(x)$ abhängig ist. Im Fall einer *unabhängigen* x -Sequenz wird die Wahrscheinlichkeit für das Ereignis, daß links von der Stelle x von insgesamt n Meßwerten r Werte liegen, offensichtlich durch die Wahrscheinlichkeitsfunktion $P(r)$ der Binomial-Verteilung

$$P(r) = \binom{n}{r} [F(x)]^r [1 - F(x)]^{n-r} \text{ für } r = 0, 1, \dots, n \quad (5.4)$$

beschrieben, siehe Gleichung (A.16).

Auf der Grundlage dieser Beziehung kann im Fall einer *unabhängigen* x -Sequenz eine Verteilungsfunktion $\tilde{F}_X(x)$ bestimmt werden. Im Fall einer *korrelierten* x -Sequenz, der in der Simulationstechnik fast immer gegeben ist, wird dieser Ansatz erweitert durch Übergang von dem hier angenommenen unabhängigen "0/1"-Zufallsprozeß auf den korrelierten "0/1"-Zufallsprozeß der 2-gliedrigen Markoff-Kette, siehe Abschnitt 3.2.2 und 5.5.

Bestimmung höherer Momente

Durch die Potenzfunktion $Z = X^\nu$ mit der inversen Funktion $X = Z^{1/\nu}$ wird X in die Zufallsvariable Z transformiert [Pap91]:

$$E[Z] = E[X^\nu] = Z^{(1)} = X^{(\nu)} \quad \text{mit } \nu = 1, 2, \dots \quad (5.5)$$

Daraus folgt, daß man in Erweiterung von Gleichung (5.1) zur Bestimmung des unbekanntenen Momentenparameters $X^{(\nu)}$ der Zufallsvariablen X aus den Meßwerten x_1, x_2, \dots, x_n die Summen-Zufallsvariable

$$\left. \begin{aligned} Y_\nu &= \frac{1}{n} \sum_{i=1}^n Z_i = \frac{1}{n} \sum_{i=1}^n X_i^\nu \quad \text{für } \nu = 1, 2, 3, \dots \\ E[Y_\nu] &= E[X^\nu] = X^{(\nu)} \end{aligned} \right\} \quad (5.6)$$

bilden und bezüglich des Erwartungswertes $E[Y_\nu] = X^{(\nu)}$ statistisch auswerten muß.

Bestimmung der Varianz

Für die statistische Bestimmung der unbekanntenen Varianz σ_X^2 der Zufallsvariablen X betrachtet man die Meßwerte x_1, x_2, \dots, x_n bzw. die Zufallsvariablen X_1, X_2, \dots, X_n , wobei jedes X_i der gleichen Verteilung wie X folgt. Gilt zusätzlich die Unabhängigkeit, dann kann man für die folgende Summen-Zufallsvariable V zeigen, daß ihr Erwartungswert der Varianz von X entspricht:

$$\left. \begin{aligned} V &= \frac{1}{n-1} \sum_{i=1}^n (X_i - Y)^2 = \frac{1}{n-1} \left(\sum_{i=1}^n X_i^2 - n \cdot Y^2 \right) \\ \text{mit } Y &= \frac{1}{n} \sum_{i=1}^n X_i \quad \text{und } E[V] = \sigma_X^2 \end{aligned} \right\} \quad n = 2, 3, \dots \quad (5.7)$$

Hier ist die Teilung durch $n-1$ (nicht durch n) erforderlich, weil in dem Summenwert Y der bei der Differenzbildung jeweils aktuelle Wert X_i bereits enthalten ist, siehe [Mor68]. In vielen Anwendungen von Gleichung (5.7) ist die Zufallsvariable X normalverteilt und die Summen-Zufallsvariable V ist daher χ^2 -verteilt, siehe Gleichung (A.7).

Dieser Schätzwert für die Varianz setzt unabhängige Stichproben voraus, was bei Simulationsdaten im allgemeinen nicht vorausgesetzt werden kann. In [LK91] heißt es: *“It has been our experience that simulation output data are always correlated. (If there are simulations with independent output data, we have never seen one.)”*

Zur Auswertung von Meßwerten eines Simulationslaufes werden meist Verfahren eingesetzt, die normalverteilte, unabhängige Sequenzen [WW93] voraussetzen. Durch

Summenbildung kann wie in Gleichung (5.1) erläutert eine approximative Normalverteilung mit gleichem Erwartungswert erzeugt werden. Da die Stichproben bei der Simulation im allgemeinen korreliert sind, wird versucht durch verschiedene Verfahren Unabhängigkeit zu erzielen. Das sind zum einen *unabhängige Simulationsläufe* mit unterschiedlichen Startwerten, oder die Summenbildung über Gruppen von einem Regenerationspunkt zum nächsten, wobei z. B. das leere Wartesystem einen Regenerationspunkt definiert, oder die Gruppenbildung (*batches*) einer hinreichend großen Anzahl von Werten. Im folgenden soll die am häufigsten benutzte Methode das sog. *Batch-Means-Verfahren* erläutert werden.

5.4 Das Batch-Means-Verfahren

Dieses typische Statistikverfahren, das auf Methoden der *Häufigkeitsstatistik* beruht, wird für die Analyse korrelierter Simulationsdaten eingesetzt. Eine ausführliche Darstellung und kritische Bewertung dieses und anderer Verfahren findet man in [LK91, BFS87, Jai91, Fis78].

Mittelwertbestimmung: Es ist die folgende allgemeine Aufgabe zu lösen: *Aus einer stationären x -Zufallssequenz, die in unbekanntem Ausmaß korreliert ist, soll der unbekannte Erwartungswert $\mu = E[X]$ bestimmt werden.*

Das Verfahren besteht im wesentlichen darin, die aus n Versuchen gewonnene *Gesamtsequenz* x_1, x_2, \dots, x_n in m *Teilsequenzen (batches)* $x_{j1}, x_{j2}, \dots, x_{jb}$ von jeweils gleicher Größe $b = n/m$ so aufzuteilen, daß die Summen-Zufallswerte

$$Y_j = \frac{1}{b} \sum_{i=1}^b X_{ji} \quad \text{mit} \quad E[Y_j] = \mu = E[X] \quad \text{für} \quad j = 1, 2, \dots, m \quad (5.8)$$

als *quasi-unabhängig* und *quasi-normalverteilt* angenommen werden können. Als typische Zahlenwerte können z. B. gelten: Aufteilung einer Gesamtsequenz mit $n = 240\,000$ x -Werten in $m = 12$ Teilsequenzen mit je $b = n/m = 20\,000$ x -Werten.

Zur weiteren Auswertung und Beurteilung, ob die vorgenommene Aufteilung zulässig war, werden aus den Y_j -Werten folgende Größen berechnet:

- Schätzwert $\tilde{\mu}$ für den gesuchten Erwartungswert $\mu = E[X]$ der x -Sequenz, vergleiche Gleichung (5.1):

$$\tilde{\mu} = \frac{1}{m} \sum_{j=1}^m y_j = \frac{1}{m b} \sum_{j=1}^m \sum_{i=1}^b x_{ji} = \frac{1}{n} \sum_{w=1}^n x_w \quad (5.9)$$

- Schätzwert $\tilde{\sigma}_Y^2$ für die Varianz σ_Y^2 der Y_j -Werte, mit $n = m$, vgl. Gleichung (5.7):

$$\tilde{\sigma}_Y^2 = \frac{1}{m-1} \sum_{j=1}^m (y_j - \tilde{\mu})^2 = \frac{1}{m-1} \left(\sum_{j=1}^m y_j^2 - m \tilde{\mu}^2 \right) \quad (5.10)$$

- Schätzwert $\tilde{\rho}_Y$ für den Korrelationskoeffizienten 1. Ordnung ρ_Y der Y_j -Werte

$$\tilde{\rho}_Y = \frac{1}{(m-1)\tilde{\sigma}_Y^2} \sum_{j=1}^{m-1} (y_j - \tilde{\mu})(y_{j+1} - \tilde{\mu}) \quad (5.11)$$

- Unter der Annahme, daß Y_j normalverteilt ist, bildet man die Zufallsvariable $t_m = \left(\frac{1}{m} \sum_{j=1}^m Y_j - \mu \right) / \sqrt{V(m)/m}$, die dann der Student-Verteilung oder t -Verteilung mit $m-1$ Freiheitsgraden genügt [LK91]. Zu einer vorgegebenen Konfidenzwahrscheinlichkeit P_C (z. B. $P_C = 0.95$) wird auf der Grundlage der t -Verteilung mit $m-1$ Freiheitsgraden und der Dichtefunktion

$$\left. \begin{aligned} f^*(t) &= \frac{\Gamma\left(\frac{m}{2}\right)}{\sqrt{\pi(m-1)} \Gamma\left(\frac{m-1}{2}\right)} \left[1 + \frac{t^2}{m-1} \right]^{-m/2} \\ \text{mit } t &= \frac{\mu - \tilde{\mu}}{\tilde{\sigma}_Y / \sqrt{m}} \text{ für } m = 2, 3, \dots \end{aligned} \right\} \quad (5.12)$$

ein Konfidenzintervall der Größe $\mu_1 \leq \tilde{\mu} \leq \mu_2$ gewonnen, siehe [Sch80] und die tabellierte Verteilungsfunktion $F^*(t)$ in [Kre73]. Dieses Konfidenzintervall kann als *relatives Konfidenz-Fehlermaß*

$$\varepsilon = (\mu_2 - \mu_1) / \tilde{\mu} \quad (5.13)$$

für den nach Gleichung (5.9) berechneten Wert $\tilde{\mu}$ verwendet werden.

Der Wert $\tilde{\mu}$ aus Gleichung (5.9) wird als Schätzwert für den hier zu bestimmenden Erwartungswert $\mu = E[X]$ akzeptiert,

- wenn der nach Gleichung (5.10) berechnete Schätzwert $\tilde{\rho}_Y$ als ausreichend klein beurteilt wird (Idealwert für Unabhängigkeit: $\tilde{\rho}_Y = 0$)
- und aufgrund der Batchgröße b die Annahme der Normalverteilung für die Zufallswerte Y_j gerechtfertigt erscheint
- und das relative Konfidenz-Fehlermaß ε eine vorgegebene Fehlerschranke ε_{max} unterschreitet, wenn also z. B. die Bedingung $\varepsilon < \varepsilon_{max} = 3\%$ erfüllt ist.

Falls die erste und/oder zweite Bedingung nicht erfüllt sind, kann man z. B. die Batchgröße b durch Zusammenlegen von je zwei benachbarten Teilsequenzen verdoppeln und n zusätzliche x -Werte produzieren, um so die benötigten m Teilsequenzen und Y_j -Werte aus Gleichung (5.8) zu gewinnen. Die Anzahl der insgesamt erzeugten x -Werte hat sich bei dieser Vorgehensweise verdoppelt. Falls jedoch nur die dritte Bedingung nicht erfüllt ist, genügt es, so viele weitere Teilsequenzen der bisherigen Größe b zu produzieren, bis die Anzahl m groß genug ist, so daß das Fehlermaß unter der vorgegebenen Fehlerschranke liegt.

Da eine Auswertung von Konfidenzintervallen für die Verteilungsfunktion aufgrund der Binomialverteilung nach Abschnitt 5.3 für große Stichproben sehr aufwendig ist, wird das Batch-Means-Verfahren in modifizierter Form auch zur Bestimmung von Konfidenzintervallen für die Werte der Verteilungsfunktion [LO89] eingesetzt.

Da das Batch-Means-Verfahren und auch Alternativvorschläge wie etwa die *Autoregressive Methode* [Fis78, LK91] in theoretischer und praktischer Hinsicht problematisch bleiben [Dev86, LK91], wurden neue Verfahren, wie das LRE-Verfahren entwickelt, welches im folgenden Abschnitt beschrieben wird.

5.5 Der LRE-Algorithmus

Die Auswertung von Meßdaten aus Simulationsläufen mit Hilfe des LRE-Algorithmus (LRE – *Limited Relative Error*) [SG96b] liefert die stationäre Verteilungsfunktion $F(x)$ oder die komplementäre Verteilungsfunktion $G(x)$ der untersuchten x -Sequenz und zusätzlich den sog. lokalen Korrelationskoeffizienten $\rho(x)$, $-1 \leq \rho(x) \leq 1$, der in ein mittleres quadratisches Fehlermaß $d(x)$ für die Steuerung der Simulationslaufzeit einbezogen wird. Der Gebrauch dieser Fehleraussage anstelle von Aussagen mittels Konfidenzintervallen und das Prinzip, Korrelationserscheinungen zu messen, statt sie zu eliminieren, sind wichtige Elemente, um die bekannten Unzulänglichkeiten des Batch-Means-Verfahrens zu beseitigen. Der LRE-Algorithmus ist mit Erfolg bei vielen Simulationsuntersuchungen [GKS91] eingesetzt worden.

In dieser Arbeit wird ein vereinfachter LRE-Algorithmus (LRE-III) erläutert, der speziell zur Auswertung von *diskreten* x -Sequenzen wie z. B. der Ankunftsbelegung eines endlichen Wartesystems vom Typ G/G/1/N geeignet ist. Die Vorteile des LRE-Algorithmus III gegenüber anderen Verfahren werden durch Simulation von analytisch darstellbaren Wartesystemen gezeigt und validiert.

Diese spezielle LRE-Version ist Teil des neuen Simulationssystems zur Untersuchung *seltener Ereignisse* [SG94, GS96] durch Kombination des sog. RESTART-Prinzips (RESTART – *REpetitive Simulation Trials After Reaching Threshold*) [VAVA91, Bay70] mit den LRE-Auswertungs-Prinzipien. In Kapitel 8 wird gezeigt, wie dieser Algorithmus mit der RESTART-Methode als RESTART/LRE-Algorithmus für

die Simulation seltener Ereignisse zur Steuerung der Länge des Simulationslaufes kombiniert werden kann.

Aufgrund seiner Einfachheit ist der LRE-Algorithmus III am besten geeignet, um die grundlegenden Eigenschaften des LRE-Verfahrens zu zeigen: das Messen der Korrelation und die Simulationssteuerung durch eine mittlere quadratische Fehlerformel.

Lokale Korrelation: Es wird im folgenden angenommen, daß die Wertefolge einer diskreten Zufallsvariable α auf die Zustandsübergänge der zufälligen Zustandssequenzen einer rekurrenten $(k+1)$ -gliedrigen Markoff-Kette mit den Übergangswahrscheinlichkeiten $p_{ji} = P\{j|i\}$ abgebildet wird. Die stationären Zustandswahrscheinlichkeiten $P_j = P_\alpha(j)$ werden eindeutig durch die Lösung des folgenden Gleichungssystems festgelegt:

$$\left. \begin{aligned} 0 &= \sum_{i=0}^k p_{ji} P_i - P_j & \text{für } j = 1, 2, \dots, k \\ P_0 &= 1 - \sum_{j=1}^k P_j & \text{für } k = 1, 2, \dots \end{aligned} \right\} \quad (5.14)$$

Die komplementäre Verteilungsfunktion $G(x)^1$, siehe Abbildung 3.10, ist damit:

$$\left. \begin{aligned} G(x) = G_i &= \sum_{j=[x]}^k P_j = \sum_{j=i}^k P_j & \text{für } i-1 \leq x < i \\ \text{und } i &= 1, 2, \dots, k & \text{mit } G_0 = 1 \text{ und } G_{k+1} = 0 \end{aligned} \right\} \quad (5.15)$$

Hier und im folgenden wird vorzugsweise eine Beschreibung der komplementären Verteilungsfunktion $G(x)$ verwendet statt der Verteilungsfunktion $F(x) = 1 - G(x)$, da in vielen Anwendungen insbesondere der Bereich für kleine Werte von $G(x)$, d. h. das rechte Ende der Verteilung von Interesse ist.

An jedem Punkt x der reellen Achse kann die Markoff-Kette in zwei Teile aufgespalten werden und man kann eine äquivalente x -bezogene 2-gliedrige Markoff-Kette definieren, deren Übergangswahrscheinlichkeiten $p_0(x)$ und $p_1(x)$ eindeutig durch die Gleichungen [Sch87a] gegeben sind:

$$\left. \begin{aligned} p_0(x) &= \frac{1}{1 - G_i} \sum_{r=0}^{i-1} P(r) \sum_{j=i}^k p_{jr} \\ p_1(x) &= \frac{1}{G_i} \sum_{r=i}^k P(r) \sum_{j=0}^{i-1} p_{jr} \end{aligned} \right\} \begin{aligned} i-1 \leq x < i \\ i = 1, 2, \dots, k \end{aligned} \quad (5.16)$$

Der lokale Korrelationskoeffizient $\varrho(x)$ dieser 2-gliedrigen Markoff-Kette und die Zustandswahrscheinlichkeiten $Q(x) = P_0(x)$ des linken Zustands $S_0(x)$ bzw. $1 - Q(x) =$

¹Man beachte, daß nach dieser Definition $G_i \neq G(i)$ ist, es gilt $G_i = P\{X \geq i\} = G(i-1)$.

$P_1(x)$ des rechten Zustands $S_1(x)$ lauten:

$$\left. \begin{aligned} \varrho(x) &= 1 - [p_0(x) + p_1(x)] \\ Q(x) &= p_1(x)/[1 - \varrho(x)] \equiv F(x) \\ 1 - Q(x) &= p_0(x)/[1 - \varrho(x)] \equiv G(x) \end{aligned} \right\} i - 1 \leq x < i \quad (5.17)$$

Daraus folgt die Gleichung

$$p_0(x) F(x) = p_1(x) G(x) \quad (5.18)$$

und man kann folgern, daß die beiden Summenausdrücke in Gleichung (5.16) gleich sind.

Die Funktion $\varrho(x)$, die erstmalig 1967 von BLOMQUIST [Blo67, Rey75] verwendet wurde, wird *lokal* genannt, da sie vom Ort x abhängt und sich dadurch vom globalen Korrelationskoeffizienten ϱ unterscheidet, der eine Art Mittelwert in Bezug auf alle x -Werte [DS90] darstellt, siehe Abschnitt 3.3.

Der LRE-Algorithmus III: Die komplementäre Verteilungsfunktion $G(x)$ einer $(k + 1)$ -gliedrigen Markoff-Kette mit bekanntem Wert für k und bekannten Übergangseigenschaften, aber unbekanntem Werten der Übergangswahrscheinlichkeiten p_{ji} soll durch Simulation bestimmt werden. Es ist also bekannt, daß eine diskrete Sequenz vorliegt mit einem bekannten Wertebereich. Die einzige Speicheranforderung für eine Auswertung der komplementären Verteilungsfunktion ist ein Feld zur Speicherung von 2 Zählern bzw. Häufigkeiten für jeden Zustand i der $(k + 1)$ -gliedrigen Kette nachdem diese Kette n Zustandswechsel durchlaufen hat:

- a) Die **Zustandshäufigkeit** h_i gibt an, wie oft jeder Zustand i in n Versuchen aufgetreten ist. Daraus erhält man die Summenhäufigkeit

$$v_i = \sum_{j=i}^k h_j \quad \text{für } i = 0, 1, \dots, k \quad \text{mit } v_0 = n \quad (5.19)$$

die angibt, wie oft der rechte Zustand $S_1(x)$ mit $i - 1 \leq x < i$ der 2-gliedrigen Kette aufgetreten ist.

- b) Die **Übergangshäufigkeit** c_i , $i = 1, 2, \dots, k$ gibt an, wie oft ein Übergang von $S_1(x) \rightarrow S_0(x)$, d. h. ein Übergang von jedem Zustand j im Bereich $j \geq i$ über die Trennlinie am Punkt x im Intervall $i - 1 \leq x < i$ stattgefunden hat.

Die Häufigkeiten v_i und c_i bzw. die entsprechenden Zähler $r_i = n - v_i$ und $a_i \approx c_i$ gehören zur äquivalenten x -bezogenen 2-gliedrigen Markoff-Kette wie in Abbildung 5.1 gezeigt.

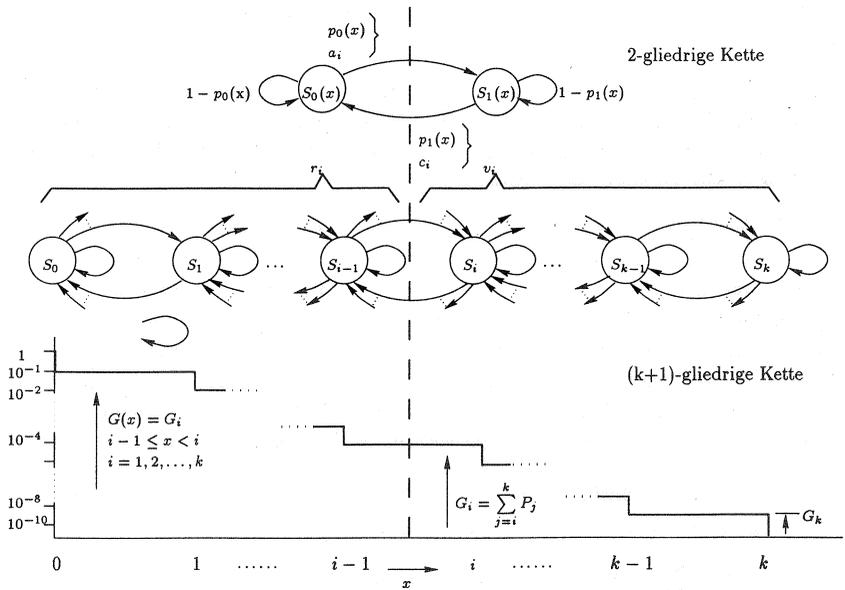


Abbildung 5.1: $(k+1)$ -gliedrige und äquivalente x -bezogene 2-gliedrige Markoff-Kette

Die Statistik der zweigliedrigen Markoff-Kette: Nachdem die Häufigkeiten v_i und c_i in n Versuchen gemessen wurden, können die Ergebnisse der statistischen Untersuchungen aus [Sch86, Sch87b] bzgl. der 2-gliedrigen Markoff-Kette angewandt werden. Falls die *Large-Sample*-Bedingungen:

$$n \geq 10^3; (r_i, v_i) \geq 10^2; (a_i, c_i, r_i - a_i, v_i - c_i) \geq 10 \quad (5.20)$$

erfüllt sind, können die benötigten Aussagen bezüglich

- der gemessenen komplementären Verteilungsfunktion $\tilde{G}(x)$
- des gemessenen Zustandsmittelwertes $\tilde{\alpha}$
- des gemessenen lokalen Korrelationskoeffizienten $\tilde{\varrho}(x)$
- des zugehörigen Korrelationsfaktors (*correlation factor*) $\tilde{c}f(x)$

$$\tilde{c}f(x) = \frac{1 + \tilde{\varrho}(x)}{1 - \tilde{\varrho}(x)} \quad (5.21)$$

- und der relative Fehler $d_G(x) = \sigma_G(x)/\tilde{G}(x)$ in Bezug auf die gemessene Funktion $\tilde{G}(x)$

durch folgende Formeln beschrieben werden:

$$\left. \begin{aligned} \tilde{G}(x) = \tilde{G}_i &= v_i/n \\ \tilde{\alpha} &= \frac{1}{n} \sum_{i=1}^k v_i \\ \tilde{\varrho}(x) = \tilde{\varrho}_i &= 1 - \frac{c_i/v_i}{1 - v_i/n} \\ \tilde{c}f(x) = \tilde{c}f_i &= (1 + \tilde{\varrho}_i)/(1 - \tilde{\varrho}_i) \\ d_G(x) = d_i &= \left(\frac{1 - v_i/n}{v_i} \cdot \tilde{c}f_i \right)^{1/2} \end{aligned} \right\} \begin{array}{l} \text{für } i - 1 \leq x < i \\ \text{und } i = 1, 2, \dots, k \end{array} \quad (5.22)$$

Aufgrund der großen Versuchszahl ergibt sich, daß die Niveaus \tilde{G}_i der gemessenen Funktion $\tilde{G}(x)$ einfach durch die relativen Häufigkeiten v_i/n angegeben werden können. Der wesentliche Fortschritt dieser Methode wird durch die Einbeziehung der Übergangshäufigkeiten c_i in die Auswertungsroutine erreicht. Man erhält dadurch nicht nur die Korrelationsfunktion $\tilde{\varrho}(x)$ sondern auch den mittleren quadratischen relativen Fehler $d_G(x)$ als Funktion von $\tilde{\varrho}(x)$ in einer durchgehenden nicht unterteilten Auswertung, die eine einfache Erhöhung der Anzahl der Versuche n erlaubt bis die *Fehlerbedingung* $d_i \leq d_{max}$ erfüllt ist, wobei d_{max} der vorgegebene maximale Fehler ist.

Die Gesamtversuchszahl n wird bestimmt durch d_{max} und den kleinsten gewünschten Wert der komplementären Verteilungsfunktion, der etabliert werden soll, nämlich $G_{min} = \tilde{G}_k$ im vorliegenden Fall mit $d_i = d_k = d_{max}$ und $\tilde{c}f_i = \tilde{c}f_k$ nach Gleichung (5.22):

$$n = \frac{(1 - G_{min}) \tilde{c}f_k}{G_{min} d_{max}^2} \approx \frac{\tilde{c}f_k}{G_{min} d_{max}^2} \quad \text{mit} \quad \tilde{c}f_k = \frac{1 + \tilde{\rho}_k}{1 - \tilde{\rho}_k} \quad (5.23)$$

Diese Formel kann zur approximativen Bestimmung der Simulationslaufzeit für vorgegebene Werte von G_{min} und d_{max} sowie für eine geschätzte Vorgabe von $\tilde{\rho}_k$ benutzt werden, wobei in der Praxis meist $0 \leq \tilde{\rho}_k < 1$ gilt.

Nimmt man z. B. $G_{min} = 10^{-4}$ und $d_{max} = 0.05$ oder 5 % bei $\tilde{\rho}_k = 0.8$, d. h. der Korrelationsfaktor ist $\tilde{c}f_k = 9$, so ergibt sich, daß ungefähr $n = 3.6 \times 10^7$ Versuche notwendig sind. Diese Versuchszahl ist in Abhängigkeit vom Modell bei der inzwischen verfügbaren Rechenleistung meist in weniger als einer Stunde durchführbar, siehe Abbildung 8.1.

Eine Verringerung des kleinsten Wertes auf z. B. $G_{min} = 10^{-9}$ würde die Zahl der Versuche auf den Wert $n = 3.6 \times 10^{12}$ anheben. In diesem Fall führt die direkte Simulation nicht mehr in angemessener Zeit zum Ziel, so daß Verfahren zur Beschleunigung der Simulation eingesetzt werden müssen, wie sie z. B. in Kapitel 7 und 8 beschrieben werden.

Die Durchführung des LRE-Algorithmus III: Die Parameter der Simulation seien mit d_{max} , α_{min} und α_{max} vorgegeben. α_{min} und α_{max} müssen nicht notwendig mit dem Wertebereich von α übereinstimmen. Der LRE-Algorithmus III wertet die α -Sequenz $\alpha_1, \alpha_2, \dots, \alpha_n$ aus, die auf die $(k + 1)$ -gliedrige Markoff-Kette abgebildet werden kann, und erhöht die Anzahl der Versuche n solange, bis alle Niveaus der gemessenen komplementären Verteilungsfunktion $\tilde{G}(x)$ mit einem relativen Fehler $d_i \leq d_{max}$ etabliert wurden. Bei der Laufzeitkontrolle bezeichnet s die Grenzmarke oder den Index, der als nächster bzgl. der Fehlervorgaben überprüft werden soll. In der Beschreibung wird $s = 1$ initialisiert, d. h. die Überprüfung erfolgt bzgl. der betrachteten komplementären Verteilungsfunktion von links nach rechts. In der folgenden Version des Algorithmus bezeichnet ω den vorhergehenden Wert der Sequenz, und es wird $\alpha_{min} = 0$ und $\alpha_{max} = k$ angenommen.

Parameter und Variablen des LRE-Algorithmus III:

- Parameter d_{max} : Vorgabe des maximalen LRE-Fehlers
- Parameter α_{min} : untere Grenze des Auswertungsbereichs
- Parameter α_{max} : obere Grenze des Auswertungsbereichs
- Variable s : Index, von dem an die Werte zu überprüfen sind
- Variable ω : vorhergehender Wert von α

LRE-Algorithmus III:**Initialisierung:**

Setze alle Zähler h_i und c_i auf Null für $i = 0, \dots, k$;

Setze $\alpha := 0$ und $n := 0$;

Setze Index der Laufzeitsteuerung $s := 1$;

Auswertung und Laufzeitsteuerung:

begin loop L_1

Setze $\omega := \alpha$; Übernehme neuen Wert für α ; Erhöhe Zähler h_α ;

if $\alpha < \omega$

then Erhöhe Zähler $c_{\alpha+1}$ bis c_ω

begin loop L_2

Setze $v_s := \sum_{i=s}^k h_i$; $r_s := n - v_s$; $a_s := c_s$;

if (Bedingung Gleichung (5.20) für $i = s$ erfüllt)

then

begin

Berechne \tilde{q}_s und rel. Fehler d_s nach Gl. (5.22);

if $d_s \leq d_{max}$

then if $s = k$ **then** Verlasse Schleife L_1 **else** Setze $s := s + 1$

else Verlasse Schleife L_2 ;

end

end loop L_2

end loop L_1

Terminierung:

for $i = 1, 2, \dots, k$

begin

Berechne Summenhäufigkeiten v_i nach Gl. (5.19);

Berechne Ergebniswerte \tilde{G}_i , \tilde{q}_i und d_i nach Gl. (5.22);

end

In einer Implementierung des LRE-Algorithmus III kann weiter optimiert werden, indem die Abfrage der Laufzeitsteuerung nicht nach jedem Wert sondern erst nach einem ganzen Block von Werten erfolgt. Die Blocklänge kann fest vorgegeben oder aufgrund der bisher gemessenen Werte dynamisch bestimmt werden. Außerdem kann ein Wert G_{min} vorgegeben werden, so daß nur gemessene Werte $\tilde{G}_i \geq G_{min}$ ausgewertet werden.

Simulationsergebnisse: In diesem Abschnitt werden beispielhaft Simulationsergebnisse vorgestellt, die mit dem LRE-Algorithmus III ausgewertet wurden. Ausgewählt wurden zwei Warteräume, die eine Validierung der Ergebnisse durch analytische Berechnung erlauben.

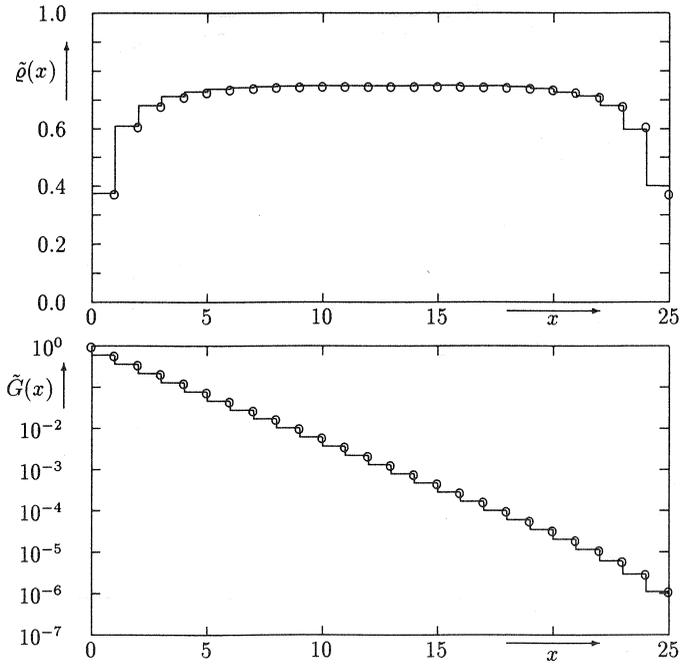
Zu dem elementaren Wartesystem M/M/1/N mit endlicher Anzahl von Wartepätzen N kann die komplementäre Verteilungsfunktion $G(x)$ und der lokale Korrelationskoeffizient $\varrho(x)$ der Belegung $\alpha = 0, 1, \dots, k$ mit $k = N + 1$ berechnet werden, siehe Abschnitt 3.4.4.

$$\left. \begin{aligned} G(x) = G_i &= \frac{\eta^i - \eta^{k+1}}{1 - \eta^{k+1}} \\ \varrho(x) = \varrho_i &= 1 - \frac{(1 - \eta)(1 - \eta^{k+1})\eta^i}{(1 + \eta)(1 - \eta^i)(\eta^i - \eta^{k+1})} \\ &\text{für } i - 1 \leq x < i \text{ und } i = 1, 2, \dots, k \end{aligned} \right\} \quad (5.24)$$

Abbildung 5.2 zeigt die Ergebnisse eines Simulationslaufs mit Auswertung durch den LRE-Algorithmus III. Die Genauigkeit der gemessenen Werte von $\tilde{G}(x)$ und $\tilde{\varrho}(x)$ werden durch den vorgegebenen maximalen LRE-Fehler $d_{max} = 0.05$ bestimmt. Diese beiden diskreten Treppenfunktionen stimmen mit den in Gleichung (5.24) angegebenen analytischen Funktionen überein. In Abbildung 5.2 wurden die analytisch berechneten Werte als Punkte eingetragen, um den Vergleich zu erleichtern. Nur die Funktion des lokalen Korrelationskoeffizienten $\tilde{\varrho}(x)$ zeigt kleine Abweichungen. Die analytische Kurve ist symmetrisch mit $\varrho_1 = \varrho_k$. Abweichungen bei $x = 24$ und $x = 25$ erklären sich durch die sehr kleinen Wahrscheinlichkeiten für diese Zustände. Diese Ergebnisse können mit dem in dieser Arbeit entwickelten RESTART/LRE-Verfahren, siehe Kapitel 8, und einem kleineren Wert von d_{max} verbessert werden. Eine RESTART/LRE-Simulation mit $d_{max} = 0.01$ ergab für $\tilde{\varrho}_{24}$ und $\tilde{\varrho}_{25}$ eine prozentuale Abweichung von unter 1 %.

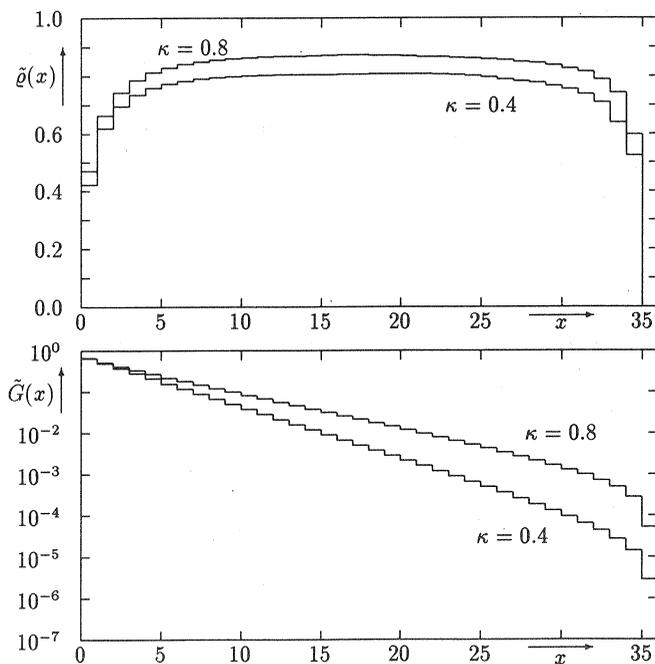
Unter Verwendung der Werte $d_{max} = 0.05$, $G_{min} \approx 1.1 \times 10^{-6}$ und $\varrho_k \approx 0.4$ kann die Versuchsanzahl mit Formel (5.23) zu ungefähr 8.4×10^8 berechnet werden. Wenn man die in dieser Arbeit verwendete realistische Rate von ca. 10^4 Versuche/sec für ein einfaches Modell zugrundelegt, so führt dies auf eine Simulationslaufzeit von ca. einem Tag (23.3 h).

Abbildung 5.3 zeigt die gemessenen Funktionen $\tilde{G}(x)$ und $\tilde{\varrho}(x)$ eines endlichen Wartesystems SSMP^[X](Geo,Geo)/D/1/N, wobei der Ankunftsprozeß ein *Special Semi-Markov Process*, siehe Abschnitt 3.3.2, mit Batchankünften bezeichnet, der einen korrelierten Ankunftsstrom z. B. eines ATM-Netzes beschreibt [DD90]. Die Bediendauer ist konstant (D) entsprechend der konstanten Länge einer ATM-Zelle.



komplementäre Verteilungsfunktion $\tilde{G}(x)$	lokaler Korrelationskoeffizient $\tilde{\varrho}(x)$
Simulation: Treppenkurve	Analytische Werte für $G(x)$ und $\varrho(x)$: \circ
Warteplätze: $N = k - 1 = 24$	Maximale Belegung: $B = k = 25$
Verkehrslast: $\eta = 0.6$	Verluste: $P_B = G_B \approx 1.1 \times 10^{-6}$
Max. Fehler: $d_{max} = 0.05$	Versuchszahl: $n \approx 8.4 \times 10^8$

Abbildung 5.2: LRE-Simulation: Warteraum M/M/1/N



komplementäre Verteilungsfunktion $\tilde{G}(x)$

Warteplätze: $N = k - 1 = 35$

Verkehrslast: $\eta = 0.7$

Korrelationskoeffizient τ_a : $\kappa = 0.4$

Verluste: $P_B = G_B \approx 3 \times 10^{-6}$

Versuchszahl: $n \approx 6 \times 10^6$

Analytisch berechnete Werte: siehe [Her94]

Lokaler Korrelationskoeffizient $\tilde{g}(x)$

Bediendauer: $D = 2$

Max. Fehler: $d_{max} = 0.05$

bzw. $\kappa = 0.8$

bzw. $\approx 5 \times 10^{-5}$

bzw. $n \approx 90 \times 10^6$

Abbildung 5.3: LRE-Simulation: Wartesystem $SSMP^{[X]}(\text{Geo}, \text{Geo})/D/1/N$

Dieses komplexe Wartesystem wurde in [Her94] analytisch untersucht und ein Vergleich der Diagramme 4 und 5 in [Her94] mit Abbildung 5.3 zeigt, daß der LRE-Algorithmus III auch dieses komplexe System – innerhalb der durch $d_{max} = 0.05$ vorgegebenen Genauigkeit – richtig auswertet.

Aufgrund seiner elementaren Struktur zeigt der LRE-Algorithmus III auf einfache Weise das Prinzip des *Limited Relative Error*, nämlich wie aus den Meßdaten der lokale Korrelationskoeffizient $\rho(x)$ bestimmt wird und wie daraus das korrelationsabhängige Fehlermaß zur Steuerung der Länge der Simulation in Abhängigkeit von der Genauigkeit gewonnen wird.

Die Entwicklung des LRE-Algorithmus: Der LRE-Algorithmus wurde zuerst für unabhängige Sequenzen entwickelt (LRE-I,[Sch84]). Die Auswertung beliebiger, korrelierter, kontinuierlicher und diskreter – auch gemischter – Sequenzen wird mit dem LRE-II ermöglicht [Sch88a], dessen Haupteigenschaften folgendermaßen charakterisiert werden können:

- Der Algorithmus bestimmt die Verteilungsfunktion $F(x)$ bzw. die komplementäre Verteilungsfunktion $G(x) = 1 - F(x)$ einer stationären, aber sonst nicht bekannten x -Sequenz nicht durch Unterteilung der x -Achse, sondern durch Unterteilung der F - bzw. G -Achse entsprechend der vorgegebenen Auflösung, d. h. der gewünschten Anzahl von Intervallen k und dem kleinsten Wert F_{min} bzw. G_{min} , der bei der Simulation bestimmt werden soll (z. B. $G_{min} = 10^{-3}$).
- Der Charakter der x -Sequenz kann diskret, kontinuierlich oder gemischt kontinuierlich-diskret sein. Der LRE-Algorithmus II kann innerhalb der gewählten Auflösung Ereignisse und isolierte diskrete Punkte sowie diskrete Punkte innerhalb einer kontinuierlichen Umgebung als Teil der Funktion $F(x)$ bzw. $G(x)$ aufdecken.
- Anstatt wie bei anderen Verfahren zu versuchen die Korrelation zu eliminieren, wird als Korrelationsmaß ständig die sog. lokale Korrelation $\rho(x)$ gemessen und in das Fehlermaß miteinbezogen, siehe Abschnitt 3.3.
- Zu jedem Zeitpunkt einer Simulation wird $\tilde{F}(x)$ bzw. $\tilde{G}(x)$ durch Auswertung der x -Sequenz – unter Einhaltung der *Large-Sample*-Bedingungen – bestimmt. Das Fehlermaß wird nicht durch Konfidenzintervalle angegeben, sondern durch ein mittleres quadratisches Fehlermaß, welches unter anderem von dem gemessenen Wert des lokalen Korrelationskoeffizienten $\tilde{\rho}(x)$, der gemessenen Varianz $\sigma_{\tilde{F}}^2(x)$ bzw. $\sigma_{\tilde{G}}^2(x)$ und daraus abgeleitet der *relativen Fehlerfunktion* $d_F(x) = \sigma_F(x)/\tilde{F}(x)$ bzw. $d_G(x) = \sigma_G(x)/\tilde{G}(x)$ abhängt.
- Ein Simulationslauf wird beendet, wenn im vorgegebenen Bereich der Untersuchung für F bzw. G der relative Fehler $d_F(x)$ bzw. $d_G(x)$ unter einem vorgegebenen Fehlerlimit d_{max} liegt. Um dies zu erreichen, wird die Anzahl der Versuche n ständig erhöht bis $d_F(x) \leq d_{max}$ bzw. $d_G(x) \leq d_{max}$ erfüllt ist.
- Die gemessenen Funktionen $\tilde{F}(x)$, $\sigma_{\tilde{F}}^2(x)$ bzw. $\tilde{G}(x)$, $\sigma_{\tilde{G}}^2(x)$ implizieren Aussagen über die Erwartungswerte (Momente) $E\{X^\nu\}$, $\nu = 1, 2, \dots$ der x -Sequenz.

Der LRE-Algorithmus II wurde als Teil der CNCL-Simulationsbibliothek [JG93] implementiert. Die Simulationsergebnisse z. B. in [PS93] zeigen, daß diese Auswertemethode zusätzliche Einsichten in das Verhalten eines Wartesystems ermöglicht, da neben der Verteilungsfunktion $F(x) = 1 - G(x)$ auch die Funktion $\varrho(x)$ als quantitative Aussage über die Korrelation der untersuchten x -Sequenz zur Verfügung steht.

Die Erfahrung aus vielen Simulationsuntersuchungen zeigt, daß bisherige Auswerteprobleme korrelierter x -Sequenzen weitgehend durch die Benutzung des LRE-Algorithmus eliminiert werden können. Der lokale Korrelationskoeffizient $\varrho(x)$ stellt die relevante Korrelationsinformation zur Verfügung.

Die Einführung der Funktion $\varrho(x)$ in die Statistik und Simulation einerseits und in die Wahrscheinlichkeitstheorie und analytischen Verfahren andererseits ermöglicht die Untersuchung der grundlegenden Abhängigkeiten und Gemeinsamkeiten dieser Gebiete.

Eine Hauptanwendung des LRE-Algorithmus, insbesondere des LRE-III, bezieht sich auf die Simulation seltener Ereignisse. Durch Angabe der Fehlerformel beim RESTART/LRE-Algorithmus, wie in Kapitel 8 beschrieben, wird eine analytische Betrachtung der optimalen Durchführung des Verfahrens ermöglicht.

Modelle für Kommunikationsnetze und zugehörige Ergebnisse

Ein Hauptanliegen dieser Arbeit ist die Beschreibung verschiedenartiger Modelle, die als Grundlage für analytische und simulative Untersuchungen von Kommunikationsnetzen gewählt werden. Genauso wie sich die Technik weiterentwickelt, so müssen auch die Modelle ständig angepaßt und auch neue Analyse- und Simulationstechniken entwickelt werden. Insbesondere die Entwicklung von Breitband- und Mobilfunknetzen hat hier wesentliche neue Impulse gegeben.

In den folgenden Abschnitten werden anhand von Beispielen Modelle für Kommunikationsnetze dargestellt, wobei das Hauptgewicht entsprechend der Komplexität der Anwendung auf den simulativen Untersuchungen liegt. Als Beispiel für ein lokales Hochgeschwindigkeitsnetz wird das FDDI-System vorgestellt. Als Beispiel einer Untersuchung für Weitverkehrsnetze werden neue Telekommunikationsdienste basierend auf der Technik der Intelligenzen Netze vorgestellt und untersucht. Abschließend wird die Problematik bei Mobilfunknetzen am Beispiel des DECT-Systems kurz umrissen.

Modelle von ATM-Vermittlungssystemen zur Behandlung der Problematik beim neuen Breitband-ISDN werden in Abschnitt 8.7 nach der Vorstellung der Techniken zur Behandlung seltener Ereignisse präsentiert.

6.1 Hochgeschwindigkeitsnetz FDDI

Das Haupteinsatzgebiet von FDDI (FDDI – *Fiber Distributed Data Interface*) liegt zur Zeit im Bereich der LAN-Kopplung. FDDI-Glasfasernetze mit 100 *Mbit/s* werden als *Backbone* zur Verbindung von Ethernet- oder Token-Ring-Inseln installiert. An der RWTH Aachen wurde 1990 eines der ersten großen FDDI-Verbindungsnetze¹ installiert. Teilweise wird bei sehr hohen Übertragungsanforderungen von Multimedia-Anwendungen FDDI auch für die direkte Verbindung leistungsfähiger Workstations und Mainframes eingesetzt. Vor allem netzweite Dateisysteme (z. B. NFS – *Network File System*), die den Einsatz plattenloser Workstations ermöglichen, erfordern eine leistungsfähige Rechnerkopplung.

¹[Hau91] beschreibt den Stand von 1991. Zur Zeit (Stand: Februar 1997) gibt es 23 Anschlußpunkte mit 40 Routern bzw. Bridges mit Übergängen zu Lokalen Netzen in Instituten, worüber insgesamt 250 Institutsnetze mit ca. 5000 Endgeräten angeschlossen sind.

FDDI stellt eine direkte Weiterentwicklung des Token-Ring dar. Das Kanalzugriffsprotokoll ist ein *Timed Token Protokoll* mit Prioritäten. Das Senderecht des Tokenhalters wird von einem *Token Holding Timer* (THT) begrenzt, der mit der *Target Token Rotation Time* (TTRT) verglichen wird. Zusätzlich gibt es in jeder Station einen *Token Rotation Timer* (TRT), der die Zeit vom letzten Token bis zum nächsten Token aus Sicht der Station mißt. Sobald der TRT den Wert des TTRT überschreitet, wird mit einem sog. *Late Count* für den laufenden Zyklus die Häufigkeit der Überschreitung in der Station gezählt. Eine Station erhält nur dann das Senderecht, wenn bei Erhalt des Tokens der *Late Count* im letzten Zyklus nicht gesetzt wurde, siehe Abbildung 6.4.

Mittlere Wartezeit des FDDI-Zugriffsprotokolls

Für den Modus *Exhaustive Service*, d. h. daß eine Station solange überträgt bis keine Pakete mehr zur Übertragung anstehen ($TTRT \rightarrow \infty$), und bei geringer Auslastung auch für das Protokoll mit Timersteuerung (d. h. $TTRT \ll \infty$) gilt für die mittlere Wartezeit $E[\tau_w]$ nach [HO86] bei N Stationen und einer Ringlänge l_R die Formel:

$$E[\tau_w] = \frac{z_0(1 - \frac{\eta}{N})}{2(1 - \eta)} + \frac{\eta(c_b^2 + 1)E[\tau_b]}{2(1 - \eta)} \quad (6.1)$$

Die Auslastung η des Gesamtsystems, die Zykluszeit des leeren Rings z_0 und die Ringlatenz L_R sind gegeben durch:

$$\eta = N \cdot \lambda \cdot E[\tau_b] \quad (6.2)$$

$$z_0 = \frac{L_R + N \cdot L_T}{v} \quad (6.3)$$

$$L_R = \frac{l_R}{c_M} v + N \cdot L_S \quad (6.4)$$

Hierin bezeichnet c_M die Ausbreitungsgeschwindigkeit, L_S die Stationslatenz und v die Übertragungsrate, siehe Tabelle 6.1.

Die Berechnung des Mittelwertes nach obenstehender Gleichung kann zur Überprüfung von Simulationsergebnissen herangezogen werden. In [SI95] findet man eine Übersicht über analytische Ergebnisse zu verschiedenen Modellen von Lokalen Netzen, z. B. Token Ring, Token Bus, CSMA/CD-Netze, Sternnetze, sowie deren Simulationsmodelle. Approximationsformeln für den Durchsatz und die mittlere Wartezeit des FDDI-Zugriffsprotokolls mit Timersteuerung werden in [Tan91] abgeleitet und in [Meh92] mit Simulationsergebnissen validiert.

Modellierung und Simulation des FDDI-Zugriffsprotokolls

Vereinfacht kann man sich das Zugriffsverfahren des FDDI-Rings als geschlossenes Gleisoval vorstellen, auf dem maximal eine freie Lokomotive kreist, siehe Abbildung 6.1. Sobald die freie Lokomotive eine Station erreicht, kann die Station im Rahmen der Zeitvorgaben Pakete senden, wobei auch die Lokomotive einen Datenrahmen trägt und somit belegt ist. Die Datenrahmen, die zusammenhängend von einer Station gesendet werden, entsprechen einem Zug. Das Senderecht gibt die Station in Form einer freien Lokomotive (Freitoken) an die nächste Station weiter. Abbildung 6.2 zeigt das Modell als Warteschlangensystem. Abbildung 6.4 und 6.5 beschreiben in Abhängigkeit vom Zustand der Station den Ablauf des Protokolls innerhalb der Station formal in einer SDL-ähnlichen Notation. Das Beispiel zeigt die Modellierung der asynchronen Klasse ohne Prioritäten wie sie im Simulationspraktikum [GDJL96] behandelt wird.

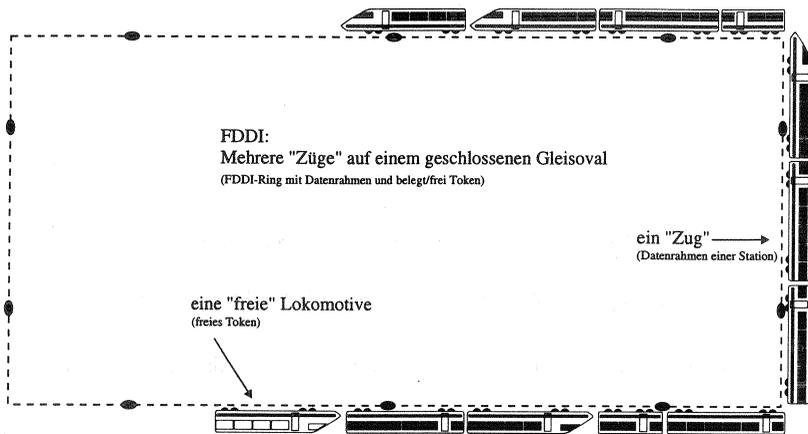


Abbildung 6.1: FDDI als Zugmodell

Abbildung 6.3 zeigt einen Vergleich der mittleren Transferzeiten als Funktion des Verkehrsangebots η für verschiedene Strategien [SG89, Sch91b] beim FDDI-Zugriffsprotokoll. Ausgewertet wurde ein Verkehrsgemisch verschiedener Dienste in einem realistischen Netz mit 86 Stationen der Länge 100 km. Man sieht, daß die mittlere Transferzeit $E[\tau_d]$ reduziert werden kann, wenn man das SRPT-Prinzip einsetzt. Bezogen auf Nachrichten und Pakete bedeutet das, daß als nächstes das erste Paket der Nachricht mit der kleinsten Anzahl Restpakete übertragen wird. Damit werden insgesamt kurze Nachrichten vor langen bevorzugt. Man kann das Prinzip nur auf die Stationswarteschlangen beziehen (SRPN – *Shortest Rest Packet Number*) oder auf das gesamte Netz ausdehnen, indem man die Prioritäten des FDDI entsprechend

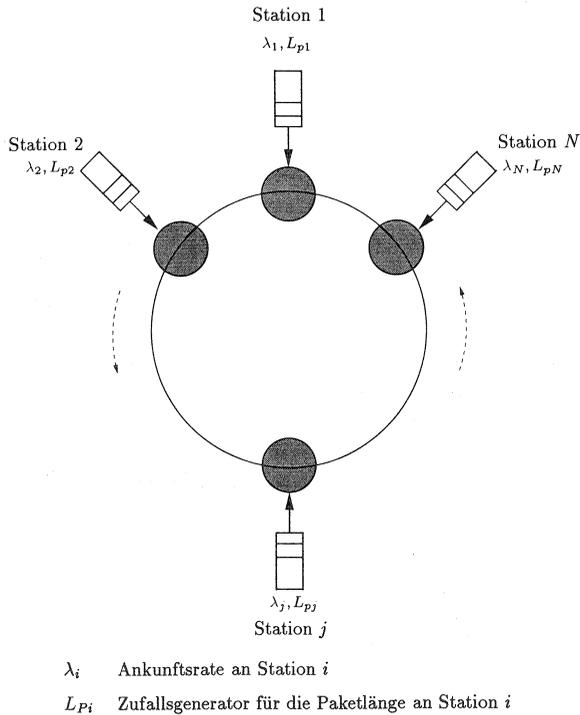


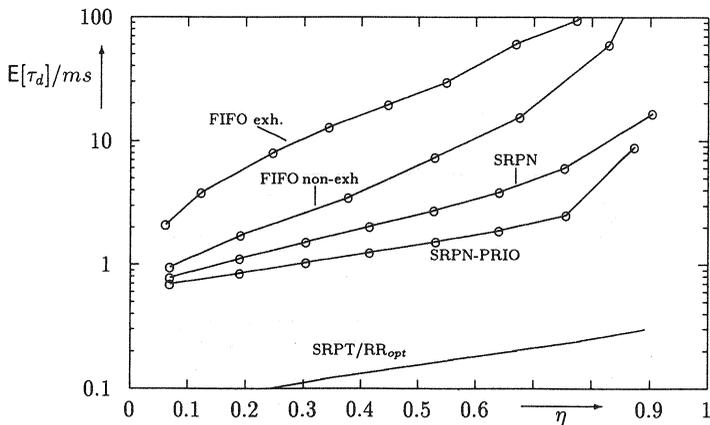
Abbildung 6.2: Token-Passing-Netz als Warteschlangensystem

der Restpaketnummer vergibt. Die unterste Kurve zeigt das theoretische Minimum, welches nach [Gör90] berechnet wurde.

Abbildung 6.6 zeigt eine Momentaufnahme einer CNCL-Simulation und Abbildung 6.7 zeigt eine analoge Momentaufnahme für eine OPNET²-Simulation des FDDI-Modells.

Das Modell benutzt die in Tabelle 6.1 angegebenen Parameter.

²OPNET [MIL97] ist ein kommerzielles Simulationswerkzeug der Firma MIL3, Washington, USA, welches zu Vergleichszwecken zusätzlich zur Klassenbibliothek CNCL im Simulationspraktikum des Lehrstuhls eingesetzt wird.

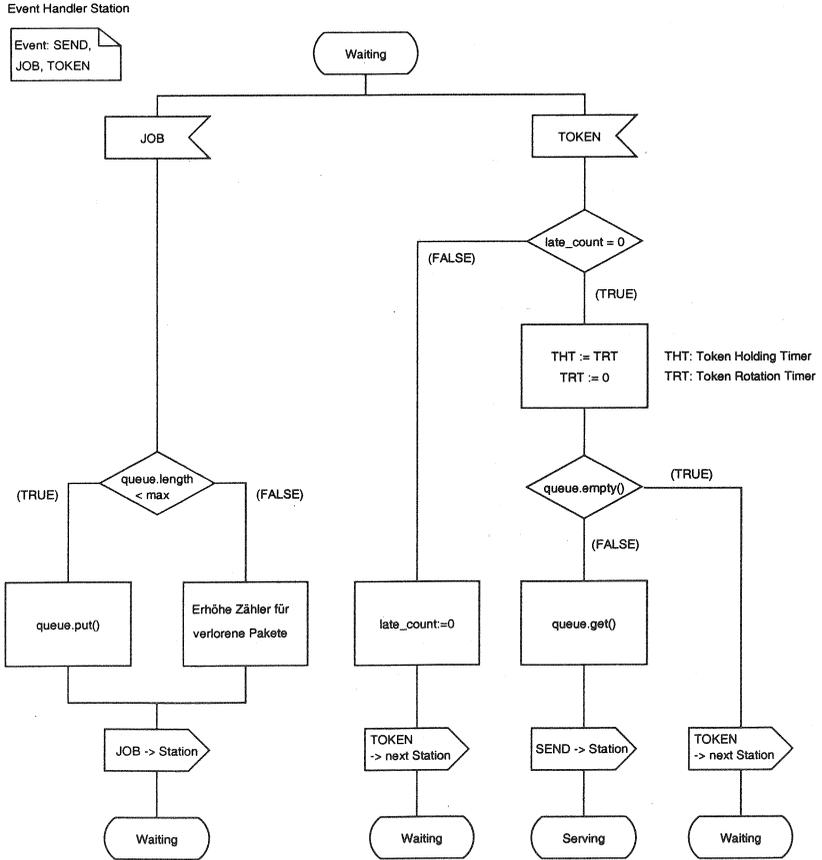


FIFO exh.:	Station: FIFO Netz: <i>exhaustive service</i>
FIFO non-exh.:	Station: FIFO Netz: <i>non-exhaustive service</i> , Timed Token Protocol
SRPN:	Station: <i>Shortest Remaining Packet Number</i> Netz: <i>non-exhaustive service</i> , Timed Token Protocol
SRPN-PRIO:	Station: <i>Shortest Remaining Packet Number</i> Netz: Prioritäten nach Anzahl der Restpakete
SRPT/RR _{opt} :	SRPT mit Zeitscheiben, theor. Vergleich nach [Gör90]

Abbildung 6.3: FDDI-Simulation: Mittlere Transferzeit $E[\tau_d]$ verschiedener Strategien

Tabelle 6.1: Modellparameter

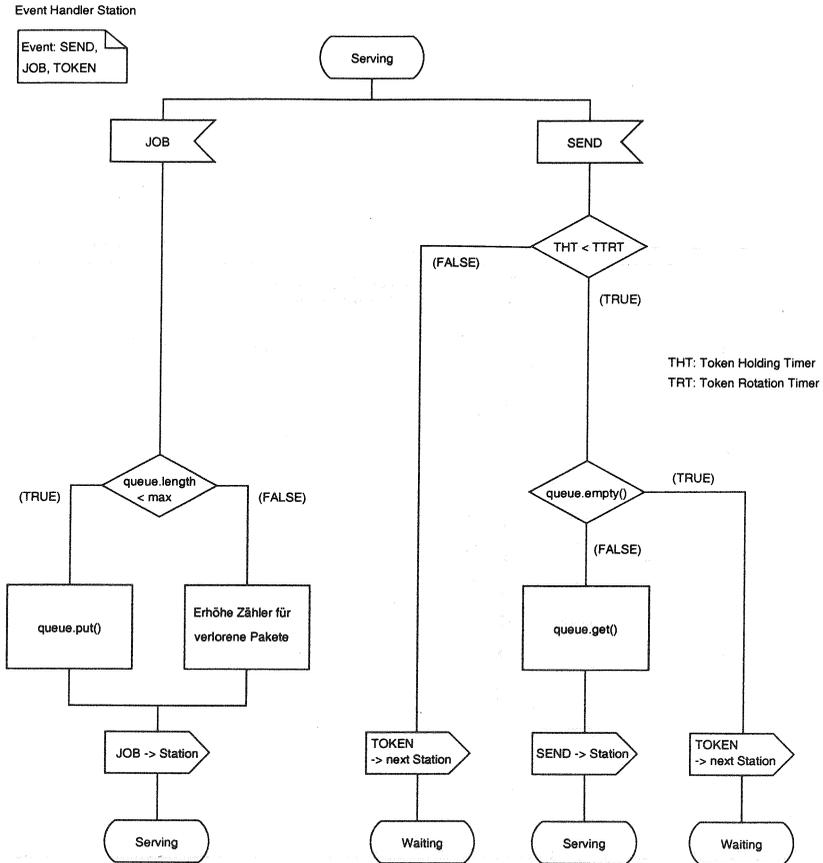
Übertragungsrate	v	100 Mbit/s
Ausbreitungsgeschwindigkeit	c_M	200 000 km/s
Stationslatenz	L_S	60 bit
Tokenlänge	L_T	88 bit



Formale Spezifikation des Zustands *Waiting* in einer SDL-ähnlichen Notation

Die Legende der Symbole wird in Abbildung 2.8 beschrieben.

Abbildung 6.4: FDDI: Ablaufbeschreibung der Station im Zustand *Waiting*



Formale Spezifikation des Zustands *Serving* in einer SDL-ähnlichen Notation

Die Legende der Symbole wird in Abbildung 2.8 beschrieben.

Abbildung 6.5: FDDI: Ablaufbeschreibung der Station im Zustand *Serving*

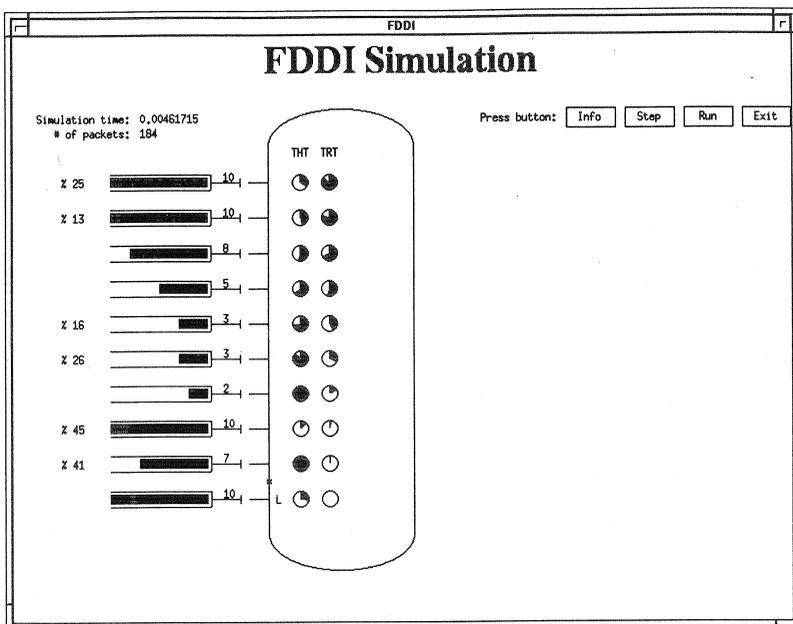


Abbildung 6.6: Momentaufnahme einer FDDI-Animation (CNCL)

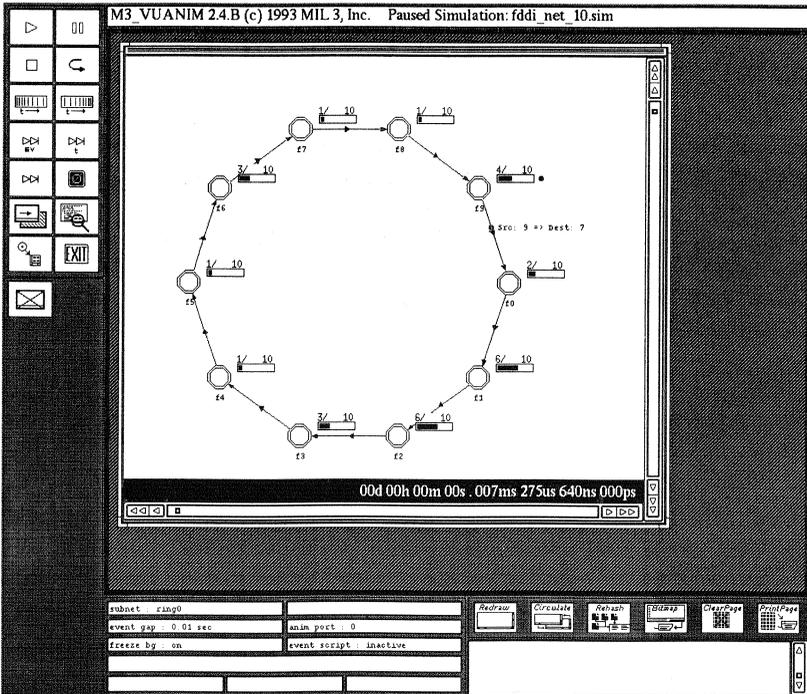


Abbildung 6.7: Momentaufnahme einer FDDI-Animation (OPNET)

6.2 Intelligente Netze (IN)

Das sog. *Intelligente Netz* (IN – *Intelligent Network*) als neues weltweites Konzept spezifiziert die Funktionen, Schnittstellen und dienstunabhängigen Basisbausteine (SIB – *Service Independent Building Block*), die für die Implementierung von Mehrwertdiensten in modernen digitalen Telekommunikationsnetzen erforderlich sind. Ein Beispiel ist der standardisierte UPT-Dienst (UPT – *Universal Personal Telecommunication*), ein Dienst für die Teilnehmer-Mobilität in Festnetzen [NA-92a, NA-92b].

In diesem Abschnitt wird die Implementierung und Bewertung eines erweiterten UPT-Dienstes [GGK97] beschrieben, der eine Erreichbarkeitssteuerung für ISDN anbietet und die Integration einer Vielzahl von Kommunikationsdiensten erlaubt. Die Basis für diesen Dienst bilden das Intelligente Netz und das Signalisiersystem Nr. 7 (SS.7) [MS90]. Der erweiterte UPT-Dienst erlaubt dem Benutzer eine Steuerung seiner Erreichbarkeit in Abhängigkeit von seiner aktuellen Situation (z. B. normaler Arbeitstag, Besprechung, Dienstreise, Urlaub). In [WK90, Baf95] wird eine Grundlage für die Leistungsanalyse neuer Dienste (z. B. ISDN, IN, Mobilkommunikation) basierend auf dem Signalisiersystem Nr. 7 beschrieben. Der vorliegende Abschnitt beschreibt die Ergebnisse einer umfangreichen Projektstudie zur Leistungsanalyse des erweiterten UPT-Dienstes. Es werden Modelle des Kommunikationsverhaltens des Benutzers, der Systemkomponenten und der Dienste entwickelt, und daraus Warteschlangenmodelle für die Verkehrsanalyse abgeleitet, die approximativ analytisch und mit Hilfe der stochastischen Simulationstechnik untersucht werden. Zielgrößen der Untersuchung sind die Verbindungsaufbaudauer der neuen Dienste sowie die Belastung des Signalisiernetzes. Die hier vorgestellten Arbeiten sind Ergebnisse der Mobilise-Projektgruppe³ des europäischen RACE⁴ II Forschungsprogramms. Verschiedene Szenarien für die Verteilung der Funktionalität des Intelligenten Netzes werden untersucht.

Zwei Prototypen des erweiterten UPT-Dienstes werden beschrieben: eine Implementierung im Netz und eine Lösung am Netzrand, d. h. außerhalb des Netzes. Beide Prototypen wurden auf verschiedenen Konferenzen und Messen vorgestellt und werden teilweise inzwischen als Produkte verschiedener Firmen angeboten.

Schätzt man die Zahl der potentiellen Benutzer dieser neuen Dienste ab, so fällt auf, daß die zusätzliche Belastung des Signalisiernetzes und seiner Knoten dramatisch sein kann, wenn die Implementierungsalternativen nicht sorgfältig ausgewählt werden. Bei der hier vorgeschlagenen Variante wird ein Numerierungsschema mit der heute häufig geforderten Eigenschaft gewählt, daß die UPT-Nummer eine wirklich

³Die Projektgruppe – sowie innerhalb des Gesamtprojekts die *Demonstrator Core Task* – wurde von der Autorin von 1992-1995 geleitet.

⁴Research and technology development in Advanced Communications technologies in Europe

persönliche, vom Netzbetreiber unabhängige Nummer ist.

Die Untersuchungen basieren auf einer realistischen Mischung von lokalen, regionalen, nationalen und internationalen Rufen. Jeder Ruftyp wird getrennt bewertet und außerdem wird der Gesamtmittelwert berechnet. Zusätzlich wird angenommen, daß

- die Mobilität eines Benutzers mit der Rufverteilung korrespondiert,
- ungefähr die Hälfte der Benutzer sich im lokalen Bereich aufhalten,
- während ungefähr ein Viertel im Regionalbereich,
- ein Fünftel im nationalen Bereich
- und ein geringer Prozentsatz der Benutzer international unterwegs sind.

Diese Werte treffen näherungsweise für das Netz der Deutschen Telekom AG zu, welches als Beispielanwendung für diese Untersuchungen gewählt wurde.

Im folgenden Abschnitt wird der erweiterte UPT-Dienst beschrieben. Dem Abschnitt über IN-Bewertungsaspekte und Kenngrößen folgt ein Abschnitt über das Verkehrsmodell für die analytischen bzw. simulativen Untersuchungen. Die Hauptergebnisse⁵ sind in Abschnitt 6.2.4 zusammengestellt. Vor der Zusammenfassung werden die Prototypimplementierungen kurz beschrieben.

6.2.1 Erweiterte UPT-Dienste

UPT ermöglicht persönliche Mobilität für Telekommunikationsdienste. Jeder UPT-Teilnehmer kann Dienste auf der Basis einer persönlichen netztransparenten UPT-Nummer an jedem Endgerät netzübergreifend initiieren oder empfangen. Im Mobilise Projekt wurde ein erweiterter UPT-Dienst unter dem Namen PSCS (PSCS – *Personal Services Communication Space*) definiert und als Demonstrationssystem implementiert und vorgeführt, siehe [GKG⁺95, KG95, PFJ95]. Beispiele des erweiterten UPT-Dienstes sind: Rufaufzeichnung, Interaktionen mit Hilfe von DTMF-Erkennung (DTMF – *Dual Tone Multiple Frequency*), Rufumlenkung und Aktivierung von Benachrichtigungsdiensten. Zusätzlich wird das sog. *Kommunikationsmanagement* eingeführt, welches eine vordefinierte Erreichbarkeitssteuerung für eingehende Verbindungswünsche zur Verfügung stellt. Der Benutzer definiert ein sog. Kommunikationsfilter (*Routing Scheme*), um die Weiterleitung der Dienste, wie z. B. Telefon, Telefax oder elektronische Post festzulegen. Die eingehenden Rufe können abhängig von der Nummer des Anrufers, der angerufenen UPT-Nummer, der Tageszeit und dem ausgewählten Dienst gesteuert werden.

Abbildung 6.8 zeigt die verschiedenen Möglichkeiten des Kommunikationsmanagements: Eingehende Kommunikationswünsche können ignoriert werden (z. B. anonyme

⁵Die hier vorgestellten Ergebnisse sind Auszüge aus den Dissertationen von S. KLEIER [Kle96] und M. GUNTERMANN [Gun96], die von der Autorin mitbetreut wurden.

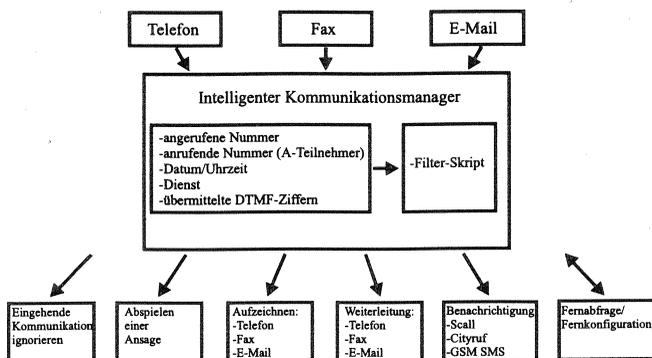


Abbildung 6.8: PSCS-Funktionen bzgl. eingehender Kommunikationsanforderungen

Anrufe), sog. *Ansagen* erlauben eine allgemeine oder gezielte Unterrichtung (CLIP – *Calling Line Identification Presentation*) von Anrufern oder Absendern elektronischer Post. Nachrichten des Anrufers können aufgezeichnet werden. Der Benutzer kann das Kommunikationsfilter so konfigurieren, daß wichtige Kommunikationswünsche direkt an ihn weitergeleitet werden, z. B. auf ein beliebiges von ihm benanntes Telefon, typisch das Mobilfunktelefon des Benutzers. Für den Telefax-Dienst kann die Nummer eines anderen Telefax-Gerätes, z. B. in einem Hotel, angegeben werden. Die Benutzung von Prototypen des Kommunikationsmanagers hat gezeigt, daß die Benachrichtigung des Benutzers über ausgewählte Aktionen des Kommunikationsmanagers über einen Benachrichtigungsdienst (Scall, Cityruf, GSM-Kurzmitteilungsdienst) eine der wichtigsten Funktionen ist.

Der Benutzer kann seinen Kommunikationsmanager fernkonfigurieren. Ruft er den Kommunikationsmanager z. B. über sein GSM-Telefon mit Unterstützung der Nummernanzeige (CLIP) an, so wird er über die rufende Nummer identifiziert und sofort ohne weitere Authentifikation mit einem Interaktionsmenü verbunden. Erfolgt der Anruf über ein anderes Telefon, so muß er sich über eine spezielle DTMF-Sequenz identifizieren. Abbildung 6.9 zeigt eine graphische Oberfläche an einem PC oder Arbeitsplatzrechner für den Zugriff des Benutzers auf seinen Kommunikationsmanager vor Ort oder über ein Netz. Der Benutzer kann die aktuellen Einstellungen seines Kommunikationsfilters durch “*drag and drop*” mit Hilfe der verschiedenen Symbole verändern. Eintreffende Telefax-Dokumente oder elektronische Briefe können am Bildschirm angezeigt und ggf. an ein beliebig wählbares Telefax-Gerät weitergeleitet werden. Die verschiedenen Nummern für die Weiterleitung von Anrufen, Telefax-Dokumenten oder elektronischen Briefen können verändert bzw. aktiviert werden.

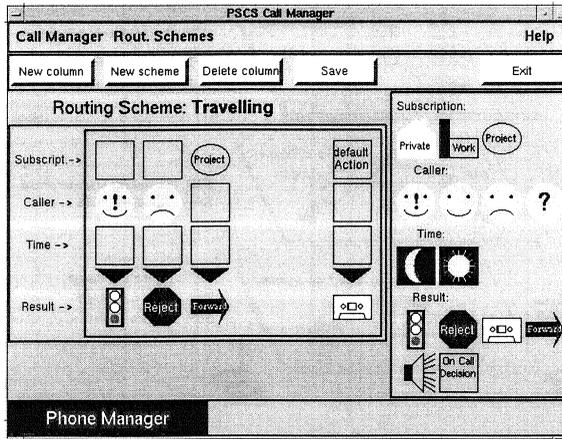


Abbildung 6.9: Graphische Benutzeroberfläche der Erreichbarkeitssteuerung

6.2.2 Bewertungsaspekte und Einflußgrößen

Das von der ITU-T (ITU-T – *International Telecommunication Union Telecommunications standardization sector*) standardisierte Konzept der Intelligenten Netze (IN) wurde spezifiziert, um den Prozeß der Einführung neuer Dienste in Telekommunikationsnetzen und auch die Modifikation und Anpassung dieser Dienste an Benutzerbedürfnisse zu standardisieren und damit zu beschleunigen.

Abbildung 6.10 zeigt die Struktur eines modernen Telekommunikationsnetzes mit seinen IN-Komponenten. Die Hauptidee des IN besteht darin, die Dienstlogik in sog. Dienststeuerungsfunktionen (SCF – *Service Control Function*) zusammenzufassen, die mit Hilfe leistungsfähiger Rechner realisiert werden. Die Dienststeuerungsfunktionen greifen auf Dienst- und Benutzerdaten über sog. Dienstdatenfunktionen (SDF – *Service Data Function*) zu, die einem Datenbankzugriff entsprechen. Diese Funktionen können separat oder physikalisch zusammen mit den Dienststeuerungsfunktionen realisiert werden.

Zur Interaktion mit dem Benutzer werden sog. Betriebsmittelfunktionen (SRF – *Specialized Resource Function*) zur Verfügung gestellt, um Ansagen abzuspielen oder DTMF-Sequenzen zu erkennen. Die einzige dienstspezifische Funktion, die im Übertragungsnetz verbleibt, ist die sog. Dienstvermittlungsfunktion (SSF – *Service Switching Function*), die die Dienstanforderungen erkennt. Über das Signalisierungsnetz kommuniziert die Dienstvermittlungsfunktion mit der Dienststeuerungsfunktion unter Verwendung des INAP-Protokolls (INAP – *Intelligent Network Application Protocol*), siehe [Thö94, Jab92, Rec93a].

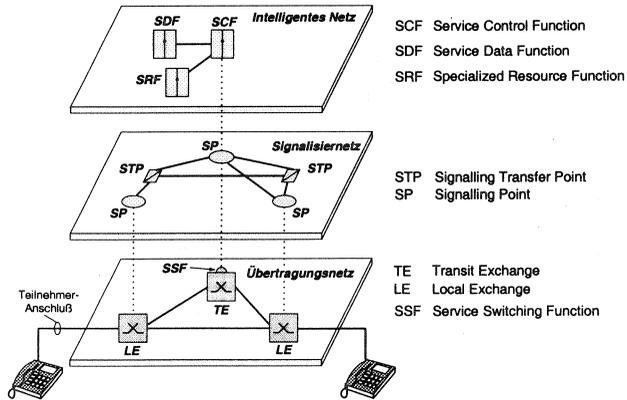


Abbildung 6.10: Architektur eines modernen Telekommunikationsnetzes

Das Netz und seine Komponenten werden modelliert, um die Dienstgüte des erweiterten UPT-Dienstes und die Belastung des Signisieretzes anhand objektiv meßbarer Leistungsgrößen vorherzusagen. Die Verbindungsaufbaudauer ist ein Hauptkriterium, welches beim Entwurf der neuen Dienste berücksichtigt werden muß. Sie wird hier für ISDN-Rufe definiert als Zeitspanne von der letzten Zifferneingabe des Anrufers bis zur zugehörigen Antwort, d. h. bis zum Empfang des Ruftones bzw. einer Ansage oder Besetztnachricht beim Anrufer.

Die Einführung des ISDN im Fernbereich hat dazu geführt, daß Benutzer sehr kurze Verbindungsaufbauzeiten erwarten. Neue Dienstplattformen wie das Intelligente Netz werden die Anforderungen an neue Dienste erfüllen, aber die Verbindungsaufbauzeiten durch zusätzliche Bediendauern und Übertragungszeiten verlängern. Untersuchungen haben gezeigt, daß Benutzer eine Verlängerung der Verbindungsaufbaudauer um etwa 2.5 s im Mittel bei normalen Rufen akzeptieren, siehe [YD94]. Dieser Wert kann als Ausgangswert für die Planung neuer Architekturen und Dienste dienen. Zusätzlich ist zu beachten, daß der Wert vom Typ des Anrufs abhängig ist, z. B. ob es sich um einen lokalen oder internationalen Ruf oder einen speziellen Dienst handelt.

6.2.3 Verkehrsmodell

Das *Verkehrsmodell* beschreibt den Verkehrsfluß in einem Telekommunikationsnetz. Abbildung 6.11 gibt eine Übersicht des Verkehrsmodells, das zur Bewertung neuer Telekommunikationsdienste benutzt wurde. Das Modell umfaßt:

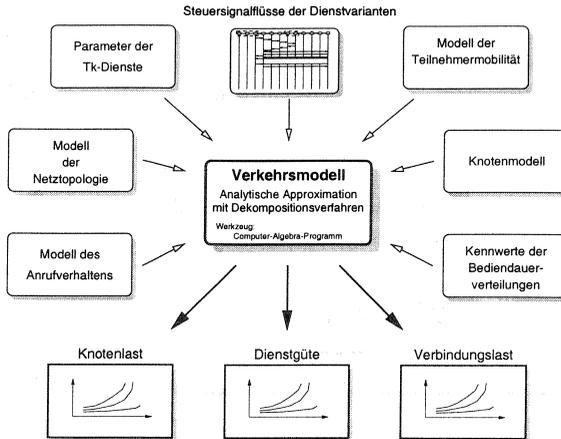


Abbildung 6.11: Überblick Analysemethode

- das **Modell des Anrufverhaltens**, welches die Häufigkeit und den Zeitabstand zwischen Rufen des Benutzers beschreibt.
- das **Modell der Teilnehmermobilität**, welches den Einfluß des Aufenthaltsortes des Teilnehmers auf die Auswahl der Dienstvarianten beschreibt.
- das **Modell der Netztopologie**, welches die Anordnung der Knoten beschreibt und den Aufbau des Netzes mit seinen physikalischen Eigenschaften definiert. Das Netzmodell basiert auf dem hierarchischen Signalisiernetz der Deutschen Telekom AG. Dieses Modell enthält die physikalische Verteilung der verschiedenen IN-Knoten, Datenbanken und Zwischenspeicher (*data cache*) im Signalisiernetz.
- **Steuersignalflüsse der Dienstvarianten**, welche die verschiedenen Varianten der Dienste beschreiben, insgesamt ca. 64 verschiedene Steuersignalflüsse.
- die **Parameter der Telekommunikationsdienste**, die die Signale, die eine Dienstvariante realisieren, beschreiben.
- das **Knotenmodell**, welches die Struktur eines IN- oder Vermittlungsknotens beschreibt. Die IN-Funktionalität kann innerhalb des Netzes oder in einem Endgerät oder einer Nebenstellenanlage außerhalb des Netzes realisiert werden.
- die **Kennwerte der Bediendauer-Verteilungen**, dazu gehören z. B. die mittlere Bediendauer und der Variationskoeffizient der Bediendauer.

Die Auswertung umfasst:

- die **Knotenlast**, womit die Auslastung jedes Knotens bestimmbar und die Berechnung der benötigten Kapazität zur Realisierung der verschiedenen Dienste möglich wird.
- die **Dienstgüte** als Hauptergebnis der simulativen und analytischen Untersuchung, die die Mittelwerte und Verteilungen der Verbindungsaufbaudauer der verschiedenen Netzarchitekturen auswerten.
- die **Verbindungslast**, die die Berechnung der benötigten Kapazität zwischen den Signalisierknoten erlaubt.

Vermittlungsknoten: Das Modell der Vermittlungsknoten wird von einem typischen modernen Multiprozessor-Vermittlungsknoten abgeleitet, siehe Abbildung 6.12 und [AE91].

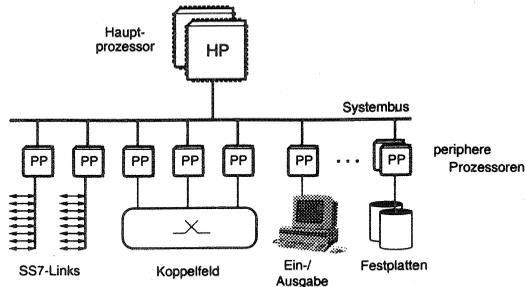


Abbildung 6.12: Schematische Vermittlungsknoten-Architektur

Einer oder mehrere *Hauptprozessoren* (HP) verarbeiten die Dienstlogik, siehe Abbildung 6.12. In der SS.7-Protokollhierarchie entspricht dies der Bearbeitung von Anwendungen der Ebene 4, siehe [MS90].

Die *peripheren Prozessoren* (PP) entlasten den Hauptprozessor von Routinearbeiten und bilden die Schnittstelle zu den Koppelfeldern, den Eingabe- und Ausgabeeinheiten, den Festplatten und zum Signalisiersystem (SS.7). Bezüglich des Signalisiersystems findet eine Vorverarbeitung in den peripheren Prozessoren statt. In der SS.7-Protokollhierarchie entspricht dies der Bearbeitung durch die MTP (MTP – *Message Transfer Part*) Ebene 1-3, siehe [Rec93b].

Im folgenden werden die wichtigsten Annahmen zusammengefaßt:

- Es werden nur Rufe zwischen zwei Teilnehmern modelliert. Telefonkonferenzen werden nicht betrachtet.
- Es wird von folgender Verteilung der Rufe ausgegangen: 54 % lokal, 26 % regional, 17.5 % national und 2.5 % international. Außerdem wurden leichte Variationen dieser Verteilung benutzt.
- Die Last im realen Netz ist nahezu symmetrisch. Das vereinfachte analytische Modell ist eine gute Approximation für diesen Fall.
- Die Rate der eingehenden und ausgehenden Rufe ist gleich.
- Allen Teilnehmern wird die gleiche Anrufrate zugeordnet.
- Die berechneten oder simulierten Signalisiertransferzeiten müssen zumindest die Vorgaben der ITU erfüllen, die in Abbildung 6.13 als Kurven dargestellt sind. Die Last wird hier auf die Hauptverkehrsstunde bezogen, die mit 100 % bezeichnet wird.
- Das untersuchte Netz ist für den ISDN-Verkehr der Hauptverkehrsstunde ausreichend dimensioniert.
- Es gibt keine Blockierungen im Netz.
- Es wird von idealen Signalisierverbindungen ausgegangen, d. h. Bitfehler werden ignoriert.
- Die Laufzeitverzögerung ist 3 ms für alle Signalisierverbindungen.
- Das gerufene Endgerät hat eine konstante Antwortverzögerung von 200 ms.
- Es wird von $N = 39.2$ Millionen Teilnehmern ausgegangen, siehe [AG94].
- Erfolgreiche Rufe erzeugen eine höhere Signalisierlast als nicht erfolgreiche Rufe, z. B. Besetzzeichen. Im analytischen Modell werden unterbrochene Rufe wie erfolgreiche Rufe behandelt, d. h. es wird die Maximallast betrachtet.
- Die simulativen Modelle enthalten eine PUI-SDP-Datenbank (PUI – *Personal User Identification*, SDP – *Service Data Point*) sowie einen Zwischenspeicher (PUI-SDP cache), die die Zuordnung der UPT-Nummer zur PUI der Teilnehmer des erweiterten UPT-Dienstes enthalten. Es wird angenommen, daß der Anteil der UPT-Rufe, die eine Abfrage in der PUI-SDP-Datenbank benötigen dem Anteil an internationalen Rufen entsprechen, d. h. die Anteile sind 97.5 % bzw. 2.5 %, siehe [Lin96].

Simulationsmodell: Das Modell des hierarchischen Netzes, welches den simulativen Untersuchungen zugrundeliegt, ist in Abbildung 6.14 dargestellt. Die Vermittlungsstellen der obersten Ebene sind vollvermascht. Eine lokale Vermittlungsstelle steht stellvertretend für alle lokalen Vermittlungsstellen dieses Bereichs.

In jeder Vermittlungsstelle werden M/G/S-Prozessormodelle mit Rückführungen entsprechend der Round-Robin-Bedienstrategie eingesetzt, siehe Abbildung 6.14. Die ankommenden Signalisiernachrichten werden gleichmäßig auf die Warteschlangen verteilt.

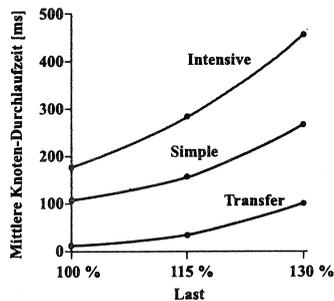


Abbildung 6.13: ITU-Transferzeiten: Maximalwerte der SS.7-Nachrichten [Rec93c]

Abhängig vom Typ der Signalisiernachricht – *Transfer* ($x = 1$), *Simple* ($x = 3$) oder *Intensive* ($x = 5$) – müssen die Nachrichten x -mal das gesamte System durchlaufen. Dies entspricht in etwa dem Verhältnis der Transferzeiten für die unterschiedlichen Datentypen aus dem ITU-Standard von $35\text{ ms} : 110\text{ ms} : 180\text{ ms}$, siehe Abbildung 6.13.

Dieses Prozessormodell erlaubt die Approximation der im Standard vorgegebenen Werte für die Knotendurchlaufzeiten, wobei die Nachrichten von einem Prozessor behandelt werden und so auch die gegenseitige Beeinflussung modelliert wird, siehe [FKG95, Kle96].

Analytisches Modell: Neben dem Simulationsmodell wurde ein analytisch berechenbares Modell des gesamten Netzes entwickelt und untersucht, um größere Parameterstudien zu ermöglichen.

Es basiert im wesentlichen auf dem geplanten Netz der Deutschen Telekom AG und benutzt die dafür diskutierten Netztopologien. Die gegenseitige Validierung der beiden Modelle wurde erfolgreich für ein spezielles Szenario durchgeführt.

Modellierung der Netzknoten: Die Netzknoten können in zwei Klassen eingeteilt werden: die Vermittlungsknoten und die IN-Knoten. Alle Knoten werden durch das Signalisiernetz (SS.7) verbunden. Die Bedeutung der Variablen in den folgenden Abschnitten ist wie folgt: k Knotennummer, s SS.7 Signaltyp, x Anteil der Teilnehmer des erweiterten UPT-Dienstes, τ_{MP} MTP Bediendauer (siehe *processing handling time*, τ_{ph} , Q.706), τ_{MS} MTP Sendeverzögerung (siehe *outgoing link delay*, T_{od} , Q.706), τ_{An} Verzögerung der MTP Ebene 4 (siehe *handling time* T_{ht} , Q.706), $\tau_{w,k,x}$ Wartezeit eines SS.7-Signals im Knoten k .

IN Knoten: Die IN-Dienststeuerknoten (SCP – *Service Control Point*) und Dienstdatenknoten (SDP – *Service Data Point*) werden durch das M/G/1-FIFO Wartesystem modelliert, siehe Abbildung 6.15.

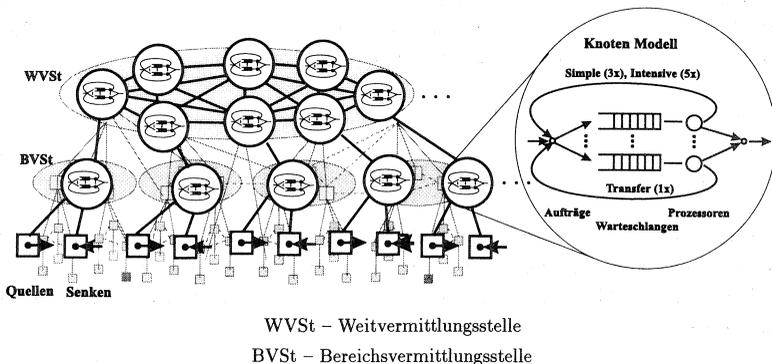


Abbildung 6.14: Simulatives Modell des hierarchischen Telekommunikationsnetzes

Die Ankünfte von Signalsnachrichten werden als Poisson-Prozeß modelliert. Die Bediendauer jedes Signaltyps wird durch das erste und zweite Moment der Bediendauerverteilung beschrieben. Die mittlere Bediendauer $E[\tau_{b,k,x}]$ entsprechend der Mischung der Signaltypen wird wie folgt mit Hilfe der Mischwahrscheinlichkeiten $p_{k,s,x}$ berechnet:

$$E[\tau_{b,k,x}] = \sum_{s=1}^S p_{k,s,x} \cdot E[\tau_{b,k,s}] \tag{6.5}$$

Das zweite Moment der Bediendauerverteilung $\tau_{b,k,x}^{(2)}$ ergibt sich zu:

$$\tau_{b,k,x}^{(2)} = \sum_{s=1}^S p_{k,s,x} \cdot \tau_{b,k,s}^{(2)} \tag{6.6}$$

Damit berechnet sich der Variationskoeffizient $c_{b,k,x}$ zu:

$$c_{b,k,x} = \sqrt{\frac{\tau_{b,k,x}^{(2)}}{E[\tau_{b,k,x}]^2} - 1} \tag{6.7}$$

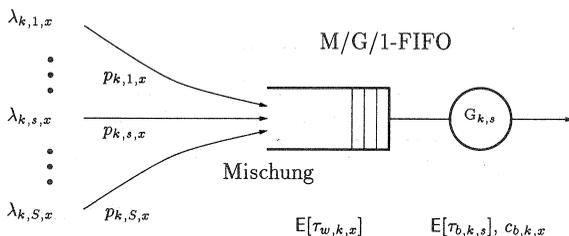


Abbildung 6.15: Analytisches Modell der IN-Knoten

Mit dem Variationskoeffizienten $c_{b,k,x}$, dem Mittelwert der Bediendauer $E[\tau_{b,k,s}]$ des Signaltyps s sowie der Auslastung $\eta_{k,x}$ und dem Mittelwert der Bediendauer $E[\tau_{b,k,x}]$ ergibt sich die *mittlere Verweildauer je Signaltyp* $E[\tau_{d,k,s,x}]$ zu:

$$\begin{aligned} E[\tau_{d,k,s,x}] &= E[\tau_{w,k,x}] + E[\tau_{b,k,s}] \\ &= \frac{E[\tau_{b,k,x}] \cdot \eta_{k,x} \cdot (1 + c_{b,k,x}^2)}{2 \cdot (1 - \eta_{k,x})} + E[\tau_{b,k,s}] \end{aligned} \quad (6.8)$$

Vermittlungsknoten: Für die Modellierung dieser Architektur wird das in Abbildung 6.16 gezeigte Wartnetz benutzt. Das Wartnetz besteht aus hintereinandergeschalteten Warteschlangen, die die peripheren Prozessoren und den Zentralprozessor modellieren.

Berechnung der Verzögerungen: Das Knotenmodell in Abbildung 6.16 berücksichtigt die unterschiedlichen Bediendauern der Steuersignale in den verschiedenen Ebenen des SS.7-Architekturmodells. Alle Signale werden durch die peripheren Prozessoren in den MTP-Ebenen 1-3 vorverarbeitet. Dabei muß zum einen die mittlere Prozessorverzögerung $E[\tau_{MP,k,s,x}]$ und zum anderen die MTP-Sendeverzögerung $E[\tau_{MS,k,s}]$ berücksichtigt werden. $\tau_{MS,k,s}$ hängt von der Leistungsfähigkeit des Knotens, vgl. Tabelle in [Rec93b], der Länge l_m der Protokolldateneinheit vom Typ m und der Auslastung $Y_{k,l}$ sowie der Übertragungsgeschwindigkeit $B_{k,l}$ des Kanals ab und berechnet sich wie folgt:

$$\tau_{MS,k,s} = \frac{l_m}{B_{k,l}} \cdot \left(1 + \frac{3}{l_m} + \frac{Y_{k,l}}{1 - Y_{k,l}}\right) \quad (6.9)$$

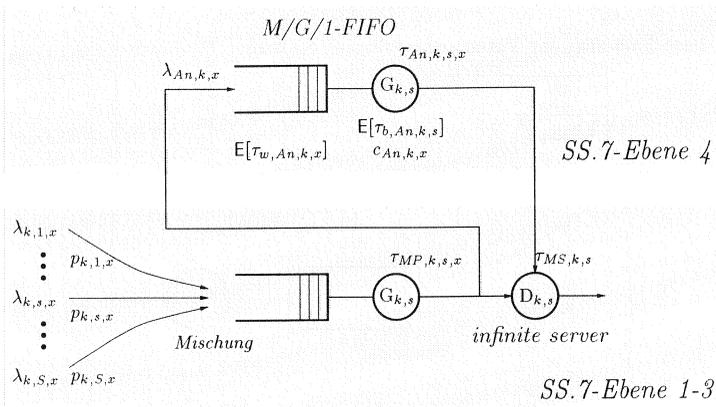


Abbildung 6.16: Analytisches Wartnetzmodell der Vermittlungsknoten

Die MTP-Prozessorverzögerung $\tau_{MP,k,s,x}$ ist abhängig von der relativen Auslastung η_x/η_0 und der Länge l_m der MTP-Protokollateneinheit vom Typ m :

$$\tau_{MP,k,s,x} = 115.67 \cdot \frac{\eta_x}{\eta_0} - 100.767 + (0.83 \cdot \frac{\eta_x}{\eta_0} - 0.693) \cdot l_m \quad (6.10)$$

Das Modell der SS.7-Ebenen 1-3 wurden der ITU-Empfehlung Q.706 [Rec93b] entnommen.

Alle Steuersignale, die eine Verarbeitung in der SS.7-Ebene 4 erfordern, müssen durch den Zentralprozessor verarbeitet werden. Dabei hängt die Verweildauer $\tau_{An,k,s,x}$ vom Signaltyp ab. Die Berechnung der mittleren Anwenderteilverzögerung $E[\tau_{An,k,s,x}]$ entspricht der Berechnung der mittleren Verweildauer $E[\tau_{d,k,s,x}]$ beim elementaren Knotenmodell, vgl. Gleichung (6.8), wobei dieser Verarbeitungsschritt für die Transfersignale (STP-Nachrichten) entfällt. Damit ergibt sich die Knotendurchgangsdauer $E[\tau_{d,k,s,x}]$ beim Wartnetzmodell der Vermittlungsknoten aus der Summe der verschiedenen Verweildauern zu:

$$E[\tau_{d,k,s,x}] = \begin{cases} E[\tau_{MP,k,s,x}] + E[\tau_{MS,k,s}] & : \text{ Transfersignale} \\ E[\tau_{MP,k,s,x}] + E[\tau_{An,k,s,x}] + E[\tau_{MS,k,s}] & : \text{ sonst} \end{cases} \quad (6.11)$$

Die Teilmodelle (Knotenmodelle) werden getrennt analysiert, und es werden Poisson-Ankunftsprozesse angenommen. Der Ausgangsprozess eines M/G/1-FIFO-Knotens ist im allgemeinen kein Poisson-Prozeß und auch kein Erneuerungsprozeß. Daher ist die Poisson-Annahme eine Approximation, die aber zu sinnvollen Ergebnissen führt, wenn eine genügend große Anzahl von Signalen unterschiedlicher Rufe angenommen werden kann. Dies ist für ein öffentliches Telekommunikationsnetz in der Hauptverkehrsstunde zutreffend.

6.2.4 Bewertungsergebnisse

Abbildung 6.17 zeigt die verschiedenen untersuchten Szenarien. Die drei Szenarien in der oberen Reihe wurden mit den Methoden der stochastischen Simulation untersucht und die der unteren Reihe mit approximativen analytischen Methoden. Die Szenarien unterscheiden sich bezüglich der Zuordnung der IN-Funktionalität zur Realisierung der erweiterten UPT-Funktionalität. Die zentralen Szenarien stellen den Ausgangspunkt für eine mögliche UPT-Einführungsstrategie dar. In allen Fernvermittlungen (WVST – *Weitvermittlungsstellen*) wird die gesamte IN-Funktionalität installiert (SSF, SCF, SDF, SRF). In den dezentralen Szenarien wird die IN-Funktionalität im wesentlichen auf der Ebene der lokalen Vermittlungen (BVSt – *Bereichsvermittlungsstellen*) angeordnet. Die IKM-Szenarien beschreiben die Lösung am Nettrand. In den Szenarien A-Zentral und A-Dezentral sind im Unterschied zu S-Zentral und S-Dezentral die Funktionen SCF und SDF zu einem SCP zusammengefaßt. Zusätzlich wird in den simulativ untersuchten Szenarien die PUI-Datenbank eingeführt.

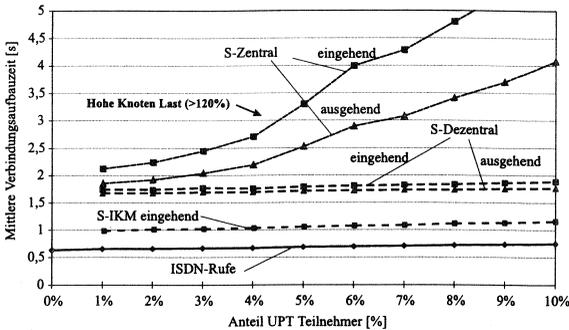


Abbildung 6.18: Simulationsergebnisse: Mittlere Verbindungsaufbauzeit

UPT-Teilnehmern ist die Last im Signalisieretz zu hoch. Kürzere Verbindungsaufbauzeiten werden erreicht durch die Verteilung der Funktionen, aber dies erfordert den IN-gemäßen Ausbau in jeder der betroffenen Vermittlungsstellen.

Die kürzesten mittleren Verbindungsaufbauzeiten werden bei den hier getroffenen Annahmen bzgl. der Benutzermobilität mit der Lösung außerhalb des Netzes erreicht, die die Weiterleitung an den Aufenthaltsort des Benutzers mit dem ISDN-Dienstmerkmal *Rufumlenkung* realisiert. In diesem Fall sind keine aufwendigen Anfragen an zentrale Datenbanken zu berücksichtigen.

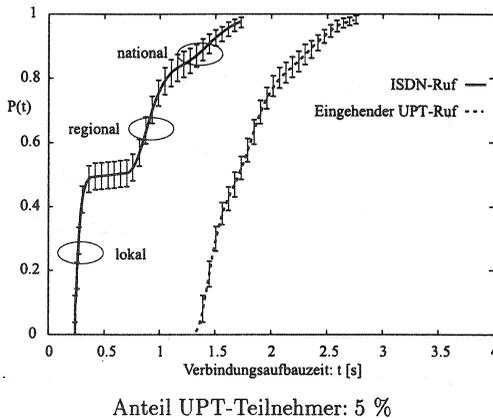


Abbildung 6.19: Simulationsergebnisse: Verteilung der Verbindungsaufbauzeiten

Abbildung 6.19 zeigt die Verteilung der Verbindungsaufbauzeit für das dezentrale Szenario (S-Dezentral) mit Konfidenzintervallen für 95 % statistische Sicherheit nach der Batch-Means-Methode. Die linke Kurve zeigt die Verteilung der normalen

ISDN-Rufe. Man sieht die drei Ruftypen (lokal, regional, national). Die Einführung eines UPT-Dienstes (5 % UPT) führt zusätzliche Verzögerungen von ungefähr einer Sekunde ein und die Benutzermobilität gleicht die Unterschiede zwischen den verschiedenen Ruftypen aus.

Analytische Bewertung der Szenarien [Gun96]: Der Hauptunterschied zwischen den drei analytisch bewertete Szenarien in Abbildung 6.17 ist die Zuordnung der Dienstvermittlungsfunktion (SSF – *Service Switching Function*). Im Szenario *A-Zentral*⁶ ist die Dienstvermittlungsfunktion in der oberen Ebene der Vermittlungshierarchie, im Szenario *A-Dezentral* in der unteren Ebene angeordnet. Im Szenario *Intelligenter Kommunikationsmanager (A-IKM)* außerhalb des Netzes ist auch die Dienstvermittlungsfunktion in einem privaten Endgerät oder einer Nebenstellenanlage implementiert.

Verbindungsaufbaudauer der Mehrwertdienste im Vergleich zu ISDN: Das linke Diagramm in Abbildung 6.20 zeigt die mittlere Verbindungsaufbaudauer für den erweiterten UPT-Dienst über dem Anteil der UPT- bzw. PSCS-Teilnehmer im Netz. Das Szenario *A-Dezentral* erreicht die kürzeste mittlere Verbindungsaufbaudauer von 1.2 s bis 1.7 s. Die längsten mittleren Verbindungsaufbaudauern ergeben sich im Szenario *A-Zentral* von 1.4 s bis 2.5 s. Das rechte Diagramm in Abbildung 6.20 zeigt den Einfluß des erweiterten UPT-Dienstes auf die Verbindungsaufbaudauer des ISDN-Dienstes, da beide Dienste das gleiche Signalisieretz benutzen. Auch hier ist das Szenario *A-Dezentral* das günstigste, d. h. es wird am wenigsten beeinflusst.

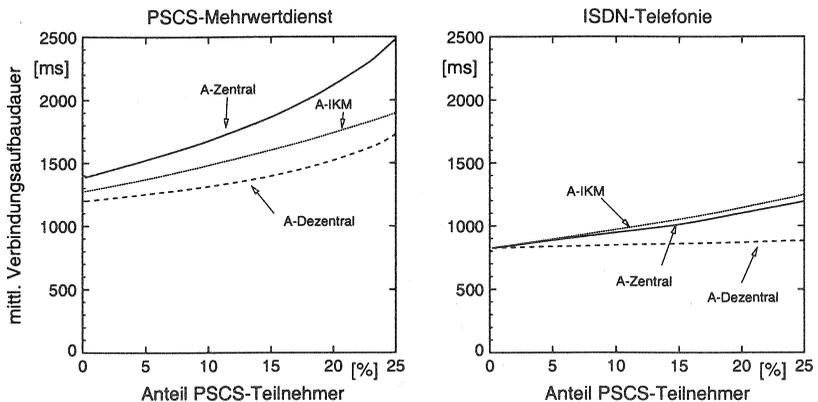
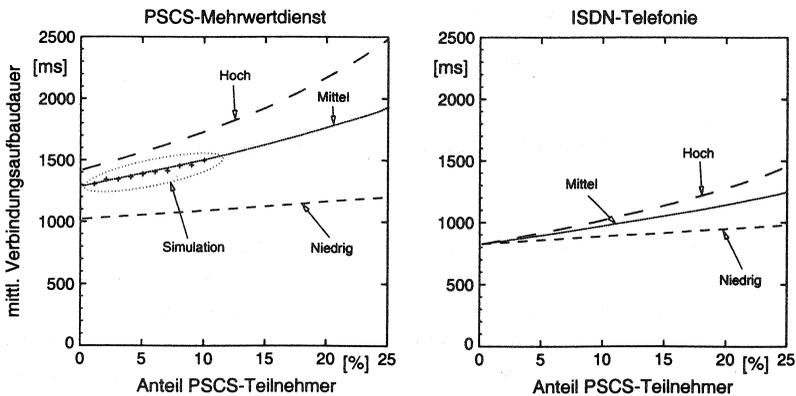


Abbildung 6.20: Analytische Ergebnisse: Mittlere PSCS-Verbindungsaufbaudauer

⁶A bedeutet analytisches Model.

Einfluß der Benutzermobilität: Höhere Benutzermobilität führt zu längeren Verbindungsaufbauzeiten. Die Untersuchungen zeigen den Einfluß der Benutzermobilität im Szenario *A-IKM* in Abbildung 6.21. Derselbe Effekt kann beim Einfluß auf den ISDN-Dienst im rechten Diagramm beobachtet werden.

Außerdem zeigt das linke Diagramm von Abbildung 6.21 einen Vergleich mit Simulationsergebnissen. Die Übereinstimmung ist gut, wenn man einen Versatz von 300 ms hinzufügt, der die unterschiedliche Modellierung der Vermittlungsfunktion (CCF) berücksichtigt, die einmal im Netz (S-IKM) und einmal im Endgerät (A-IKM) platziert ist.



Szenario A-IKM: Variation der Benutzermobilität

Abbildung 6.21: Analytische Ergebnisse: Mittlere PSCS-Verbindungsaufbauzeit

Die Untersuchungen zeigen, daß die IN-Knoten mit Bedacht dimensioniert werden müssen. Die Verbindungen im Signalisierernetz stellen in der Einführungsphase keinen Engpaß dar.

Prototypen des UPT-Dienstes im Netz und am Netzrand: Die Leistungsuntersuchungen wurden parallel zur Implementierung von Prototypen für die neuen UPT-Dienste durchgeführt. Zwei Prototypen wurden als Demonstrationssysteme implementiert: eine integrierte Lösung im Netz und eine netzunabhängige Lösung am Netzrand als Endgerätelösung.

Integrierte Lösung im Netz: Abbildung 6.22 zeigt die Struktur der netz-integrierten Lösung. Der Diensteknoten enthält den Rahmendienst, der die Dienstfunktionen des Kommunikationsmanagers – wie Authentifizierung, Registrierung und Erreichbarkeitssteuerung – zur Verfügung stellt. Als ein Beispiel eines Anwendungsdienstes steuert der Rahmendienst den Zugriff auf einen Multimedia-Maildienst (*Media-Box-Modul*).

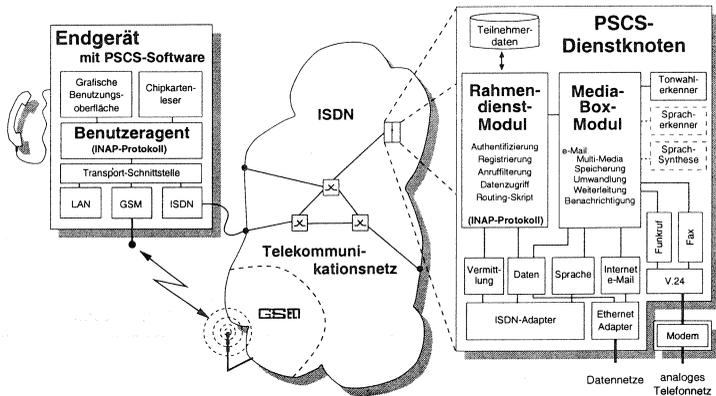


Abbildung 6.22: Struktur des Demonstrationssystems

Das Endgerät ist ein persönlicher digitaler Assistent (PDA – *Personal Digital Assistant*) mit einer graphischen Benutzeroberfläche und Verarbeitungsleistung für das Kommunikationsprotokoll des Rahmendienstes. Abbildung 6.23 zeigt einen Prototyp des Endgerätes. Auf der linken Seite ist ein GSM-Telefon als Netzanschluß und auf der rechten Seite ist der Kartenleser (*smart card reader*) für die Authentifizierung (RSA 3-Message-Authentication) zu sehen. Auf dem Bildschirm sieht man die Erreichbarkeitssteuerung, siehe Abschnitt 6.9.

Unabhängige Lösung am Netzrand: Zur Realisierung der unabhängigen Lösung am Netzrand wurde ein Anwendungsprogramm auf einem PC implementiert, das eine Vielzahl von verschiedenen Diensten anbietet und die Erreichbarkeitssteuerung unter der Regie des Benutzers realisiert.

Abbildung 6.24 zeigt den sog. *Intelligenten Kommunikationsmanager* im heutigen Umfeld der Kommunikationsmöglichkeiten. Der Kommunikationsmanager enthält als Hauptmerkmale: Anrufaufzeichnung, Interaktionen durch DTMF-Erkennung, Rufumlenkung, Benutzerbenachrichtigung, siehe Abschnitt 6.2.1.

Der Kommunikationsmanager wird über eingehende Kommunikationswünsche über den Steuerkanal (D-Kanal) des ISDN-Basisanschlusses informiert. Die Signalisierungsdaten werden mit den Einstellungen des aktuellen Kommunikationsfilters verglichen und die entsprechenden Aktionen gestartet, z. B. wird ein wichtiger Anruf an das GSM-Telefon des Benutzers weitergeleitet oder ein weniger wichtiger Anruf aufgezeichnet und der Benutzer über den GSM-Kurzmitteilungsdienst informiert. Diese Realisierung des IKM wird z. Zt. von verschiedenen Firmen in intelligenten Kommunikationsendgeräten angeboten.

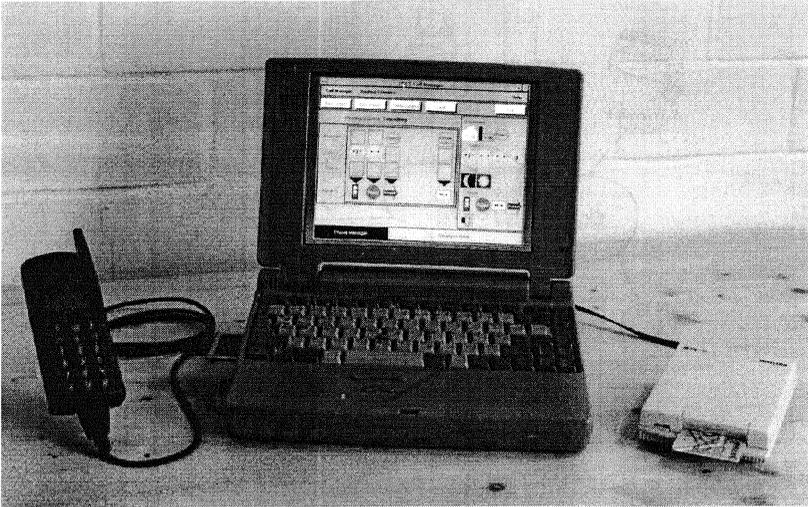


Abbildung 6.23: Endgerät des Demonstrationssystems

6.2.5 Zusammenfassende Bewertung

Die globale Teilnehmer-Mobilität in Festnetzen und Mobilfunknetzen steht am Anfang einer Entwicklung. Erste Standards basierend auf dem Intelligenten Netz für UPT-Dienste wurden von der ITU-T definiert und werden z. Zt. getestet bzw. eingeführt.

Verschiedene Szenarien und Prototypimplementierungen für erweiterte UPT-Dienste wurden untersucht. Die Vor- und Nachteile der Implementierungs-Szenarien wurden diskutiert und der Einfluß auf existierende und zukünftige Signalisiernetze in bezug auf Verbindungsaufbaudauern und Belastung der Knoten untersucht. Die Szenarien unterscheiden sich bezüglich der Plazierung der IN-Funktionalität, die zentral oder dezentral im Netz oder außerhalb des Netzes möglich ist. Die zentrale Lösung ist geeignet zur Einführung eines UPT-Dienstes mit einer kleinen Anzahl von UPT-Teilnehmern. Die dezentralen Szenarien ermöglichen kürzere Verbindungsaufbaudauern und eine größere Anzahl UPT-Teilnehmer. Der Engpaß bei diesen Lösungen ist die Verarbeitungskapazität der Netzknoten – nicht das Verbindungsnetz. Der IKM als Lösung außerhalb des Netzes ist sowohl zur Einführung eines UPT-Dienstes geeignet als auch leicht skalierbar und verwendet nur bereits verfügbare Funktionen des ISDN.

Beide Lösungen können nebeneinander existieren und weiterentwickelt werden. Ein sehr mobiler Teilnehmer mit wenig Anforderungen an ein Kommunikationsmanage-

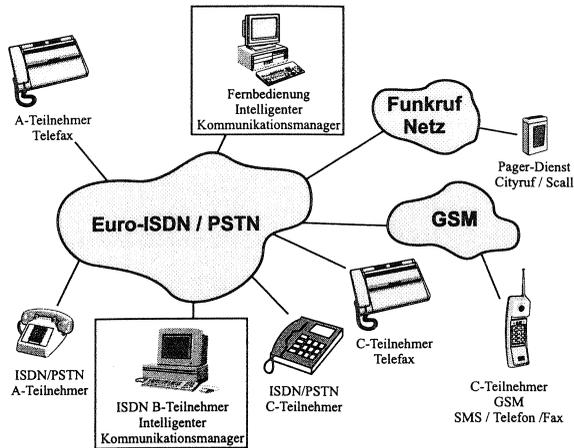


Abbildung 6.24: IKM: Intelligenter Kommunikationsmanager am Netzrand

ment wird einen vom Netz angebotenen UPT-Dienst benutzen. Firmen bzw. Benutzer, die ihre Daten und ihr Kommunikationsmanagement nicht dem Netz anvertrauen wollen oder Gebühren sparen wollen, werden die Lösung außerhalb des Netzes bevorzugen. Die endgültige Entscheidung und Verbreitung der Dienste wird nicht zuletzt auch von der Tarifpolitik der Netzbetreiber abhängen.

6.3 Mobilfunknetze

Durch die Liberalisierung des Telekommunikationsmarktes, welche in Deutschland schrittweise mit der Einführung zweier Mobiltelefon-Netze (D1 und D2) begonnen wurde, sind die Mobilfunknetze in das Zentrum des Interesses gerückt. Auch im Hinblick auf die weitere Liberalisierung des Festnetzes sind Mobilfunktechniken für den Anschluß des Endkunden über ein drahtloses Zugangsnetz (*RLL – Radio in the Local Loop*) eine technisch interessante Möglichkeit. Hier gibt es bislang wenig Erfahrungen, so daß umfangreiche Untersuchungen notwendig sind. Für einfache Szenarien sind z. T. analytische Modelle und Lösungen bekannt, aber bei komplexeren Szenarien ist die Simulationstechnik die einzig heute anwendbare Methode. Im folgenden werden die allgemeinen Anforderungen von Mobilfunknetzen kurz beschrieben und durch Beispiele aus dem Bereich Fahrzeug-Baken-Kommunikation ergänzt. Im anschließenden Abschnitt wird eine Kapazitäts-Studie des DECT-Systems vorgestellt.

6.3.1 Allgemeine Anforderungen

Aufbauend auf der allgemeinen CNCL-Bibliothek, siehe Abschnitt 2.4, wurde eine spezielle Bibliothek SIMCO3++ (*SIMCO – Simulation of Mobile COmmunication systems*) [RBGW94] entwickelt, die zusätzlich die speziell für Mobilfunksimulationen benötigten Klassen zur Verfügung stellt, z. B. Mobilitätsmodelle, Wellenausbreitungsmodelle, Kanalmodelle und Szenariobeschreibungen. Darauf bauen die Werkzeuge für verschiedene Anwendungsbereiche von Mobilfunknetzen auf, z. B. DESI (*DEct SIMulator*) [Ple96] oder SIMCO3++/DSRC (*DSRC – Dedicated Short-Range Communication system*) [Wie97]. Im folgenden werden die neuen Problemstellungen funkgestützter Netze kurz genannt.

Kommunikation zwischen:

- Mobilstation und Basisstation
(z. B. Fahrzeug-Bake, zellulare Netze: GSM, DECT, MBS)⁷
- Mobilstation und Mobilstation
(z. B. Fahrzeug-Fahrzeug)

Spezielle Problemstellungen:

- Realitätsnahe Nachbildung der Mobilität
- Modellierung der dynamischen Kanaleigenschaften
- Synchronisation der Mobilstationen
- Verträglichkeit verschiedener Funksysteme

⁷GSM – *Global System for Mobile Communication*,
DECT – *Digital Enhanced Cordless Telecommunication*
MBS – *Mobile Broadband System*

Mobilitätsmodell:

- Auflösung
(z. B. mikroskopische Nachbildung der Mobilität von Fahrzeugen im cm-Bereich)
- Umgebung
(z. B. Geographie, Bebauung, Abschattung durch andere Fahrzeuge)
- Bewegungsregeln
(z. B. Mobilitätsklassen, Fahrzeuge beachten Abstandsregeln, können überholen, abbremsen, ausweichen)
- Mobilstation
Art der Station (z. B. Handy, Notebook, Kamera)
Art der Anwendung (z. B. Text, Video, Sprache)
- Meßdaten zur Simulationssteuerung und Entwicklung von Modellen
- Parameter
(z. B. Verkehrsintensität der Mobilstationen, Anzahl der Fahrspuren, Reaktionszeiten, aktuelle Geschwindigkeit, Wunschgeschwindigkeit)

Kanalmodell:

- Antennendiagramme
(z. B. Strahlungs-Charakteristik einer Bakenantenne)
- Umgebung
(z. B. Geographie, Bebauung, Abschattung, Mehrwegeausbreitung, Dämpfungs- und Reflexionseigenschaften, Rauschquellen)
- Datenübertragungsraten und Bitfehlerraten
- Gleich- und Nachbarkanalstörungen
- Parameter
(z. B. Modulationsverfahren, Senderleistung, räumliche Verteilung der Bitfehlerraten, geometrische Eigenschaften der Hindernisse, z. B. Häuser, Wald)

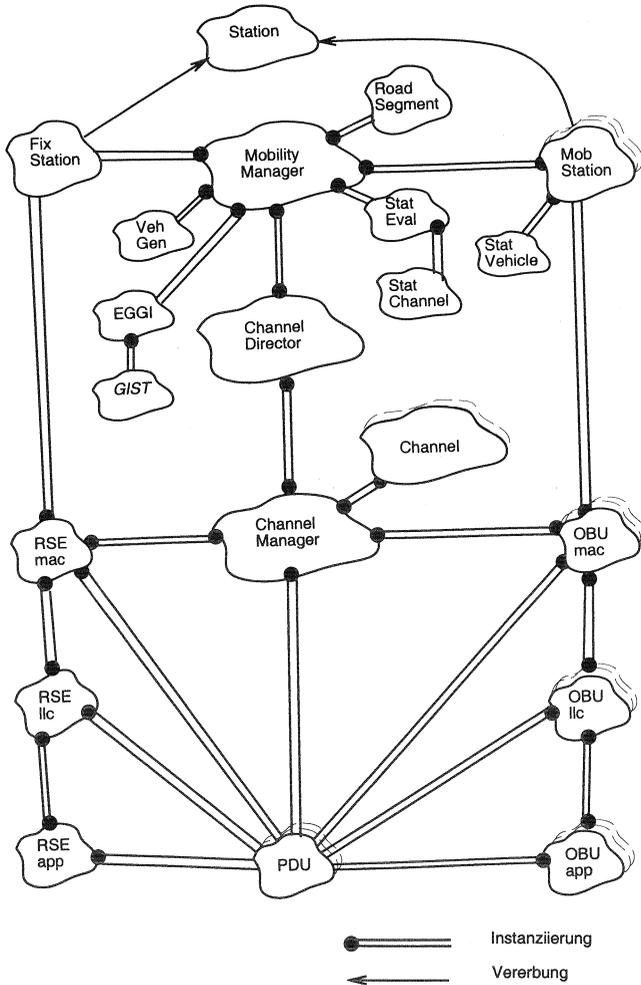
Untersuchungsziele:

- Verkehrsleistung von Protokollen
(Kanalzugriff, Transport, Routing)
- Verkehrsleistung der Fehlerbehandlung
(FEC – *Forward Error Correction*, ARQ – *Automatic Repeat Request*)
- Synchronisationsprotokolle
- Handoveralgorithmen
- Transaktionsfehlerraten
- Kanalwahl
- Kollisionsauflösungsalgorithmen

Abbildung 6.25 zeigt die Struktur der Klassenbibliothek zur Unterstützung der Simulation von Mobilfunknetzen am Beispiel der Fahrzeug-Baken-Kommunikation. Der

Mobilitätsmanager (MobilityManager) überwacht die Bewegungen der mobilen Stationen (MobStation) auf Straßenabschnitten (RoadSegment), welche die möglichen Bewegungsrichtungen der Stationen innerhalb eines Szenarios beschreiben. Der Kanal (Channel) beinhaltet eine Abbildung der Eigenschaften des Übertragungskanals, z. B. ortsbezogene Bitfehlerraten. Die Bake (RSE – *Road Side Equipment*) und die Einheit im Fahrzeug (OBU – *On Board Unit*) sind Beispiele für die Implementierung einer Kommunikationsarchitektur für Fahrzeug-Baken-Systeme und der dazugehörigen Protokolle (PDU – *Protocol Data Unit*). Der Kanal ermittelt unter Rückgriff auf die Positionen der Mobilstationen, ob die Datenübertragung zwischen der Bake und der Mobilstation gestört ist. Abbildung 6.26 zeigt eine Momentaufnahme einer animierten Simulation. Abbildung 6.27 zeigt unterschiedliche Verkehrssituationen: einzelnes Fahrzeug, Abschattung durch LKW und Umgebungsstörungen. In Abbildung 6.28 wird beispielhaft ein Schnittbild und ein Empfangspegelprofil für die Durchfahrt eines PKWs gezeigt. Die Bake ist bei der Position 10 m. Diese Beispiele stammen aus [Wie97].

Mit Hilfe der Simulation kann die Leistungsfähigkeit eines Mobilfunksystems vor einer flächendeckenden Einführung beurteilt werden. Es genügen Messungen in repräsentativen Testszenarien. Die Untersuchung von Alternativen kann kostengünstig per Simulation erfolgen.



Beispiel: Fahrzeug-Baken-Kommunikation

Abbildung 6.25: Klassenstruktur Mobilfunk-Simulationsbibliothek

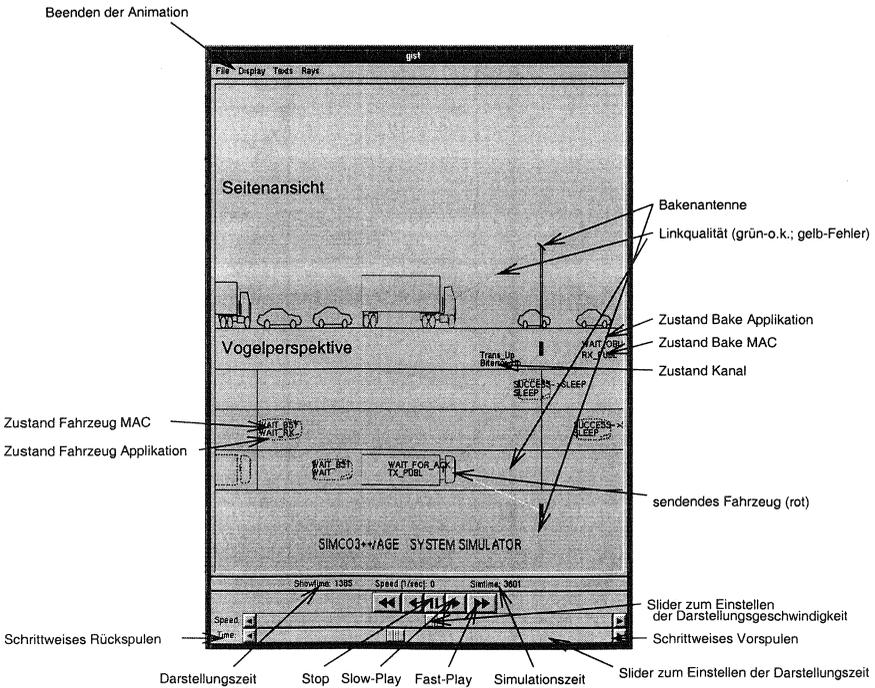
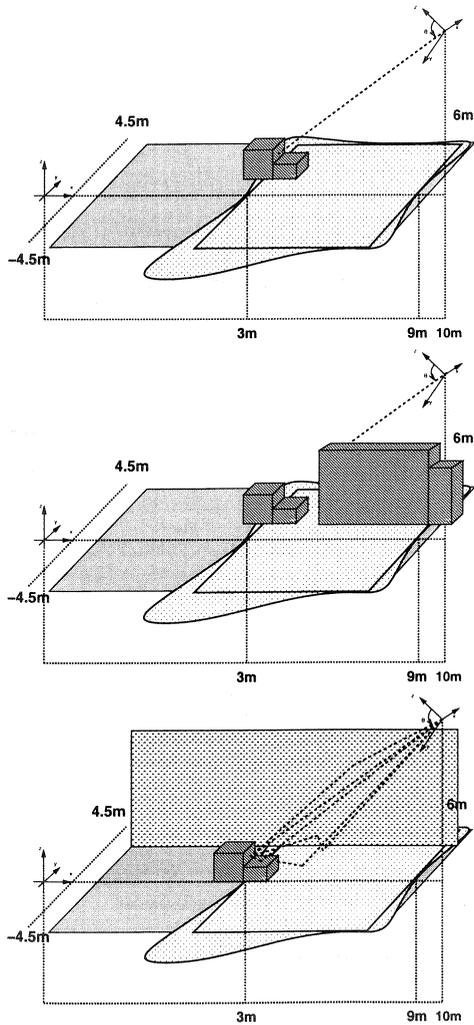
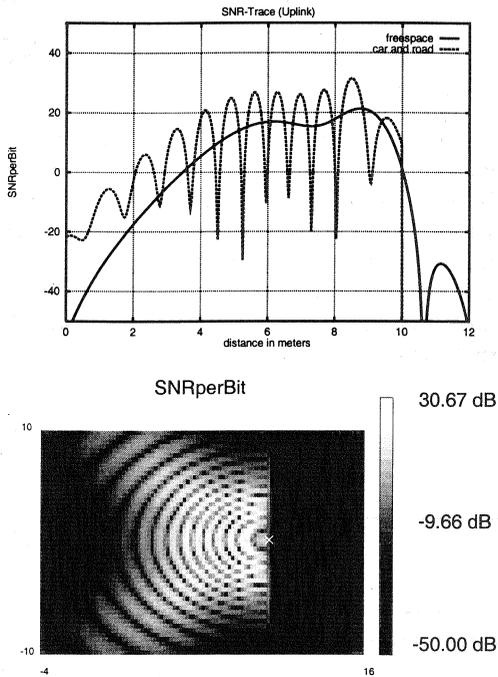


Abbildung 6.26: Momentaufnahme: Fahrzeug-Baken-Kommunikation



- Oben: einzelnes Fahrzeug
 Mitte: Abschattung durch LKW
 Unten: Umgebungsstörung

Abbildung 6.27: Unterschiedliche Verkehrssituationen



Oben: Empfangspegelprofil

Unten: Schnittbild

Abbildung 6.28: Durchfahrt eines PKWs in einem Autobahnszenario

6.3.2 DECT

DECT (*Digital Enhanced Cordless Telecommunication*) ist ein seit 1992 europaweit standardisiertes System für drahtlose Telefonie. Ursprünglich wurde es für die drahtlose Kommunikation in Gebäuden konzipiert. Im Zuge der Liberalisierung des Telekommunikationsmarktes und der Planungen neuer Netzbetreiber, drahtlose Zugangsnetze (RLL – *Radio in the Local Loop*) zu realisieren, wird es aufgrund seiner Eigenschaften – dezentrale Organisation, nahtloser Handover (*seamless handover*) gesteuert durch die Mobilstation, kleine Zellen, kostengünstige Realisierungen – für neue Einsatzbereiche erschlossen.

Im folgenden sollen die Haupteigenschaften des DECT-Systems kurz zusammengefaßt werden. Daran schließt sich eine Übersicht über Untersuchungen des DECT-Systems an. In [Ple96] wurden DECT-Systeme sowohl durch eine verkehrstheoretische Analyse als auch durch stochastische Simulation verschiedener Szenarien untersucht. Ziel dieser Untersuchungen war die Fragestellung: Können DECT-Systeme bzgl. Kapazität und Dienstgüte mit Festnetzen konkurrieren? Eine zweite Fragestellung betrifft die Beeinflußung von DECT-Systemen konkurrierender Betreiber (auch öffentlich und privat) im gleichen geographischen Bereich.

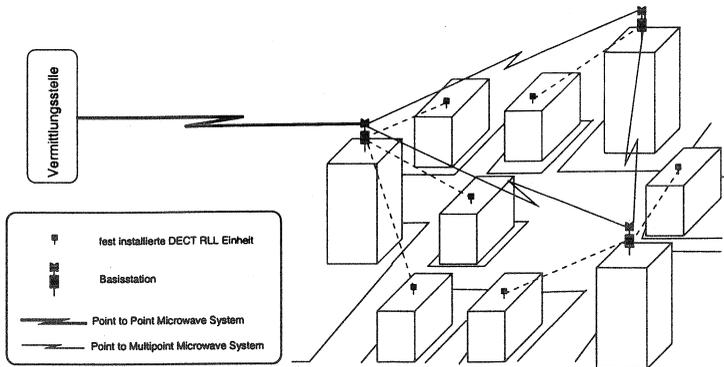


Abbildung 6.29: DECT-Zugangsnetz oberhalb der Dächer

Eigenschaften des DECT-Systems: Das DECT-System ist ein mikrozellulares, digitales Mobilfunknetz für Telefonie und Datendienste bzw. ISDN-Dienste, welches für den Einsatz auf privaten Grundstücken allgemein genehmigt und bereits eingeführt ist. Die ETSI-Standardisierung für den Einsatz als Zugangsnetz (RLL, siehe Abbildung 6.29) und öffentliches Mobilfunksystem (PCS – *Personal Communication System*, siehe Abbildung 6.30) sind weitgehend abgeschlossen und sehen den Betrieb im gleichen Frequenzbereich von 1880 – 1900 MHz vor.

Die Eigenschaften des DECT-Systems hängen stark von der Einsatz-Umgebung ab, so kann z. B. der maximale Abstand einer Mobilstation zu einer Feststation (RFP – *Radio Fixed Part*) im Freien sehr viel größer sein (ca. 250 m bzw. 2500 m mit Sektorantennen) als in einem Gebäude (ca. 30 m). Dies liegt an der unterschiedlichen Dämpfung der Signale. Durch die Möglichkeit des Handovers sind der Abmessung des Netzes und damit der räumlichen Mobilität der Teilnehmer keine Grenzen gesetzt.

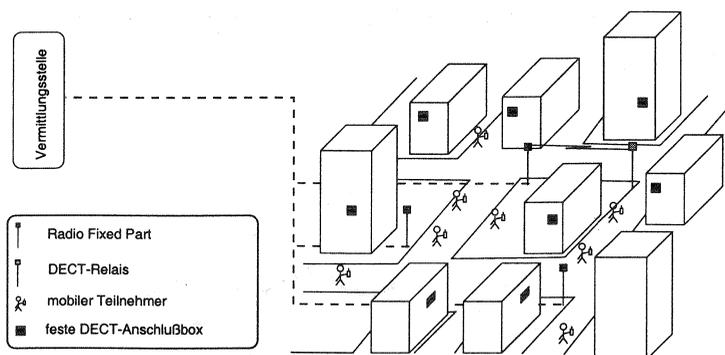


Abbildung 6.30: DECT als Mobilfunknetz und Zugangsnetz unterhalb der Dächern

Die Bitübertragungsschicht (physical layer) des DECT-Systems teilt das Funkspektrum in Kanäle ein. DECT verwendet eine Mischung aus Zeit- (TDMA – *Time Division Multiple Access*) und Frequenz-Multiplex-Zugriffsverfahren (FDMA – *Frequency Division Multiple Access*). Das Frequenzband 1880 – 1900 MHz wird in 10 Teilbänder bzw. Trägerfrequenzen aufgeteilt (Trägerabstand 1.728 MHz). Die maximale Sendeleistung ist auf 250 mW begrenzt. Jede der 10 Frequenzen wird in 10 ms lange Rahmen (*frame*) eingeteilt, die wiederum in 24 Zeitschlitze (*slot*) aufgeteilt sind. Der Rahmen hat die Bruttoübertragungsrate von 1152 kbit/s. Ein Zeitschlitz trägt 480 bit und ist 416 μ s lang.

Da für eine Duplexverbindung zwei Kanäle benötigt werden, ist die erste Rahmehälfte der Richtung von der Feststation zur Mobilstation (*downlink*) und die zweite der umgekehrten Richtung (*uplink*) zugeordnet. Die Zuordnung erfolgt nach der Zeitlagentrennung (TDD – *Time Division Duplexing*), wobei jeweils die gleiche Zeitlage innerhalb der ersten und der zweiten 12 Zeitschlitze für eine Duplexverbindung benutzt wird. Ein Zeitschlitz besteht aus einem Synchronisationsfeld (32 bit), einem Datenfeld (388 bit) und einer Schutzzone (*guard period*, 60 bit).

In der MAC-Schicht (MAC – *Medium Access Control*) wird das Datenfeld weiter unterteilt in ein A-Feld (64 bit) und ein B-Feld (324 bit). Im B-Feld werden im sog. ungeschützten Modus nochmal 4 bit für Fehlererkennungszwecke abgetrennt, so daß man in diesem Modus eine Datenrate von 320 bit/10 ms = 32 kbit/s erhält, wie man sie

für digitale Sprachübertragung benötigt. Eine genauere Beschreibung des DECT-Systems findet man in [ETCRE92, ETCRE94, ETS91a, ETS91b, ETS91c, ETS95] und [Rap96, Wal96b].

Die Kapazität von DECT-Systemen im Freien ist abhängig von der Dämpfung des Signals und von der Form der Zellen (hexagonal, sektorisiert, isoliert durch die Morphostruktur, siehe Abbildung 6.31 und 6.32). Innerhalb von Gebäuden ist die Position der Feststationen und die Beschaffenheit der Hindernisse entscheidend, z. B. Wände, Decken, Möbel.

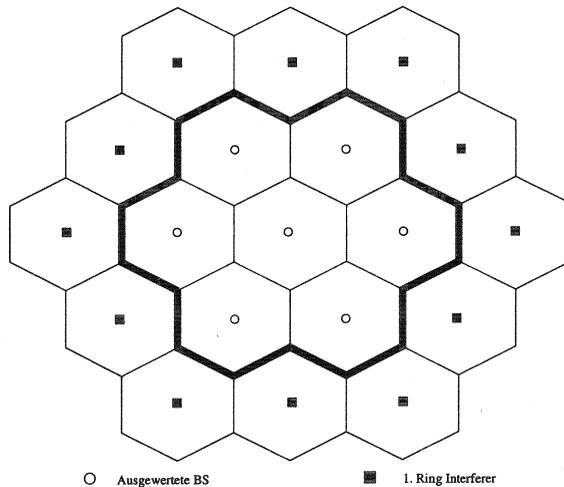


Abbildung 6.31: DECT-Szenario mit hexagonalen Zellen und 19 Feststationen

Die Verwaltung der Funkressourcen in DECT-Systemen unterscheidet sich grundsätzlich von anderen zellularen Systemen, z. B. dem GSM-System. Zum einen sind in allen Zellen eines DECT-Systems grundsätzlich alle Kanäle verfügbar (DCA – *Dynamic Channel Assignment*) während bei GSM eine feste Zuordnung benutzt wird (FCA – *Fixed Channel Assignment*). Im DECT-System erfolgt die Steuerung der Handover-Funktion durch die Mobilstation, während im GSM-System die Feststation diese Steuerung durchführt. Das DECT-System ist damit flexibler und kann sich besser auf wechselnde Lasten einstellen.

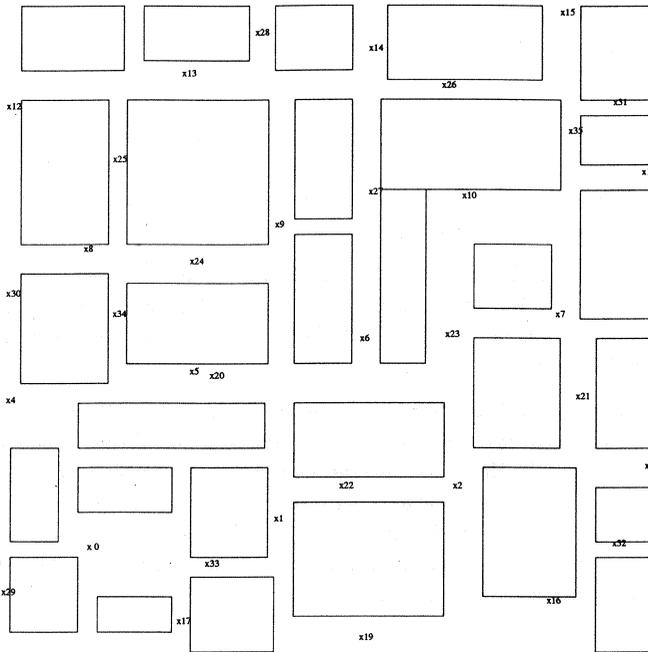


Abbildung 6.32: DECT-Szenario mit Morphostruktur und 36 Feststationen

Modelle für die Funkausbreitung: Sowohl analytische als auch simulative Untersuchungen von Funksystemen hängen entscheidend von der Realitätsnähe der Modellierung der Funkausbreitung ab.

Es kommen dabei einerseits deterministische Modelle zur Anwendung, die die Funkausbreitung mathematisch berechnen. Da eine genaue Berechnung zu aufwendig ist, werden sog. Strahlverfolgungsmethoden (*ray-tracing*) eingesetzt, bei denen die Funkwellen als Strahlen modelliert werden, die im betrachteten Szenario verfolgt, d. h. an Oberflächen reflektiert bzw. transmittiert, und an Ecken evtl. gebeugt werden. Der Mehrwegeeinfluß wird durch Mehrfachreflektionen und -transmissionen bis zu einer gewissen Tiefe berücksichtigt.

Bei der *Strahlaussendungsmethode (ray-launching)* werden Strahlen vom Sender aus in unterschiedlichen Winkeln abgestrahlt und verfolgt, um dann festzustellen, welche Strahlen in einer gewissen Umgebung des Empfängers (Empfangskugel) liegen.

Bei der *Spiegelungsmethode (imaging)*, die aufwendiger als die Strahlverfolgungsmethode ist, werden alle Spiegelungspunkte n -ten Grades des Empfängers und damit die exakten Strahlenwege zum Empfänger bestimmt.

Eine zweite Klasse von Methoden befaßt sich mit empirischen Modellen, die auf Messungen und daraus abgeleiteten Näherungsformeln für die Funkausbreitung beruhen. Die Modelle sind abhängig von den verwendeten Frequenzen, d. h. bei GSM werden z. B. andere Modelle als bei DECT verwendet. In [Ple96] werden zwei empirische Modelle der COST 231 Gruppe (*COST – European COoperation in the field of Scientific and Technical research*) und von CNET (*CNET – Centre Nationale d'Etude de Telecommunication*) zusammen mit der France-Telecom vorgestellt und mit den Strahlverfolgungs- bzw. Spiegelungsmethoden verglichen.

Die Signalleistung in einem DECT-System hängt aufgrund des benutzten Frequenzbereichs stärker von der Morphostruktur ab als bei GSM-Systemen, die im Bereich von 900 MHz arbeiten (z. B. das D1- und D2-Netz), d. h. die Morphostruktur muß genauer modelliert werden.

Zielgrößen des DECT-Systems: Ziel der Bewertung ist die Wahrscheinlichkeit, daß ein Verbindungswunsch abgewiesen wird (Blockierung, Verlust) und die Wahrscheinlichkeit, daß ein Handover-Vorgang und damit die Verbindung abgebrochen wird. Verkehrstheoretische Ansätze versuchen diese Größen (z. B. die Blockierungswahrscheinlichkeit) direkt zu bestimmen, während bei der Simulation die Anzahl der Ereignisse gemessen wird und dann mit der Gesamtzahl der Verbindungen im betrachteten Zeitraum in Relation gesetzt wird.

Bei Mobilfunksystemen wird als Zielgröße ein sog. *Servicegrad* (GOS – *grade of service*) definiert, der Blockierungen und Abbrüche gewichtet bewertet:

$$GOS = \frac{\text{Blockierungen} + 10 \cdot \text{abgebrochene Verbindungen}}{\text{Gesamtanzahl der Verbindungen bzw. Verbindungsversuche}} \quad (6.12)$$

Ein niedriger GOS -Wert bedeutet eine hohe Dienstgüte im System, da nur wenige Verbindungen abgewiesen werden und nahezu keine Verbindungen abgebrochen werden. Der GOS -Wert kann theoretisch zwischen 0 und 10 liegen. In den folgenden Untersuchungen wird ein Servicegrad von maximal 1 % gefordert ($GOS \leq 0.01$), da in Konkurrenz zum Festnetz nur DECT-Netze mit sehr hoher Dienstgüte in Betracht kommen.

Verkehrstheoretische Analyse des DECT-Systems: Abbildung 6.33 zeigt das DECT-System als Modell. In Mobilfunksystemen mit vielen Teilnehmern und damit Überlagerung vieler unabhängiger Verkehrsquellen kann der Ankunftsprozeß als negativ-exponentiell verteilt angenommen werden. Damit kann zur Berechnung der Verlustwahrscheinlichkeit P_B die Erlang-B-Formel benutzt werden [Kle75, Jai91], die für negativ-exponentiell verteilte aber auch für beliebig verteilte Verbindungsdauern, d. h. für das M/G/S Verlustsystem ohne Wartemöglichkeiten gilt, wobei A dem Verkehrsangebot und S der Anzahl der nutzbaren Kanäle entspricht:

$$P\{\text{Verlust}\} = P_B = \frac{A^S}{S!} \frac{1}{\sum_{i=0}^S \frac{A^i}{i!}} \quad (6.13)$$

Bei überlappenden Funkzellen muß das Modell erweitert werden. Die Verlustwahrscheinlichkeit bei zwei sich überlappenden Funkzellen mit je einem Transceiver hängt von dem Grad der Überlappung ab. Mit wachsender Überlappung wird die Verlustwahrscheinlichkeit kleiner, da immer mehr Mobilstationen zwei Feststationen zur Verfügung stehen, bis sie bei 100 %-iger Überlappung der Verlustwahrscheinlichkeit für ein System mit zwei Transceivern entspricht. Bei Kapazitätsbegrenzung wird also bei überlappenden Zellen aufgrund der dynamischen Kanalvergabe ein Kapazitätsgewinn erzielt.

Simulative Analyse des DECT-Systems: Das Simulationstool DESI (*DEct Simulator*), welches gezielt für die Untersuchung von DECT-Systemen entwickelt wurde, siehe Abbildung 6.34, basiert auf den SIMCO3++/CNCL-Klassenbibliotheken. Darauf aufbauend wurden Klassen für die DECT-Protokolle, die Kanalmodellierung, die Darstellung des untersuchten Szenarios sowie für die Lastgenerierung implementiert.

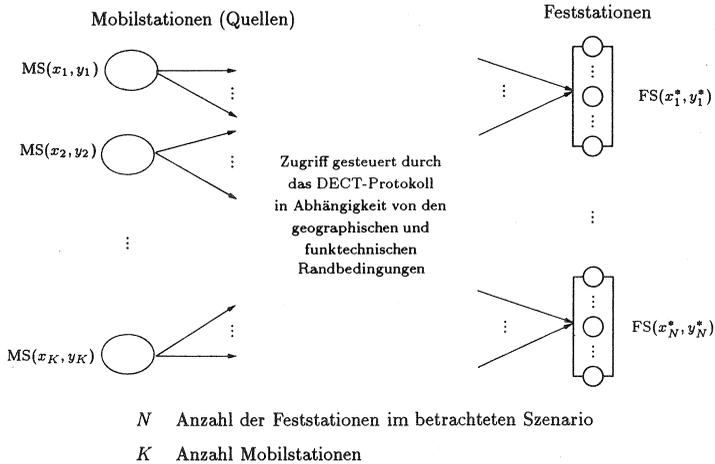


Abbildung 6.33: Modell des DECT-Systems

Ablauf einer Simulation:**Eingabedaten:**

z. B. Szenario, Position und Anzahl der Feststationen, Mobilitätsmodell und Lastverhalten der Mobilstationen

Initialisierung:

Berechnung der Empfangspegel aller Feststationen für jeden Punkt des Szenarios (Strahlaussendung oder Spiegelung)

Haupt-Simulationsschleife:

Zeitinkrement: Rahmentakt (10 ms)
 Aktualisierung aller Mobilstationen und Feststationen
 Meßwertaufbereitung, Statistische Auswertung
 graphische Ausgabe (GIST)
 Abbruchkriterium: maximale Modellzeit

Ausgabe der Ergebnisse:

z. B. Handover-Raten, Blockierraten, Verteilungsfunktion der C/I-Werte (C/I – Carrier-to-Interference Ratio, Träger-zu-Interferenz-Verhältnis)

Um einen Eindruck des Umfangs dieser Untersuchungen zu vermitteln, werden im folgenden einige Daten genannt. Es wurden für jeden Parametersatz 100 Minuten Modellzeit simuliert. Für jedes Szenario wurden ungefähr 5 bis 10 Läufe durchgeführt, um den tragbaren Verkehr für den vorgegebenen Wert $GOS \leq 1\%$ zu bestimmen. Einige Simulationsläufe benötigten auf leistungsfähigen Workstations (SUN/SPARC20) bis zu 2 Wochen. Insgesamt wurden ca. 70 unterschiedliche Szenarien ausgewertet.

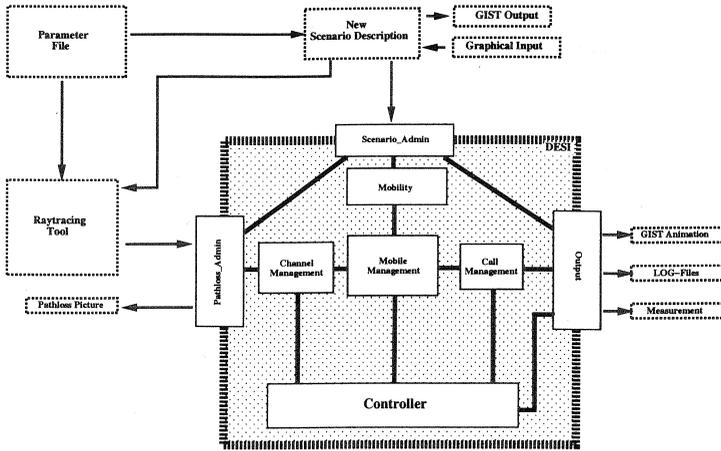


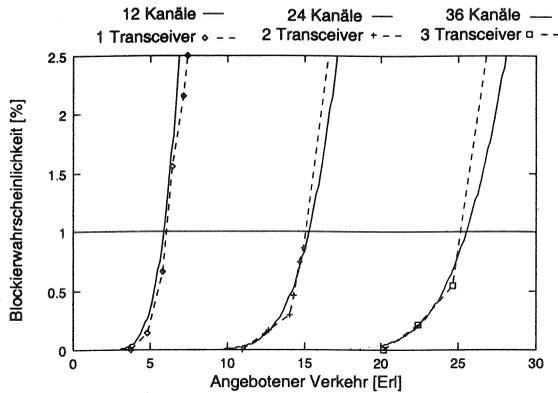
Abbildung 6.34: DESI-Simulator

In Abbildung 6.35 wird beispielhaft die Übereinstimmung von Simulationsergebnissen mit den Ergebnissen aus der verkehrstheoretischen Analyse gezeigt.

Mit vereinfachenden Annahmen führen Richtantennen bei festinstallierten Teilnehmeranschlüssen (ober- oder unterhalb der Dächer) dazu, daß meist nur eine Feststation genutzt wird und die Systeme sich dann ähnlich wie isolierte Zellen verhalten. Damit wird der tragbare Verkehr hauptsächlich durch die Anzahl der Transceiver der Feststation bestimmt, man spricht dann von Kapazitätsbegrenzung.

In der Simulation wurden die Teilnehmeranschlüsse der nächstgelegenen Feststation zugeordnet und Handover zwischen Zellen (Inter-Cell Handover) wurden unterbunden, d. h. Handover konnten nur innerhalb derselben Zelle durchgeführt werden (Intra-Cell Handover). Das Simulationsszenario entspricht Abbildung 6.31 mit 19 Zellen. Als Verkehrsquellen dienen Poissonquellen und die Verbindungsdauern sind negativ-exponentiell verteilt.

Das Ergebnis zeigt deutlich, daß der tragbare Verkehr durch Hinzufügen von Transceivern erhöht werden kann. Mit einem Transceiver können bis zu 12 Duplex-Kanäle gleichzeitig bereitgestellt werden, bei 2 Transceivern entsprechend 24, und 3 Transceiver ergeben 36 gleichzeitig nutzbare Kanäle. In diesem Szenario kommen aufgrund der gerichteten Antennen keine Handover zu anderen Zellen vor, so daß der Servicegrad hier im wesentlichen der Blockierungswahrscheinlichkeit entspricht und mit den Ergebnissen der Erlang-B-Formel mit der entsprechenden Anzahl von Bedienstationen (12,24,36) verglichen werden kann. Abbildung 6.35 zeigt, daß die Übereinstimmung für dieses Szenario gut ist, so daß hier eine Berechnung mit der Erlang-



Vergleich der Simulationsergebnisse mit den Resultaten der Erlang-B-Formel

Abbildung 6.35: Validierung des Szenarios für DECT-Teilnehmeranschlüsse

B-Formel erfolgen kann. Der tragbare Verkehr⁸ erhöht sich für einen Servicegrad von 1 % von 6 *Erl/RFP* über 15.3 *Erl/RFP* auf 25.2 *Erl/RFP*. Das heißt, es ist im Rahmen der Möglichkeiten günstiger eine Feststation mit mehreren Transceivern auszustatten als weitere Feststationen mit einem Transceiver vorzusehen.

Die Untersuchung von Szenarien mit konkurrierenden Betreibern hat gezeigt, daß der tragbare Verkehr pro Feststation bei vorgegebenem Servicegrad abhängig ist von der Position der Feststationen und von der Synchronität der jeweiligen Systeme. Als weiteres Beispiel zeigt Abbildung 6.36 die Beeinflussung zweier konkurrierender Betreiber bei Synchronität und bei Asynchronität (Verzögerung von etwas mehr als einem halben Rahmen: 5.208 *ms*). Durch die Synchronität wird der tragbare Verkehr in diesem Szenario von 6.42 *Erl/RFP* auf 7.12 *Erl/RFP* erhöht.

Zusammenfassung: In diesem Abschnitt wurde über eine Studie [Ple96, WPS95] zur Bewertung von öffentlichen DECT-Systemen mit mehreren konkurrierenden Betreibern berichtet, die die Eignung und Leistungsfähigkeit des DECT-Systems untersucht hat. Es wurde eine Vielzahl von Szenarien für unterschiedliche Anwendungsfälle untersucht: die Gesamtkapazität im Freien, die Beeinflussung unterschiedlicher Betreiber, die Wechselwirkung zwischen privaten und öffentlichen DECT-Systemen.

⁸Der Verkehr wird in *Erl/RFP* d. h. *Erlang/Radio Fixed Part* bzw. *Erlang/Feststation* angegeben.

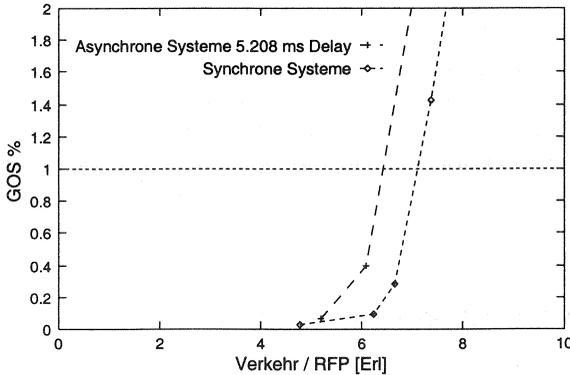


Abbildung 6.36: Vergleich von synchronen und asynchronen DECT-Systemen

Die Untersuchung wurde durch stochastische Simulation der relevanten Szenarien mit einer verkehrstheoretischen Absicherung der Ergebnisse durchgeführt.

In diesen Untersuchungen wurden die Parameter, die durch die hohe Flexibilität der DECT-Systeme und damit der Einfluß der Freiheitsgrade, wie z. B. Art und Aufstellungsort der Antennen, Anzahl der Transceiver, Dichte der Feststationen (RFP) untersucht. Dem Einfluß der Struktur der Umgebung (im Freien, unterschiedliche Materialien in Gebäuden) wurde durch unterschiedliche Funkausbreitungsmodelle Rechnung getragen.

Für alle untersuchten Szenarien lag der tragbare Verkehr deutlich über $140 \text{ Erl}/\text{km}^2$, d. h. Kleinkunden können gut versorgt werden. Die Leistungsgrenze ist einerseits durch die Anzahl der Transceiver (kapazitätsbegrenzt) und andererseits durch Belegung aller DECT-Kanäle (interferenzbegrenzt) gegeben.

Als RLL-Anschlußnetz mit einer Versorgung über den Dächern kann mit einem Servicegrad von $GOS \leq 1 \%$ mit rundstrahlenden Antennen ein Verkehr von ca. $4.8 \text{ Erl}/\text{RFP}$ und bei drei 120° Antennen ein Verkehr von ca. $10.8 \text{ Erl}/\text{RFP}$ getragen werden. Diese Ergebnisse sind vom Ausbreitungsmodell abhängig, so daß sie nicht direkt mit Ergebnissen verglichen werden können, die mit anderen Ausbreitungsmodellen erzielt wurden, wie z. B. ein Wert von ca. $13.1 \text{ Erl}/\text{RFP}$, der sich für das Mobilfunksystem (PCS-System) in den Straßen ergibt.

Bei mehreren DECT-Betreibern ist eine starke gegenseitige Beeinflussung festzustellen, die durch gemeinsame Planung verbessert werden könnte, da z. B. die Positionierung der Feststationen am gleichen Standort und eine Synchronisation der Systeme zu erheblichen Verbesserungen führen. Damit können Leistungseinbußen vermieden

werden. Weitere Untersuchungen müssen klären, ob die Synchronisation zwischen den Betreibern versetzt erfolgen sollte, da dadurch teilweise weitere Störungen vermieden werden können.

Folgearbeiten in diesem Gebiet werden durch z. Zt. laufende Feldversuche unterstützt. Dadurch können die empirischen Ausbreitungsmodelle verbessert werden und die gegenseitige Beeinflussung unterschiedlicher Anwendungen, wie z. B. DECT als Mobilfunknetz und als Zugangsnetz, und auch neue bzw. sich ändernde Anforderungen untersucht werden.

6.4 Simulationsprojekte in der Praxis

Begleitend zu dieser Arbeit wurde untersucht [Gör96], wie die Methode der stochastischen Simulation z. Zt. in der Praxis eingesetzt wird. Anhand von 50 Projekten aus 9 Ländern wurde ermittelt, welche Werkzeuge, welche Zufallszahlengeneratoren, welche statistischen Auswertemethoden, welche Abbruchkriterien und welche Validierungstechniken eingesetzt werden.

Die Projekte bezogen sich hauptsächlich auf ein weites Spektrum von zukünftigen Kommunikationsnetzen, z. B. Festnetze, Mobilfunknetze oder Lokale Netze, sowie einige ergänzende Projekte aus dem Bereich Produktionssysteme. Das Verhältnis der Simulationsuntersuchungen aus Industrie und Universität betrug in etwa 1:2.

Bei den verwendeten Simulationssystemen fiel auf, daß kommerzielle Systeme mit graphischen Oberflächen meist nur zur Untersuchung kleiner Modelle und Voruntersuchungen mit Vorführungen benutzt werden. Bei großen Projekten waren die Anforderungen bzgl. Geschwindigkeit und Speicherplatz meist so anspruchsvoll, daß auf effizientere Simulationsbibliotheken zurückgegriffen werden mußte, die überwiegend Eigenentwicklungen sind. Die Zeit- und Speicherplatzanforderungen variierten erwartungsgemäß sehr stark. Es wurden einige wenige bis zu mehrere tausend Simulationsläufe pro Projekt durchgeführt. Die meisten Simulationen wurden auf Workstations mit 2 bis 64 MB Speicherplatz durchgeführt, ein Projekt sogar mit 150 MB. Die Laufzeit rangierte von einigen Minuten bis zu mehreren Tagen. Nur sehr selten wurden Simulationsläufe durchgeführt, die länger als eine Woche dauerten, auch wenn die Genauigkeit der Ergebnisse nicht zufriedenstellend war.

Ein weiteres wichtiges Ergebnis ist, daß meist nur unzureichende oder keine statistischen Auswertemethoden angewandt werden, obwohl die Notwendigkeit besserer Methoden oft gefordert wird.

Es wurden meist lineare Kongruenz-Generatoren eingesetzt. Neue weiterentwickelte Zufallsgeneratoren kommen kaum zur Anwendung.

Zur Validierung wurde in ca. 30 % der Projekte der Vergleich mit Messungen durchgeführt. Häufig wurden Simulationsergebnisse mit analytisch berechneten Werten (ca. 50 %) bzw. publizierten Ergebnissen (ca. 5 %) verglichen. In den restlichen Fällen wurde keine explizite Validierung durchgeführt.

Techniken zur Beschleunigung der Simulationslaufzeit wurden in den ausgewerteten Projekten kaum eingesetzt, obwohl die Problematik in einigen Projekten gegeben war und die Notwendigkeit dazu ausdrücklich gesehen wurde. Im folgenden Kapitel wird eine Übersicht über derartige Techniken gegeben, um im darauf folgenden Kapitel eine spezielle Methode weiterzuentwickeln mit dem Ziel den Einsatz dieser Techniken in der Praxis zu unterstützen.

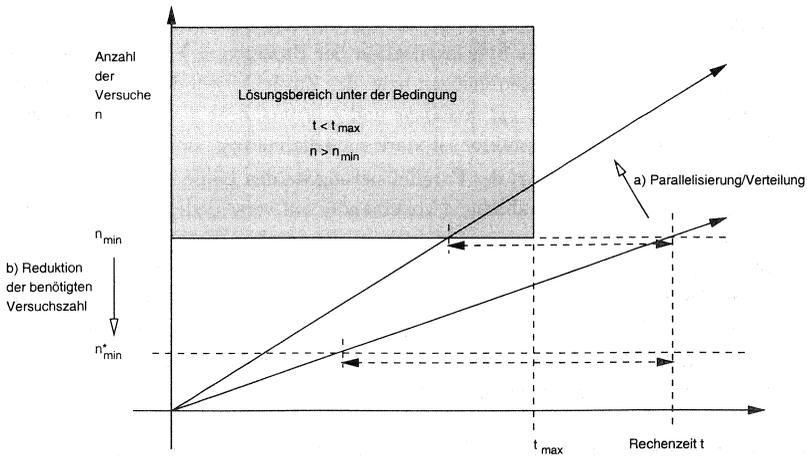
Verfahren zur Simulationsbeschleunigung

Eine bemerkenswerte Simulationsbeschleunigung ohne den Einsatz spezieller komplexer Verfahren wird bereits durch den allgemeinen technischen Fortschritt auf dem Gebiet der Mikroelektronik und die damit verbundene Steigerung der Prozessorgeschwindigkeiten erreicht. Andere Verfahren der Simulationsbeschleunigung werden dadurch jedoch nicht überflüssig, da auch die Komplexität der Modelle wächst, die simulativ untersucht werden, und inhärente Probleme, wie z. B. die Simulation seltener Ereignisse, weiterhin bestehen bzw. gerade durch den Fortschritt der Technik ins praktische Interesse rücken, z. B. extreme Anforderungen an die Dienstgüte bzw. Verkehrsgüte neuer Hochgeschwindigkeitsnetze. Hier kann die Simulationstechnik helfen, Kosten bei der Erprobung zu sparen und die Leistung durch effiziente Algorithmen zu steigern.

Abbildung 7.1 zeigt die beiden prinzipiellen Möglichkeiten eine Simulationsuntersuchung zu beschleunigen, d. h. die Rechenzeit zu verkürzen. Die Beschleunigung durch Parallelisierung bzw. Verteilung ist eine offensichtliche Methode der Beschleunigung, indem mehr Versuche pro Zeiteinheit durchgeführt werden. Die zweite Möglichkeit faßt alle Verfahren zusammen, die versuchen genauere Ergebnisse mit weniger Versuchen durch stochastische Methoden zu erzielen.

In Kapitel 8 wird das RESTART/LRE-Verfahren beschrieben, mit dem seltene Ereignisse mit vertretbaren Laufzeiten simulativ untersucht werden können. Vorher wird eine allgemeine Übersicht zu Verfahren zur Beschleunigung von Simulationsläufen gegeben, die sich teilweise auch gemeinsam einsetzen lassen. Eine Kombination der Methoden bedarf jedoch einer genauen Analyse, da sich statt einer Beschleunigung durch die Kombination auch eine Verlängerung ergeben kann.

Tabelle 7.1 gibt eine Übersicht und Einordnung möglicher Verfahren zur Beschleunigung von Simulationen. Die Prinzipien der einzelnen Verfahren werden im folgenden vorgestellt. Übersichten sowie vertiefende Kapitel zu diesem Thema findet man in [LO89], [LK91], [BFS87] und [Hee95]. Zunächst werden Verfahren zur *verteilten und parallelen Simulation* sowie *hybride Techniken* kurz skizziert. Alle weiteren Verfahren beruhen auf stochastischen Ansätzen, die z. B. Korrelationen ausnutzen, um die Varianz der Stichprobe zu reduzieren (*variance reduction*): *gemeinsame Zufallszahlen*, *antithetische Stichproben*, *Steuervariablen* und *indirekte Schätzung*. Speziell zur Behandlung seltener Ereignisse sind die Verfahren *Geschichtete Stichproben* (*stra-*



- a) Erhöhung der Versuchsrate durch parallele bzw. verteilte Ausführung
 b) Reduktion der benötigten Versuchszahl bei gleicher Genauigkeit durch stochastische Methoden

Abbildung 7.1: Möglichkeiten der Simulationsbeschleunigung

tified sampling), *Wesentliche Stichproben (importance sampling)* und *Wesentliche Partitionierung (importance splitting)* geeignet.

7.1 Verteilte und parallele Simulation

Von *paralleler Simulation* spricht man, wenn es sich um parallele Datenströme, aber nur eine Instruktion handelt (SIMD – *Single Instruction Multiple Data*). Dies ist z. B. bei Vektorrechnern der Fall. Im allgemeinen läßt sich durch diese Art der Parallelisierung nur eine geringe Simulationsbeschleunigung erreichen. Geeigneter ist die sog. *verteilte Simulation*, bei der das Simulationsprogramm auf mehrere selbständige Prozessoren bzw. Workstations (MIMD – *Multiple Instruction Multiple Data*) verteilt wird.

Bei der verteilten Simulation kann die Aufteilung des Simulationsprogramms auf die verschiedenen Prozessoren nach folgender Systematik erfolgen:

Funktionsaufteilung: Das Programm wird entsprechend der Simulationsfunktionen aufgeteilt: mehrere parallele Zufallsgeneratoren, mehrere parallele statistische Auswertungen und graphische Ausgaben, aber nur *eine* Ablaufsteuerung. Der Beschleunigungsfaktor bei dieser Art der Parallelisierung ist nicht sehr hoch, da die Begrenzung bei der Ablaufsteuerung liegt, aber es gibt kaum Synchronisationsverluste, siehe [Leh79, Bar83].

Modellaufteilung: Die Modellprozesse werden als logische Prozesse parallelisiert.

Das Hauptproblem ist die Synchronisation der Prozesse, d. h. die Vermeidung bzw. Auflösung von Verklemmungen bzw. das Zurücksetzen der Simulation. In der Literatur werden sog. *optimistische* und *pessimistische* Strategien beschrieben. Der Beschleunigungsfaktor ist stark modellabhängig, siehe [Fuj90].

Lastaufteilung: Bei dieser Art der Parallelisierung werden Läufe mit unterschiedlichen Parametersätzen unabhängig voneinander auf verschiedenen Workstations ausgewertet. Der Beschleunigungsfaktor ist maximal im Vergleich zu sequentiellen Läufen auf einer Workstation. Diese Art der Parallelisierung wird meist in der Praxis angewandt und kann auch zur Beschleunigung einer Auswertung benutzt werden, indem man das Simulationsprogramm mit unterschiedlichen unabhängigen Zufallsströmen startet und anschließend die Ergebnisse der statistischen Auswertung zusammenführt.

Die maximale Beschleunigung, die durch verteilte Simulationsläufe erreicht werden kann, wird einerseits durch die Anzahl der Prozessoren bzw. Workstations und andererseits durch die Synchronisationsverluste eingeschränkt [HW83]. Damit sind dieser Art der Beschleunigung praktische Grenzen gesetzt. Die Anzahl der Prozessoren ist meist kleiner als 100 und die effektive Beschleunigung ist z. T. stark modellabhängig und erreicht z. B. Werte von 5 bei 9 Prozessoren oder 57 bei 64 Prozessoren [Fuj90]. Andere Verfahren, wie sie in dieser Arbeit untersucht werden, erreichen für Spezialfälle Größenordnungen von z. B. 10^{15} , siehe Kapitel 8.

7.2 Hybride Techniken

Hybride Techniken kombinieren die Flexibilität der Simulation mit der Effizienz der analytischen Methoden.

Eine Möglichkeit Simulation mit analytischen Methoden zu kombinieren bietet die Methode der *Dekomposition* bzw. *Aggregation* [Cou77, Küh79] bei netzartigen Modellen, bei der ein Teil des Netzes durch ein einfacheres Ersatzsystem bzw. Teilmodell zusammengefaßt wird, welches hinsichtlich des Durchsatzes und der Durchlaufzeiten dem Teilnetz äquivalent ist. Das Verfahren ist z. B. dann sehr effektiv, wenn die Teilmodelle analytisch behandelt werden können. Auch falls keine analytischen Lösungen existieren, ist die Dekomposition noch effektiv, da die Teilmodelle nur einmal simuliert werden müssen.

Das Verfahren der Dekomposition ist stark modellabhängig und verlangt detaillierte Systemkenntnisse. Häufig läßt sich das Verfahren nur approximativ anwenden.

Eine andere Art der Aufteilung erlaubt die Methode der sog. *Bedingten Erwartung* (*conditional expectation*). Bei diesem Ansatz bildet die Aufteilung eines Erwartungs-

wertes unter Verwendung von bedingten Wahrscheinlichkeiten den Ausgangspunkt. Man nutzt spezielle Beziehungen aus und erreicht durch das Einsetzen eines *exakten analytischen Wertes* anstatt eines Schätzwertes, daß die Varianz des Zielschätzwertes reduziert wird, da man eine Quelle der Variabilität ausgeschlossen hat. Größen ohne analytische Lösung werden durch Simulation bestimmt, z. B. in einem einfachen System mit zwei Bedienern: die Wahrscheinlichkeit, daß ein bzw. zwei Bediener aktiv sind sei als analytische Lösung bekannt, der Erwartungswert unter der Bedingung "A=ein Bediener aktiv" bzw. "B=zwei Bediener aktiv" wird simulativ ermittelt:

$$E[X] = P\{A\}E[X|A] + P\{B\}E[X|B] \quad (7.1)$$

Auch diese Methode ist modellabhängig und erfordert, daß eine geeignete Bedingung identifiziert wird, die eine analytische Lösung oder eine geeignete Unterteilung erlaubt.

7.3 Varianzreduktion

Bei den Methoden der Varianzreduktion wird eine Eigenschaft des Systems ausgenutzt, um die Varianz der Zielgröße zu reduzieren. Die Anwendung dieser Methoden setzt voraus, daß man die Arbeitsweise des Modells sehr genau kennt. Ein Nachteil der Methoden ist, daß eine Vorhersage der Effizienzsteigerung im allgemeinen nicht möglich ist. Außerdem muß man berücksichtigen, daß die Methode selbst zusätzlichen Aufwand verursacht. Im folgenden werden beispielhaft mehrere Methoden skizziert, die Abhängigkeiten zwischen Beobachtungen ausnutzen, um die Varianz zu reduzieren: Gemeinsame Zufallszahlen (*common random numbers*), Antithetische Stichprobe (*antithetic sampling*), Steuervariablen (*control variables*) und Indirekte Schätzung (*indirect estimation*).

7.3.1 Gemeinsame Zufallszahlen (*common random numbers*)

Bei dieser Methode werden zwei oder mehr Systemkonfigurationen unter ähnlichen Bedingungen verglichen, so daß Unterschiede im wesentlichen auf Änderungen der Systemkonfiguration und nicht der Experimentbedingungen zurückzuführen sind. Experimentbedingungen sind bei der Simulation hauptsächlich die generierte Folge von Zufallszahlen. Benutzt man die *gleiche* Folge von Zufallszahlen in beiden Konfigurationen, so kann häufig eine positive Korrelation zwischen den Systemen erzeugt werden, die die Varianz der zu messenden Vergleichsgröße reduziert. Leider gibt es keinen allgemeinen Beweis, daß diese Methode immer zu einer Varianzreduktion führt. Doch auch wenn die Methode zu einer Varianzreduktion führt, weiß man erst im nachhinein, wie stark die Verbesserung sich auswirkt. Es gibt jedoch einige Modellklassen, für die die Varianzreduktion garantiert werden kann [BFS87].

Nachteilig wirken sich auch bei dieser Methode die häufigen Korrelationen innerhalb der beobachteten Sequenz eines Systems aus.

Wichtig ist die Synchronisation der Zufallsströme. Dies ist ein Beispiel, wo es notwendig sein kann, Zufallsströme mit verschiedenen Startwerte zu benutzen und nicht die empfohlene Methode zu benutzen, bei der alle Zufallszahlen aus einem Zufallsstrom erzeugt werden. Außerdem benötigt der Experimentator Kenntnisse über die Erzeugungsmethode der Verteilungen, z. B. ob mehr als eine Basiszufallszahl, eine feste Anzahl oder eine zufällige Anzahl benötigt wird. Gegebenenfalls müssen Zufallszahlen aus Synchronisationsgründen übersprungen werden oder es müssen alle Attribute eines Auftrags zweckmäßigerweise bei der Ankunft erzeugt werden. In jedem Fall sind sehr gründliche Kenntnisse über die Arbeitsweise des Modells und des Simulationssystems zur Anwendung dieser Methode notwendig. Andere Bezeichnungen für diese Methode sind: *correlated sampling*, *matched streams*, *matched pairs*. Vom Prinzip her ist die Methode beim *Vergleich* von Systemen nützlich und wird relativ zu anderen Methoden häufig angewandt.

Als Beispiel wird der Vergleich M/M/1 und M/M/2 [LK91] skizziert. X_{ij} sei die mittlere Durchlaufzeit der j -ten unabhängigen Replikation, d. h. des j -ten Simulationslaufs, im System M/M/ i für $i = 1, 2$. Unabhängige Replikationen werden mit Startwerten für die Zufallszahlen initialisiert, die zu unabhängigen Sequenzen von Zufallszahlen führen. Als Zielgröße zum Vergleich wird die Differenz $Z_j = X_{1j} - X_{2j}$ betrachtet. Als Fehlermaß wird der Schätzwert der Varianz $V(n)$ benutzt:

$$V(n) = \frac{\sum_{j=1}^n (Z_j - E\{Z_j\})^2}{n - 1} \quad (7.2)$$

Bei der Auswertung eines Beispiels mit $n = 100$ unabhängigen Replikationen ergibt sich für den Vergleich ohne Synchronisation als Schätzwert für die Varianz $V(100) = 18.00$ und die Korrelation beträgt $\text{Corr}(X_{1j}, X_{2j}) = -0.17$. Zur Synchronisation werden im j -ten Simulationslauf die Startwerte des Generators für die betrachtete Sequenz in beiden Systemen gleich gewählt. Werden nun die Zwischenankunftszeiten synchronisiert, so ergeben sich folgende Reduktionen: $V(100) = 9.02$, $\text{Corr}(X_{1j}, X_{2j}) = 0.33$, analog bei der Synchronisation der Bediendauern: $V(100) = 8.80$, $\text{Corr}(X_{1j}, X_{2j}) = 0.44$. Erwartungsgemäß ist die Reduktion bei der Synchronisation von Zwischenankunftszeit und Bediendauer am größten: $V(100) = 0.07$, $\text{Corr}(X_{1j}, X_{2j}) = 0.995$. In diesem Beispiel ergibt sich bei gleicher Genauigkeit ein Beschleunigungsfaktor von $18.00/0.07 = 257.14$ durch Varianzreduktion.

7.3.2 Antithetische Stichproben (*antithetic sampling*)

Bei der Methode der sog. *antithetischen Stichproben* wird versucht negative Korrelation zwischen verschiedenen Läufen eines Systems hervorzurufen: *kleine* Werte in einem Lauf sollen mit *großen* Werten in einem anderen Lauf korrespondieren. Dahinter steht die Idee, daß der Mittelwert jeweils einer *kleinen* und einer *großen* Beobachtung tendenziell näher am gemeinsamen Erwartungswert μ liegt, den es zu bestimmen gilt, als wenn unabhängige Beobachtungen benutzt würden. Eine Möglichkeit bieten sog. komplementäre Zufallszahlen ($U_k, 1 - U_k$), z. B. um jeweils die k -te Bediendauer zu erzeugen. Auch hier ist eine Synchronisation der Läufe wichtig, um die gewünschte negative Korrelation zu erzeugen, da man sonst eventuell unerwünschte entgegengesetzte Effekte erhält. Es sind ähnliche Methoden der Synchronisation anwendbar wie bei der oben beschriebenen Methode der gemeinsamen Zufallszahlen.

Man erhält bei dieser Methode n Paare von Läufen $(X_1^{[1]}, X_1^{[2]}), \dots, (X_n^{[1]}, X_n^{[2]})$ mit $X_j^{[1]}$ aus dem ersten Lauf, der U_k benutzt und $X_j^{[2]}$ aus dem zweiten synchronisierten Lauf mit $1 - U_k$. Für die Beobachtungswerte aus beiden Läufen gilt $E[X_j^{[1]}] = E[X_j^{[2]}] = \mu$, die Werte sind paarweise unabhängig und mit $X_j = (X_j^{[1]} + X_j^{[2]})/2$ und $\bar{X}(n)$ als Mittelwert der X_j , gilt insgesamt $\mu = E[X_j^{[1]}] = E[X_j^{[2]}] = E[X_j] = E[\bar{X}(n)]$. Da X_j unabhängig und identisch verteilt ist (i.i.d. - *independent and identically distributed*), gilt

$$\text{Var}[\bar{X}(n)] = \frac{\text{Var}[X_j]}{n} = \frac{\text{Var}[X_j^{[1]}] + \text{Var}[X_j^{[2]}] + 2\text{Cov}(X_j^{[1]}, X_j^{[2]})}{4n} \quad (7.3)$$

Der Gewinn bei dieser Methode ist nicht vorhersagbar, d.h. es sind Pilotstudien empfehlenswert. Für ausgewählte Modellklassen konnte die Varianzreduktion dieser Methode nachgewiesen werden, siehe [BFS87]. Erweiterungen dieser Methoden beziehen sich auf die Zielgrößen (nicht nur Erwartungswerte) und die Methoden wie die negative Korrelation erzeugt wird, siehe [LK91].

7.3.3 Steuervariablen (*control variables*)

Auch bei dieser Methode wird angestrebt, Korrelationen zwischen bestimmten Zufallsvariablen auszunutzen. Diese Korrelationen können bereits im Modell vorhanden sein oder sie können zusätzlich induziert werden.

Das Prinzip soll an folgendem Beispiel erläutert werden. Zielgröße ist der Erwartungswert $\mu = E[X]$ einer Zufallsvariablen X , z. B. der Mittelwert der ersten 100 Wartezeiten. Gesucht wird eine weitere Zufallsvariable Y mit bekanntem Erwartungswert $\nu = E[Y]$, die mit X (positiv oder negativ) korreliert ist, z. B. Mittelwert der ersten 99 abgeschlossenen Bediendauern, die bekannt sind, da sie nach Vorgabe generiert werden. Es liegt eine positive Korrelation vor, da lange Bediendauern zu

langen Wartezeiten führen und umgekehrt. Y wird als *Steuervariable* benutzt. Ist $Y > \nu$ dann wird angenommen, daß auch $X > \mu$ ist und es erfolgt eine entsprechende Anpassung. Sowohl positive als auch negative Korrelation kann ausgenutzt werden.

Die Anpassung des gesteuerten Schätzwertes X_C erfolgt unter Verwendung von $Y - \nu$, und einer Konstanten a mit gleichem Vorzeichen wie die Korrelation:

$$X_C := X - a(Y - \nu) \quad \text{mit} \quad E[X_C] = \mu \quad (7.4)$$

$$\text{Var}[X_C] = \text{Var}[X] + a^2 \text{Var}[Y] - 2a \text{Cov}(X, Y) \quad (7.5)$$

Man erhält eine Varianzreduktion für X_C genau dann, wenn $2a \text{Cov}(X, Y) > a^2 \text{Var}[Y]$ gilt. Die Erfüllung dieser Relation ist abhängig von der Wahl von Y und a . Das Hauptproblem dieser Methode ist die Bestimmung von Y und a . Theoretisch kann a optimal gewählt werden [LK91]:

$$a_{opt} = \frac{\text{Cov}(X, Y)}{\text{Var}[Y]} \quad (7.6)$$

Da die Kovarianz meist nicht bekannt ist, werden verschiedene Ansätze zur Schätzung der Parameter aus der Simulation verwendet. Die Methode kann auf mehrere Steuervariablen verallgemeinert werden. Man unterscheidet *interne* Steuervariablen, z. B. Eingabevariablen oder Funktionen davon, mit meist bekannten Erwartungswerten, die bereits als Teil des Modells generiert werden und somit keinen zusätzlichen Aufwand verursachen, und *externe* Steuervariablen, z. B. die gleiche Zufallsvariable X bei einem vereinfachten Modell, d. h. es müssen 2 Modelle simuliert werden, so daß man ein ähnliches Verfahren wie bei den gemeinsamen Zufallszahlen erhält.

7.3.4 Indirekte Schätzung (*indirect estimation*)

In Wartesystemsimulationen werden häufig stationäre Größen, wie z. B. die mittlere Durchlaufzeit, mittlere Wartezeit, mittlere Anzahl im System oder die mittlere Anzahl in der Warteschlange bestimmt, die in einer funktionalen Abhängigkeit zueinander stehen:

$$E[\text{Durchlaufzeit}] = E[\text{Wartezeit}] + E[\text{Bediendauer}] \quad (7.7)$$

und nach der Formel von Little, Gleichung (3.51), die für alle Systeme gilt:

$$E[\text{Anzahl im System}] = \lambda E[\text{Zeit im System}] \quad (7.8)$$

wobei das System sowohl die Warteschlange als auch das Gesamtsystem sein kann und die Formel sich somit sowohl auf die Anzahl der Wartenden und die Wartezeit als auch auf die Anzahl im Gesamtsystem und die Durchlaufzeit beziehen kann.

Wenn solche funktionalen Abhängigkeiten vorliegen, muß nur eine der beiden Größen ausgewertet werden. Dabei sollte die Größe mit der kleineren Varianz ausgewählt werden. Es kann gezeigt werden, daß sich die Varianz verkleinert, wenn man die Anzahl indirekt über die Zeit bestimmt [LK91]. Außerdem kann gezeigt werden, daß man besser die Wartezeit auswertet und zur Bestimmung der Durchlaufzeit dann den bekannten Mittelwert der Bediendauer addiert, statt die Durchlaufzeit direkt auszuwerten. Man benötigt also nur die mittlere Wartezeit, um daraus die mittlere Durchlaufzeit und die mittlere Anzahl im System bzw. in der Warteschlange zu bestimmen. Eine Schwäche dieser indirekten Methode ist, daß mit höherer Last ($\eta \rightarrow 1$) die Varianzreduktion gegen Null strebt. Gerade bei hoher Last wäre eine Varianzreduktion aufgrund der höheren Variabilität der Ergebnisse wünschenswert. Weitere Methoden erhält man durch die Kombination von direktem und indirektem Schätzwert. In Studien sind typischerweise Varianzreduktionen von z. B. mindestens 40 % gegenüber direkter Auswertung gezeigt worden.

7.4 Simulation seltener Ereignisse

Da die Laufzeit einer Simulationsuntersuchung immer beschränkt ist, ist insbesondere die Genauigkeit mit der ein *seltene Ereignis* in einer Sequenz aufgespürt werden kann, sehr begrenzt. Bei der simulativen Untersuchung muß man sich deshalb aller Hilfsmittel zur Steigerung der Simulationseffizienz, insbesondere der statistischen Effizienz, bedienen.

Dieses Problem ist derzeit bei der Leistungsbewertung von Hochgeschwindigkeitsnetzen, wie z. B. der Auslegung von Vermittlungsknoten des Breitband-ISDN, das nach dem asynchronen Übermittlungsverfahren (ATM – *Asynchronous Transfer Mode*) überträgt, aktuell. Ausgangsleitungen von Vermittlungsknoten werden in der Regel durch ein endliches Wartesystem vom Typ $G/G/1/N$ bzw. $G/D/1/N$ modelliert, wobei ein korrelierter, allgemeiner Eingabeprozess “G” vorliegt. Die Auslegung dieser Systeme soll garantieren, daß das Ereignis “Zellverlust” ein seltenes Ereignis mit einer Wahrscheinlichkeit in der Größenordnung $10^{-7} \dots 10^{-9}$ [Rec96, LB91] ist.

Bei sehr hohen Belastungen eines Wartesystems ist auch die Wahrscheinlichkeit für das leere System ein sehr kleiner Wert. Andere Beispiele für seltene Ereignisse sind die Randbereiche der Verteilungen der Zellverzögerung (*cell delay*) und Zellverzögerungsschwankung (*cell delay variation*) in ATM-Netzen sowie Fehlerereignisse, z. B. Bit- oder Symbolfehler oder Synchronisationsverluste.

Bei der direkten Simulation entspricht die Auftrittswahrscheinlichkeit eines Ereignisses der relativen Häufigkeit in einer endlichen Stichprobe. Soll ein seltenes Ereignis A mit z. B. $P\{A\} = 10^{-9}$ simulativ bestimmt werden, so benötigt man im Mittel für ein Auftreten des Ereignisses 10^9 Versuche. Bei einer realistischen Versuchsrate von 10^4 Versuchen pro Sekunde benötigt man dazu ca. 28 Stunden. Fordert man nach einer Daumenregel für zuverlässige Aussagen, daß das Ereignis mindestens 100 mal auftritt, so führt das zu 100×28 Stunden oder ungefähr 117 Tagen und ist damit mit direkter Simulation praktisch nicht untersuchbar.

Die Methoden, die hier zur Anwendung kommen, gehören ebenfalls zur Klasse der Varianzreduzierenden Methoden. Im folgenden wird eine Übersicht über drei Methodenklassen gegeben, mit denen insbesondere *seltene Ereignisse* untersucht werden können. Beim sog. *stratified sampling* wird die Grundgesamtheit in Schichten (*strata*) eingeteilt, so daß die Anzahl in den einzelnen Schichten getrennt erhöht werden kann. Bei den Methoden des *importance sampling* werden die zugrundeliegenden Prozesse verändert, so daß die seltenen Ereignisse häufiger auftreten, während beim sog. *importance splitting* das System selbst manipuliert wird, um seltene Ereignisse zu provozieren.

7.4.1 Geschichtete Stichproben (*stratified sampling*)

Diese Methode wird häufig bei Meinungsumfragen benutzt, um die Genauigkeit der Umfrageergebnisse verschiedener Bevölkerungsgruppen zu erhöhen.

Die Methode basiert auf der Darstellung der Zielgröße $E[Y]$ mit Schätzwert \tilde{v} als gewichtete Summe bzgl. der gewählten Aufteilung in L Schichten mit Mittelwert $E[Y_i]$:

$$E[Y] = \sum_{i=1}^L p_i E[Y_i] \quad (7.9)$$

Dies erlaubt es in den Schichten, wo die Variabilität am höchsten ist, mehr Stichproben zu sammeln,

Als Beispiel soll das Wartesystem GI/GI/s mit s Bedienstationen betrachtet werden. Die Anzahl s nimmt für jeden Tag verschiedene Werte $s = 1, 2, 3, 4$ mit Wahrscheinlichkeit p_1, p_2, p_3, p_4 an. Die Wahrscheinlichkeiten für nur eine oder zwei Bedienstationen sei relativ klein. Die Wahrscheinlichkeit $P\{\tau_w > T\}$, daß ein Kunde mindestens T Zeiteinheiten auf Bedienung warten muß, ist am wahrscheinlichsten an Tagen mit nur einer Bedienstation, d. h. diese Größe ist korreliert mit der Anzahl der Bedienstationen. Man sollte also Tage mit nur einer Bedienstation häufiger simulieren als Tage mit 4 Bedienstationen. Man kann z. B. proportional vorgehen, d. h. bei n Simulationen $n \cdot p_i$ Simulationen für das Modell mit i Bedientern durchführen.

Die Hauptprobleme bei dieser Methode bestehen in der richtigen Auswahl der Schichten, der Bestimmung der Versuchszahl pro Schicht und – falls nicht vorher bekannt – auch in der Bestimmung der Wahrscheinlichkeiten für die Schichten. Es gibt eine theoretisch optimale Versuchszahl [LO89], die aber unbekannte Größen enthält, so daß man sie bei der Experimentplanung nicht direkt benutzen kann.

Das sog. *transition splitting* ist eine Variante des *stratified sampling*, bei der in jedem Simulationslauf zuerst die Aufwärts-Transitionen simuliert werden bis das System voll ist und dann die Abwärts-Transitionen zurück zum leeren System. Bei dieser Anwendung müssen jedoch vorher die Schichtwahrscheinlichkeiten berechnet werden, und das ist im Beispiel $M/M/1/N$ anspruchsvoller als die Bestimmung der Zustandswahrscheinlichkeiten.

7.4.2 Wesentliche Stichproben (*importance sampling*)

Auch bei dieser Methode werden Stichproben proportional zu ihrer relativen Wichtigkeit bezüglich des Ergebnisses erzeugt.

Diese Methoden wurden zur effizienteren Monte-Carlo-Integration eingeführt [Kle74], [BFS87], [KM53] und haben ihre Anwendung insbesondere in der Zuverlässigkeitsanalyse und Nachrichtenverkehrstheorie. Bei der simulativen Untersuchung eines Systems, im Gegensatz zu Messungen an einem realen System, ist es besonders einfach, die Einflußgrößen eines Systems zu verändern und damit seltene Ereignisse zu provozieren.

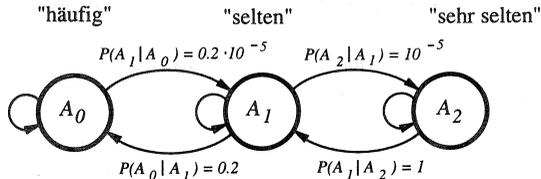
Innerhalb der Methoden des *Importance Sampling* gibt es sehr unterschiedliche Ansätze, die z. B. auf inhärenten Abhängigkeiten der zu untersuchenden Sequenz von Realisierungen für die Zielgröße beruhen.

Eine bekannte Methode bei der Bestimmung von Verlustwahrscheinlichkeiten beruht darauf, daß die Aufwärts-Übergangswahrscheinlichkeiten bzw. -Raten relativ zu den Abwärts-Übergangswahrscheinlichkeiten bzw. -Raten erhöht werden. Damit wird die Anzahl der beobachteten Verluste erhöht. Die ursprünglichen Verlustwahrscheinlichkeiten müssen mit Hilfe einer Gewichtsfunktion, die dem Quotienten der ursprünglichen Dichte und der modifizierten Dichte entspricht, umgerechnet werden. Die Methode ist sehr sensitiv bzgl. der Veränderung der zugrundeliegenden Verteilungen. Optimale Anpassungen existieren nur für spezielle Systemklassen, z. B. für das $M/M/1/N$ System, wo man die umgekehrte Zuordnung trifft, d. h. man vertauscht Ankunftsrate $\lambda^* = \mu$ und Bedienrate $\mu^* = \lambda$, so daß das Verkehrsangebot η zu $\eta^* = 1/\eta$ wird. Aufgrund der Symmetrie des Systems entspricht dann die Wahrscheinlichkeit für das leere System im neuen System der Verlustwahrscheinlichkeit im Originalsystem. Nach [Fra90] ist der Umrechnungsfaktor für dieses System η^{N+1} . Dies läßt sich mit den Formeln aus Abschnitt 3.4.4 verifizieren.

Diese Methode ist Teil eines Vergleichs in [AHH96]. Weitere Arbeiten, z. B. [Sch91a] und [Fra90], untersuchen die Techniken insbesondere im Hinblick auf die Anwendbarkeit in Kommunikationssystemen, wobei sich zeigt, daß die Anwendung der Methode nur bei sehr detaillierter Kenntnis des Systems erfolgversprechend ist.

7.4.3 Wesentliche Partitionierung (*importance splitting*)

Eine andere Methode mehr Stichproben in sog. *important regions* zu erreichen ist das *importance splitting* bei dem die zugrundeliegenden Ankunfts- und Bedienprozesse nicht verändert werden. Das System wird an einem oder mehreren Zwischenzuständen, die meist den Füllstand des Systems betreffen wieder aufgesetzt, da man erwarten kann, daß z. B. der Verlustfall ausgehend von einem höheren Füllstand schneller erreicht wird.



Zustände A_0 , A_1 und A_2 ,
die häufig, selten und sehr selten auftreten: $P(A_1) = 10^{-5}$, $P(A_2) = 10^{-10}$.

Abbildung 7.2: Beispiel einer Markoff-Kette mit drei Zuständen

Unter anderem gibt es hier das Konzept des *preferred sampling* in den sog. *important regions* einer Markoff-Sequenz [HK79], welches sich in seiner einfachsten Form mit drei Ereignissen folgendermaßen beschreiben läßt: Ein *sehr seltenes Ereignis* A_2 wird über ein *seltenes Ereignis* A_1 erreicht, welches wiederum von einem *häufigen Ereignis* A_0 aus erreicht wird, siehe Abbildung 7.2. Die Ereignisse werden den Zuständen eines Geburt- und Sterbeprozesses, der durch eine Markoff-Kette mit 3 Zuständen A_0 , A_1 und A_2 beschreibbar ist, zugeordnet, für deren stationäre Zustandswahrscheinlichkeiten gilt:

$$P(A_2) \ll P(A_1) \ll 1 \quad \text{und} \quad P(A_0) = 1 - P(A_1) - P(A_2) \quad (7.10)$$

Tabelle 7.1: Übersicht: Methoden zur Simulationsbeschleunigung

Methoden	Basis & Anwendung	Eignung für seltene Ereignisse
Gemeinsame Zufallszahlen (<i>common random numbers</i>) (andere Bezeichnungen: <i>correlated sampling</i> , <i>matched streams</i> , <i>matched pairs</i>)	Ausnutzung positiver Korrelation durch gleiche Zufallsströme, Synchronisation notwendig, anwendbar beim Vergleich zweier Systeme, Vorhersage i. allg. nicht möglich	–
Antithetische Stichproben (<i>antithetic sampling</i>)	Ausnutzung negativer Korrelation durch gleiche aber entgegengesetzte (antithetische) Zufallsströme, Synchronisation notwendig, anwendbar bei der Bewertung eines Laufs, Vorhersage i. allg. nicht möglich	–
Steuervariablen (<i>control variables</i>)	Ausnutzung negativer oder positiver Korrelation durch Einführung von Steuervariablen, anwendbar bei der Bewertung eines Laufs, Vorhersage i. allg. nicht möglich	–
Indirekte Schätzung (<i>indirect estimation</i>)	Ausnutzung theoretischer Beziehungen der Zielgröße mit anderen Größen des Systems, anwendbar bei der Bewertung eines Laufs	–
Bedingte Erwartung (<i>conditional expectation</i>)	Ausnutzung theoretischer Beziehungen der bedingten Zielgröße, sehr genaue Analyse des Modells notwendig, anwendbar bei der Bewertung eines Laufs	ja
Geschichtete Stichproben (<i>stratified sampling</i>)	Einteilung der Grundgesamtheit in Schichten	ja
Wesentliche Stichproben (<i>importance sampling</i>)	Änderung der Prozesse, z. B. Ankunfts- und Bedienrate, Vorhersage nicht immer möglich	ja
Wesentliche Partitionierung (<i>importance splitting</i>)	Provozieren von (seltenen) Ereignissen, Manipulation des Systems, Vorhersage z. T. approximativ möglich	ja

Wenn zusätzlich für die Übergangswahrscheinlichkeit $P(A_2|A_1)$ für den Übergang $A_1 \rightarrow A_2$ folgende Relation erfüllt ist:

$$P(A_2) \ll P(A_2|A_1) \quad (7.11)$$

und wenn $P(A_1|A_2) = 1$ gilt, dann kann man die sehr kleine Wahrscheinlichkeit $P(A_2)$ des seltenen Ereignisses A_2 durch eine Beschleunigung der Simulation auf folgende Art und Weise gewinnen:

- Man beginnt damit, daß man in einem ersten Schritt die Wahrscheinlichkeit $P(A_1)$ des seltenen Ereignisses A_1 bestimmt.
- Nachdem der Wert für $P(A_1)$ mit genügender Genauigkeit bekannt ist, bestimmt man die bedingte Wahrscheinlichkeit $P(A_2|A_1)$ des bedingten Ereignisses $A_2|A_1$ durch wiederholte Simulationsläufe, die im Zustand A_1 starten.
- Nachdem auch der Wert von $P(A_2|A_1)$ mit genügender Genauigkeit bestimmt wurde, erhält man die Zielgröße $P(A_2)$ als Produkt

$$P(A_2) = P(A_2|A_1)P(A_1) \quad (7.12)$$

wobei der resultierende Fehler für $P(A_2)$ gesondert bestimmt werden muß.

Diese Art des *importance sampling* wurde 1970 von A.J. BAYES [Bay70, HK79] vorgeschlagen und wurde nach 1990 im Kontext von ATM-Leistungsuntersuchungen unter dem Namen RESTART¹ [VAVA91, VAMMGFC94, VAVA94] weiterentwickelt.

Die bei weitem allgemeinste Simulationsmethode ist die direkte Simulation, da sie auf einer einfachen Übertragung des physikalischen Systemmodells beruht. Von den vorgestellten Methoden zur Behandlung seltener Ereignisse sind die Methoden des *importance splitting* die Methoden, die am nächsten am Originalmodell bleiben. Das Modell wird nicht verändert, es werden gültige Zustände des Modells benutzt, um die seltenen Ereignisse häufiger zu erzeugen und es werden Ergebnisse der Simulation benutzt, um durch Gewichtung auf die Wahrscheinlichkeit des seltenen Ereignisses zurückzuschließen.

Im folgenden Kapitel wird gezeigt, wie das Verfahren in Kombination mit dem LRE-Algorithmus vorteilhaft eingesetzt werden kann.

¹REpetitive Simulation Trials After Reaching Threshold

Simulation seltener Ereignisse mit dem RESTART/LRE-Verfahren

Betrachtet man das Referenzsystem M/M/1/N-FIFO aus Abschnitt 3.4.4 und die Fehlerformeln des LRE-Algorithmus (5.23), so kann man die benötigte Versuchszahl bei vorgegebenem maximalem Fehler d_{max} berechnen. Für eine realistische bzw. eher niedrig angenommene Versuchsrate von z. B. 10^4 Versuchen/sec und einem maximalen Fehler von z. B. 5 – 10 % ergeben sich Simulationslaufzeiten für Systeme mit Verlustwahrscheinlichkeiten in der Größenordnung 10^{-9} im Bereich von einigen Monaten bis Jahren, siehe Tabelle 8.1 und Abbildung 8.1.

Dies ist die typische Situation, in der eine direkte Simulation nicht mehr durchführbar ist und zweckmäßig durch eine Simulationstechnik basierend auf dem RESTART-Prinzip ersetzt werden sollte. Im folgenden wird dieses Prinzip mit den Berechnungsmethoden und der Fehlersteuerung durch den LRE-Algorithmus III kombiniert. Zuerst wird das Prinzip allgemein für Warteräume vom Typ G/G/1/N-FIFO erläutert. Danach werden Formeln zur Bestimmung der minimalen Versuchszahl und der optimalen Stufenzahl des Verfahrens entwickelt. Am Ende dieses Kapitels wird an konkreten Beispielen gezeigt, welche Verbesserungen durch das RESTART/LRE-Verfahren möglich sind. Der darauffolgende Abschnitt beschreibt die Erweiterung des Verfahrens für Warternetze. Im letzten Abschnitt wird das Verfahren auf Modelle für ATM-Netze angewandt.

8.1 Das RESTART/LRE-Prinzip

Das Prinzip des Verfahrens wird im folgenden am Beispiel der Ankunftsbelegung α des Systems G/G/1/N-FIFO erklärt. Der Zustand des Systems bei der Ankunft eines Auftrags wird durch die Belegung, die Bediendauern der wartenden Aufträge, sowie der Restbedienzeit des Auftrags in der Bedienstation vollständig beschrieben. Für korrelierte Ankünfte muß zusätzlich der Zustand des Ankunfts-Generators hinzugenommen werden. Der Zustandsraum ist somit im allgemeinen mehrdimensional. Durch diese Zustandsbeschreibung erhält man nach dem Prinzip der Zusatzvariablen einen Markoff-Prozeß, siehe Abschnitt 3.4.3.

Der hier untersuchte Prozeß der Ankunftsbelegung $\alpha(t)$ unterscheidet sich im allgemeinen von der Belegung $\beta(t)$, da Zustandsänderungen nur bei Ankünften erfol-

Tabelle 8.1: Direkte Simulation: Versuchszahlen für M/M/1/N

Benötigte Rechenzeit in Abhängigkeit von d_{max}
und einer angenommenen Versuchsrate von 10^4 Versuche/sec

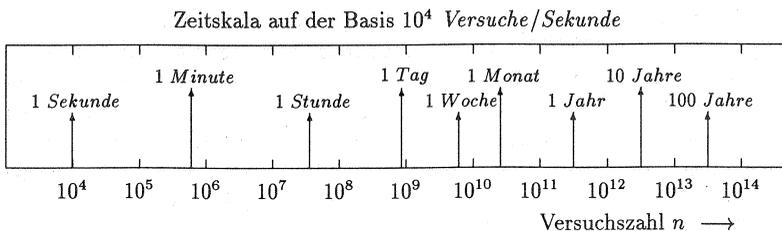
	M/M/1/34			M/M/1/85		
$B = N + 1$	35			86		
η	0.8					
$P_B = G_B$	8.1×10^{-5}			9.3×10^{-10}		
d_{max}	0.1	0.05	0.01	0.1	0.05	0.01
n	3.2×10^6	1.3×10^7	32×10^7	2.8×10^{11}	1.1×10^{12}	28×10^{12}
Rechenzeit	5.3 Min.	21.3 Min.	9 Std.	324 Tage	3.56 Jahre	89 Jahre

gen. Die Zustandsänderungen aufgrund von Abgängen erfolgen zeitverzögert bei der nächsten Ankunft. Betrachtet man die Belegung $\beta(t)$ bzw. die Ankunftsbelegung $\alpha(t)$, so liegt im allgemeinen ein Semi-Markoff-Prozess vor, da die Aufenthaltsdauer in einem der Zustände beliebig verteilt sein kann.

In Abschnitt 7.4.3 wurde eine erste Erklärung für die Arbeitsweise des RESTART-Prinzips gegeben. Abbildung 8.2 zeigt eine andere Darstellung des Prinzips bezogen auf die Bestimmung der Verlustwahrscheinlichkeit über die Messung der Ankunftsbelegung α . Die Ankunftsbelegung ist in diesen Darstellungen vereinfacht als kontinuierliche Größe dargestellt.

Dargestellt ist ein typischer Verlauf der Ankunftsbelegung α über der Zeit, wobei vom leeren Warteraum ausgegangen wird. In einem ersten Schritt wird die Wahrscheinlichkeitsfunktion für die Ankunftsbelegungen α im Bereich $0, 1, \dots, I$ mit ausreichender Genauigkeit bestimmt, wobei jeder Wert in die Messung eingeht.

Im anschließenden zweiten Schritt werden nun die *langen* Abschnitte, die sich nur

Abbildung 8.1: Simulationslaufzeit als Funktion der Versuchszahl n

unterhalb der Stufe I bewegen, nicht mehr ausgewertet bzw. ganz unterdrückt, da nur die Aufteilung der Wahrscheinlichkeiten für Belegungen oberhalb von I interessieren. Wie in Abbildung 8.2 gezeigt, kann man die Teilschritte, die jeweils mit der Ankunftsbelegung I starten, gemeinsam darstellen. Da der Zustand des Systems für die Ankunftsbelegung I nicht allein durch die Anzahl von Anforderungen im Warteraum beschrieben wird, sondern durch den Zustand des Generators und des Servers, müssen zur Durchführung des Verfahrens diese Zustände modellgerecht erzeugt werden. Dies erreicht man in der Simulationsdurchführung dadurch, daß man im ersten Schritt diese Zustände abspeichert und dann daraus im zweiten Schritt zufällig einen auswählt. Die Ergebnisse des zweiten Schritts müssen dann mit Hilfe der Ergebnisse des ersten Schritts auf die Grundgesamtheit zurückgerechnet werden:

$$P\{\alpha \geq j\} = P\{\alpha \geq j | \alpha \geq I\} P\{\alpha \geq I\} \quad \text{für } j = I + 1, \dots, B \quad (8.1)$$

Dieses Prinzip kann in natürlicher Weise auf mehrere Stufen I_0, I_1, \dots, I_{m-1} erweitert werden. Die Verlustwahrscheinlichkeit errechnet sich über alle Stufen entsprechend folgender Formel:

$$P_B = P\{\alpha \geq B\} = \quad (8.2)$$

$$P\{\alpha \geq B | \alpha \geq I_{m-1}\} P\{\alpha \geq I_{m-1} | \alpha \geq I_{m-2}\} \cdots P\{\alpha \geq I_1 | \alpha \geq I_0\} P\{\alpha \geq I_0\}$$

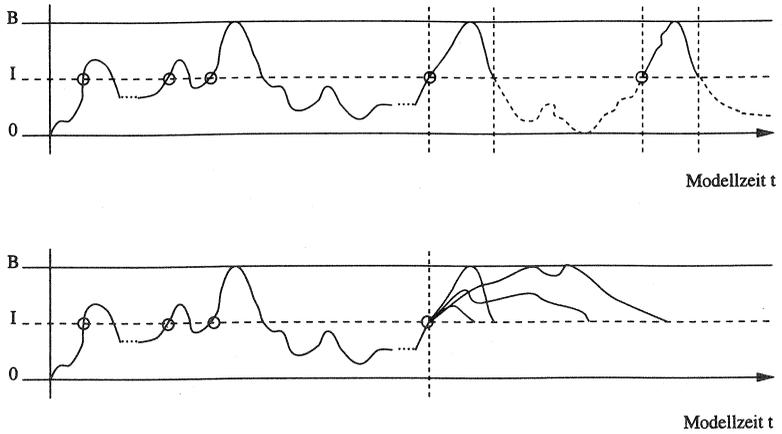
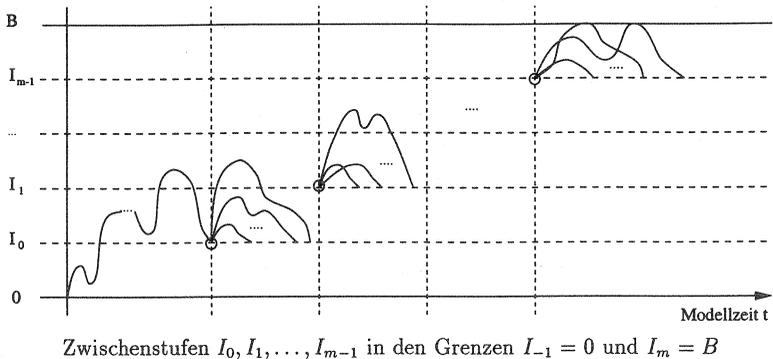
Abbildung 8.3 zeigt diese Erweiterung mit m Zwischenstufen, wobei in den Schritten 1 bis m jeweils die bedingte komplementäre Verteilungsfunktion bestimmt wird.

Prinzipiell läßt sich das Verfahren in der hier vorgestellten Art auf alle Zufallsvariablen anwenden, die einen abzählbaren, endlichen Wertebereich haben und in denen die seltenen Zustände nur über Zustände erreichbar sind, die häufiger auftreten.

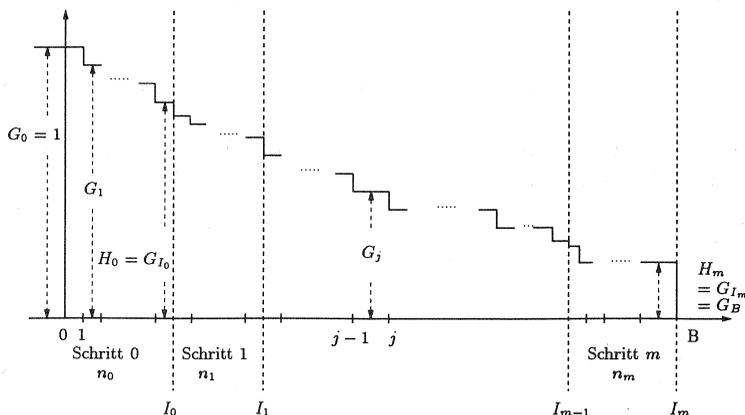
In den folgenden Abschnitten wird beschrieben, wie sich mit Hilfe des LRE-Algorithmus die minimale Versuchszahl und die optimalen Stufen in Abhängigkeit von der Stufenzahl m bestimmen lassen. Ausgehend von diesen Formeln wird approximativ die optimale Stufenzahl m^* bestimmt.

8.2 Die minimale Versuchszahl für m Stufen

Zur Herleitung der minimalen Versuchszahl und der optimalen Stufenzahl sind einige Festlegungen für das RESTART/LRE-Verfahren notwendig. Die Anzahl der Stufen wird mit m bezeichnet, d. h. die direkte Simulation ohne RESTART erhält man für $m = 0$. Mit $m = 1$ erhält man das einstufige RESTART/LRE-Verfahren mit einem Stützpunkt $I_0 = I$, siehe [SG94]. Allgemein benötigt man m Stützpunkte $I_i, i = 0, \dots, m - 1$ für $m = 1, 2, 3, \dots$ mit Ergänzung der Randpunkte $I_{-1} = 0$ und

Abbildung 8.2: Ankunftsbelegung $\alpha(t)$ über der Modellzeit t mit einer ZwischenstufeAbbildung 8.3: Ankunftsbelegung $\alpha(t)$ über der Modellzeit t mit m Zwischenstufen

$I_m = B$. Es wird für jede Stufe die bedingte komplementäre Verteilungsfunktion¹ $P\{\alpha \geq I_j | \alpha \geq I_{j-1}\}$ bestimmt.



Aufteilung der KVF $G(x)$ in $m + 1$ Intervalle:
 m Anzahl der RESTARTs (0: kein RESTART, 1: einstufiges RESTART)
 m -stufiges RESTART: $m + 1$ Läufe mit n_i Versuchen für $i = 0, \dots, m$
 Stufe i : Niveaus der KVF bis $H_i = G_{I_i}$ für $i = 0, \dots, m$
 mit $I_{-1} = 0, I_m = B, H_{-1} = 1, H_m = G_B, \gamma_i = cf_{I_i}$
 Gesamtfehler: $d_B = d_{max}$, Fehler in einer Stufe: d_S

Abbildung 8.4: RESTART/LRE-Verfahren in m Stufen

Zu jedem Punkt gehören die entsprechenden Werte bzw. Niveaus der komplementären Verteilungsfunktion $H_i = G_{I_i} = P\{\alpha \geq I_i\}$ für $i = 0, \dots, m$ und ergänzend gilt $H_{-1} = 1, H_m = G_{I_m} = G_B$. Der Korrelationsfaktor nach Gleichung (5.21) bezogen auf die Stützpunkte wird mit $\gamma_i = cf_{I_i}$ bezeichnet. Abbildung 8.4 gibt einen Überblick über die Aufteilung der Simulation in m Stufen. Im Intervall i wird ausgehend vom Stützpunkt I_{i-1} bzw. Niveau $H_{i-1} = G_{I_{i-1}}$ der Bereich bis zum Stützpunkt I_i bzw. Niveau $H_i = G_{I_i}$ für $i = 0, \dots, m$ bestimmt.

Als maximaler Gesamtfehler ist $d_B = d_{max}$ vorgegeben. Aus diesem maximalen Gesamtfehler muß die Fehlerschranke d_S des Fehlers eines einzelnen Intervalls bestimmt werden. Aus Gleichung (5.23) kann die Versuchszahl n_i im Intervall i durch Substitution des entsprechenden Korrelationsfaktors $cf_k := \gamma_i$, des Endniveaus $G_{min} := H_{i|\alpha \geq I_{i-1}}$ und des LRE-Fehlers $d_{max} := d_S$, d. h. es wird in jedem Intervall die gleiche Fehlerschranke d_S angenommen, abgeleitet werden:

$$n_i = \frac{\gamma_i}{d_S^2 H_{i|\alpha \geq I_{i-1}}} \quad \text{für } i = 0, \dots, m \tag{8.3}$$

¹Hier wird statt $G(x) = P\{\alpha > x\}$ meist der Einfachheit halber $G_i = P\{X \geq i\}$ verwendet.

Durch Multiplikation der Wahrscheinlichkeit des Niveaus i unter der Bedingung, daß alle Belegungen größer oder gleich dem entsprechenden Stützpunkt I_{i-1} sind, mit dem vorhergehenden Niveau $H_{i-1} = G_{I_{i-1}} = P\{\alpha \geq I_{i-1}\}$ erhält man das Niveau H_i :

$$H_i = H_{i|\alpha \geq I_{i-1}} \cdot H_{i-1} \quad \text{für } i = 0, \dots, m \quad (8.4)$$

Entsprechendes gilt für die Werte der komplementären Verteilungsfunktion G_j im Bereich $j = I_{i-1} + 1, \dots, I_i$:

$$G_j = G_{j|\alpha \geq I_{i-1}} \cdot H_{i-1} \quad (8.5)$$

Die Gesamtzahl n der Versuche bestimmt sich aus

$$n = \sum_{i=0}^m n_i \quad (8.6)$$

und mit Gleichung (8.3) und (8.4) kann n als Funktion der Niveaus H_i dargestellt werden:

$$n(H_0, H_1, H_2, \dots, H_m) = \frac{1}{d_2^2} \sum_{i=0}^m \frac{\gamma_i H_{i-1}}{H_i} \quad (8.7)$$

Durch sukzessive Anwendung der notwendigen Bedingung für ein Minimum auf Gleichung (8.7) werden die optimalen H_i^* -Niveaus und die minimale Gesamtzahl der Versuche in Abhängigkeit von $H_m = G_B$ bestimmt, siehe Anhang B.1 und B.2 für eine detaillierte Beschreibung der Induktionsbeweise:

$$H_i^*(H_m) = \left(\frac{\prod_{j=0}^i \gamma_j^{m-i}}{\prod_{j=i+1}^m \gamma_j^{i+1}} H_m^{i+1} \right)^{\frac{1}{m+1}} \quad (8.8)$$

$$n_{\min}(H_m) = \frac{m+1}{d_2^2} \left(\prod_{i=0}^m \gamma_i \right)^{\frac{1}{m+1}} \left(\frac{1}{H_m} \right)^{\frac{1}{m+1}} \quad (8.9)$$

Zum vollständigen Nachweis des Minimums muß zusätzlich nachgewiesen werden, daß die zweite partielle Ableitung an dieser Stelle positiv ist, siehe Anhang B.2:

$$\left. \frac{\partial^2 n}{\partial H_i \partial H_i} \right|_{H_0=H_0^*, H_1=H_1^*, \dots, H_{m-1}=H_{m-1}^*} > 0 \quad (8.10)$$

und daß folgende Matrix positiv definit ist [BS71, Zur64]:

$$\left[\frac{\partial^2 n}{\partial H_i \partial H_j} \Big|_{H_0=H_0^*, H_1=H_1^*, \dots, H_{m-1}=H_{m-1}^*} \right] \quad (8.11)$$

Dieser Nachweis wird für den in Abschnitt 8.3 betrachteten Fall in Anhang B.2 durchgeführt.

Der Fehler in m Stufen ($m+1$ Intervallschritte) beträgt, siehe Anhang B.4:

$$d_B^2 = (d_S^2 + 1)^{m+1} - 1 \rightarrow d_S^2 = (d_B^2 + 1)^{\frac{1}{m+1}} - 1 \quad (8.12)$$

8.3 Die näherungsweise optimale Stufenzahl

Der Versuch aus (8.9) und (8.12) die optimale Stufenzahl m_{opt} zu bestimmen, führt auf eine transzendente Gleichung, die nicht aufgelöst werden kann. Hier führt eine Approximation des folgenden Produkts zum Ziel:

$$\left(\prod_{i=0}^m \gamma_i \right)^{\frac{1}{m+1}} \approx (\gamma^m \gamma_m)^{\frac{1}{m+1}} \quad (8.13)$$

Die Begründung für diese Approximation leitet sich aus der Beobachtung ab, daß die Werte von γ_i analog zu den Werten von ϱ_i eine Maximumkurve bilden, siehe Abschnitt 3.4.4, so daß es sinnvoll ist, alle γ_i außer γ_m durch eine Konstante $\gamma = cf_M$, z. B. γ_{max} zu approximieren:

$$\gamma_0 \approx \gamma_1 \approx \gamma_2 \approx \dots \approx \gamma_{m-1} \approx \gamma = cf_M \quad (8.14)$$

Die daraus abgeleitete Approximation n_z^* der Versuchsanzahl beträgt abhängig vom maximalen Gesamtfehler d_B , der Stufenzahl m und der Verlustwahrscheinlichkeit $H_m = G_B$:

$$n_z^*(H_m) = \frac{(m+1) \left(\frac{\gamma^m \gamma_m}{H_m} \right)^{\frac{1}{m+1}}}{(d_B^2 + 1)^{\frac{1}{m+1}} - 1} \quad (8.15)$$

Die Güte der Approximation kann für ausgewählte Beispiele durch Vergleich mit der exakten Formel (8.7) überprüft werden. Für die optimalen Niveaus H_i^* ergibt sich approximativ:

$$H_i^*(H_m) \approx \left(\frac{\gamma}{\gamma_m} H_m \right)^{\frac{i+1}{m+1}} \quad (8.16)$$

und damit

$$H_i^*(G_B) \approx \left(\frac{\gamma}{\gamma_m} G_B \right)^{\frac{i+1}{m+1}} \quad (8.17)$$

Die Approximation der optimalen H-Niveaus der Stufe sind somit nur abhängig von G_B, m und i sowie dem Verhältnis von γ zu γ_m .

Die Stützpunkte I_i sind durch folgende Gleichung bestimmt:

$$H_i^*(G_B) = G_{I_i} \quad (8.18)$$

Zur weiteren Bestimmung kann die Approximation aus Formel (8.17) für H_i^* verwendet werden:

$$H_i^*(G_B) \approx \left(\frac{\gamma}{\gamma_m} G_B \right)^{\frac{i+1}{m+1}} = G_{I_i} \quad (8.19)$$

Mit einer weiteren Approximation des Fehlers mit Hilfe einer Reihenentwicklung:

$$d_S^2 = (d_B^2 + 1)^{\frac{1}{m+1}} - 1 \approx \frac{d_B^2}{m+1} \quad (8.20)$$

erhält man die Versuchszahl n^* als Funktion der Stufenzahl m :

$$n^*(m) = \frac{(m+1)^2 \left(\frac{\gamma^m \gamma_m}{H_m} \right)^{\frac{1}{m+1}}}{d_B^2} \quad (8.21)$$

Der absolute Fehler der Approximation in Gleichung (8.20) kann aufgrund der Reihenentwicklung durch

$$\frac{d_B^4}{(m+1)(1-d_B^4)} \quad (8.22)$$

nach oben abgeschätzt werden, wobei jeweils aufgrund des wechselnden Vorzeichens zwei aufeinanderfolgende Summanden zusammengefaßt und durch $d_B^{4k}/(m+1)$ für $k = 1 \dots \infty$ abgeschätzt werden.

Die notwendige Bedingung ($dn^*/dm = 0$) führt zur Lösung für die optimale Anzahl der Stufen m_{opt} , bei der n^* minimal wird.

$$m_{opt} = \frac{1}{2} \ln \left(\frac{\gamma_m}{\gamma H_m} \right) - 1 \quad (8.23)$$

Das Minimum kann in diesem Fall mit Hilfe der 2. Ableitung nachgewiesen werden, die an der Stelle m_{opt} auf einen positiven Wert führt.

8.3.1 Ergebnisformeln

Als Ergebnis für die optimale Zahl von Versuchen $n^*(m)$ nach Gleichung (8.21) erhält man in Abhängigkeit von m somit näherungsweise folgende Ergebnisformel mit $\gamma = cf_M, \gamma_m = cf_B, H_m = G_B$:

$$n^*(m) = \frac{(m+1)^2}{d_B^2} \left(\frac{cf_M^m cf_B}{G_B} \right)^{\frac{1}{m+1}} \quad (8.24)$$

Die optimale Zahl der Stufen bezogen auf die in Gleichung (8.24) gegebene Versuchszahl ist:

$$m_{opt} = \frac{1}{2} \ln \left(\frac{cf_B}{cf_M G_B} \right) - 1 \quad (8.25)$$

Da die Stufenzahl ganzzahlig sein muß, wird der Wert von m_{opt} gerundet. Da m zwischen 0 und maximal $B-1$ liegen muß, muß für den Fall, daß die Stufenzahl für das berechnete Minimum links von 0 oder rechts von $B-1$ liegt, dies durch einen Vergleich mit 0 bzw. $B-1$ ausgeschlossen werden:

$$m^* = \begin{cases} 0 & \text{für } m_{opt} < 0 \\ \text{round}(m_{opt}) & \text{für } 0 \leq m_{opt} \leq B-1 \\ B-1 & \text{für } m_{opt} > B-1 \end{cases} \quad (8.26)$$

Eine untere bzw. obere Schranke für die Versuchszahl erhält man, wenn man für γ den Minimalwert γ_{min} bzw. den Maximalwert γ_{max} in Gleichung (8.15) bzw. (8.21) einsetzt. Mit Gleichung (8.17) und (8.25) kann man zeigen, daß die bedingten Wahrscheinlichkeiten $P_i = H_{i|\alpha \geq I_{i-1}}$ sich im optimalen Fall zu $P_i = e^{-2}$ bestimmen lassen. Dies entspricht den Ergebnissen in [VAVA94] und kann zur automatischen Bestimmung der optimalen Stufen benutzt werden.

Wenn man den typischen Verlauf einer Maximumkurve für ϱ_i und cf_i bzw. $\gamma_j = cf_j$, nach Abbildung 3.10 annimmt, dann erhält man mit ϱ_{max} bzw. γ_{max} bereits eine gute Approximation. Im folgenden wird das Verfahren mit Hilfe des Referenzsystems M/M/1/N validiert.

8.3.2 Beispiel M/M/1/N

In Abschnitt 3.4.4 wurden die grundlegenden Formeln für das Wartesystem M/M/1/N zusammengestellt. Aufgrund der Markoff-Eigenschaft des Poisson-Prozesses gelten die dort angegebenen Formeln für die Belegung β auch für die Ankunftsbelegung α . Wendet man die in diesem Abschnitt eingeführte Bezeichnungsweise auf Gleichung (3.54) und Gleichung (3.58) an, so erhält man folgende Gleichungen für die komplementäre Verteilungsfunktion G_i und den lokalen Korrelationskoeffizienten ϱ_i :

$$\left. \begin{aligned} G_i(\eta) &= \frac{\eta^i - \eta^{B+1}}{1 - \eta^{B+1}} \quad \text{für } i = 1, 2, \dots, B \\ G_0 &= 1 \end{aligned} \right\} \quad (8.27)$$

$$\varrho_i = 1 - \frac{(1 - \eta)(1 - \eta^{B+1})\eta^i}{(1 + \eta)(1 - \eta^i)(\eta^i - \eta^{B+1})} \quad \text{für } i = 1, 2, \dots, B \quad (8.28)$$

$$\varrho_{max} \approx \frac{2\eta}{1 + \eta} \quad (8.29)$$

$$\varrho_{min} \approx \frac{\eta}{1 + \eta} \quad (8.30)$$

$$\gamma_{max} \approx \frac{1 + \varrho_{max}}{1 - \varrho_{max}} = \frac{1 + 3\eta}{1 - \eta} \quad (8.31)$$

$$\gamma_{min} \approx \frac{1 + \varrho_{min}}{1 - \varrho_{min}} = 1 + 2\eta \quad (8.32)$$

Aus Gleichung (8.19) ergibt sich für die Stützstellen approximativ:

$$I_i = \text{round} \left(\frac{\ln(H_i^*(1 - \eta^{B+1}) + \eta^{B+1})}{\ln(\eta)} \right) \quad (8.33)$$

Betrachtet man die Verlustwahrscheinlichkeit $G_B = 10^{-z}$, so kann man eine obere Grenze für die optimale Stufenzahl angeben, die etwa der Größenordnung von z entspricht, mit $cf_M \approx \gamma_{max}$ und $cf_B \approx \gamma_{min}$ und der Abschätzung $0 \leq \frac{cf_B}{cf_M} \leq 1$ folgt nach Gleichung (8.31), (8.32) und (8.25):

$$m_{opt} \leq 1.15z - 1 \quad (8.34)$$

Die Anzahl der Stufen ist somit klein gegenüber der Anzahl der Simulationsversuche, so daß der zusätzliche Aufwand, der in der Implementierung zur Initialisierung einer neuen Stufe notwendig ist, vernachlässigbar ist.

Mit Hilfe des Referenzsystems $M/M/1/N$ und der Approximationsformel (8.15) kann angegeben werden, welche Größenordnung der Wahrscheinlichkeit $P_B = G_B = 10^{-z}$ bei einem festen maximalen Fehler d_B innerhalb einer vorgegebenen Laufzeit mittels mehrstufiger Simulation bestimmt werden kann. Bei einer angenommenen realistischen Versuchsrate von 10^4 *Versuche/sec* und einer Belastung von $\eta = 0.8$ kann für $d_B = 0.01$ innerhalb eines Tages bis zur Größenordnung 10^{-23} simuliert werden. Läßt man einen Fehler von $d_B = 0.1$ zu, so kann man innerhalb eines Tages sogar bis zur Größenordnung 10^{-230} simulieren. Für einen festen Wert von η muß das Verhältnis z/d_B konstant gehalten werden:

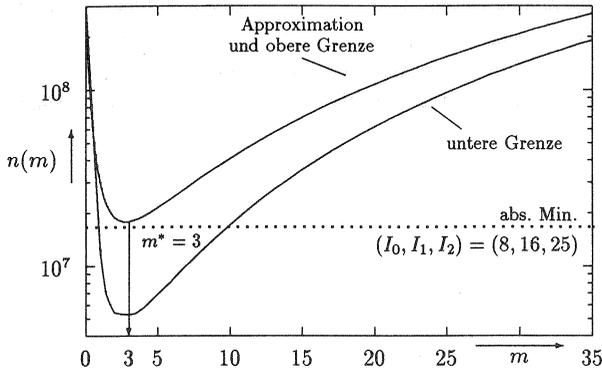
$$n^* \leq 10 \cdot \frac{1 + 3\eta}{1 - \eta} \cdot \left(\frac{z}{d_B}\right)^2 \quad (8.35)$$

Die berechnete komplementäre Verteilungsfunktion $G(x)$, siehe Abbildung 3.10, des Referenzsystems $M/M/1/N$ soll im folgenden an zwei Beispielen durch Simulation nachgewiesen werden. Das erste Beispiel wird so gewählt, daß auch eine direkte Simulation durchführbar ist. Mit $N = 34$ und $\eta = 0.8$ ergibt sich eine Verlustwahrscheinlichkeit von $P_B = G_B = 8.1 \times 10^{-5}$, siehe Tabelle 8.1. Durch Vorgabe eines maximalen Fehlers von 10 % in Gleichung (5.23), d. h. $d_{max} = 0.1$, ergibt sich die Versuchsanzahl von $n = 3.2 \times 10^6$. Nimmt man eine Versuchsrate von 10^4 *Versuche/sec* an, dann ergibt sich eine Simulationslaufzeit von $T = 320$ *sec* = 5.3 *Min*. Durch eine Vergrößerung des Warteraums von $N = 34$ auf $N = 85$ erreicht man für $\eta = 0.8$ eine Verlustwahrscheinlichkeit von $P_B = P_0 \eta^{86} = 9.26 \times 10^{-10}$. Der Nachweis durch Simulation mit gleichem maximalen Fehler $d_{max} = 0.1$ würde hier $n = 2.8 \times 10^{11}$ Versuche erfordern und die unrealistische Laufzeit von $T = 2.8 \times 10^8$ *sec* = 324 *Tage* ergeben. Die Forderung nach einem auf 5 % reduzierten Fehler $d_B = 0.05$ würde die Laufzeit für $N = 34$ auf $T = 21.3$ *Min*. und im Fall $N = 85$ auf $T = 3.56$ *Jahre* erhöhen. In den folgenden Abschnitten wird gezeigt, wie bei diesen Beispielen mit dem RESTART/LRE-Verfahren die Simulationslaufzeit effektiv reduziert werden kann.

Beispiel: M/M/1/34

Mit den Formeln (8.25), (8.26) und $\eta = 0.8$ ergibt sich für dieses Beispiel mit $B = 35$ die optimale Anzahl von Stufen zu $m_{opt} = 2.77$ bzw. $m^* = 3$. Die zugehörige Versuchszahl ist von der Indexfolge und vom vorgegebenen Fehler abhängig.

Abbildung 8.5 zeigt die Approximation für den Verlauf der Versuchszahl über der Stufenzahl m für $d_B = 0.01$. Die Approximation benutzt Formel (8.15) für n_z^* mit dem maximalen Wert für γ nach Formel (8.31). Dies ist gleichzeitig die Maximumkurve, die im vorliegenden Fall einen Wert von 1.79×10^7 ergibt. Als untere Grenze wird hier eine Kurve angegeben, die die Tatsache benutzt, daß die Stützpunkte disjunkt sind und somit aus allen γ_i die m kleinsten ausgewählt werden können. Auch für diese Kurve ergibt sich das Minimum bei $m^* = 3$. Für diesen Fall läßt sich für $m = 3$ auch das absolute Minimum durch eine vollständige Suche bestimmen. Dieses Minimum mit dem Wert 1.664×10^7 wird für die Indexfolge (8, 16, 25) erreicht. Das absolute Gesamtminimum liegt also bei $m = 3$. Die berechnete Indexfolge nach Gleichung (8.33) mit $m^* = 3$ ist (8, 17, 25) bei einer Versuchszahl von 1.665×10^7 liegt nur sehr wenig über dem exakten Minimum. Man erhält im vorliegenden Fall mit der Approximation die optimale Stufenzahl. Wenn man diese Zahlen mit denen in Tabelle 8.1 vergleicht, sieht man, daß sich auch bei Verlustwahrscheinlichkeiten, die noch mit direkter Simulation durchführbar sind, das RESTART/LRE-Verfahren lohnt, da man entweder kürzere Rechenzeiten erhält oder bei gleicher Rechenzeit genauere Ergebnisse, z. B. sind bei $m = 3$ nur 1.66×10^7 Versuche statt 32×10^7 notwendig.



Versuchszahl $n(m)$ zur Anzahl der Stützpunkte m
 Approximation und obere Grenze entspricht der Versuchszahl $n^*(m)$
 Verkehrslast $\eta = 0.8$ und maximaler Fehler $d_{max} = d_B = 0.01$

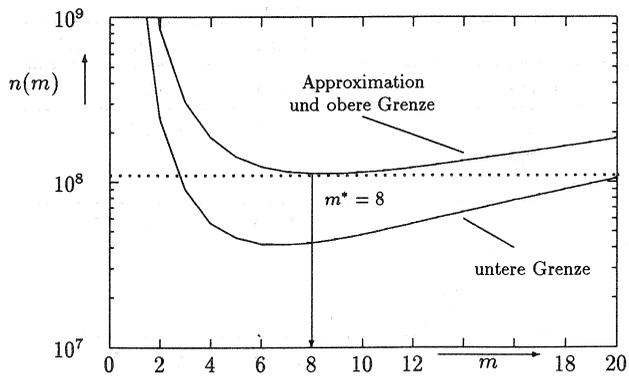
Abbildung 8.5: Wartesystem M/M/1/34: Versuchszahl

Beispiel: M/M/1/85

In Abbildung 8.6 wird das Beispiel: M/M/1/85 mit $B = 86$, $\eta = 0.8$, $d_B = 0.01$ im Ausschnitt $m = 0, \dots, 20$ dargestellt. Der Verlauf läßt sich am besten graphisch verfolgen. Die Simulationsbeschleunigung ist besser an den konkreten Versuchszahlen in Tabelle 8.2 abzulesen. Man sieht am Verlauf der Kurve und an den angegebenen Werten, daß die optimale Stufenzahl unkritisch bzw. robust ist. Für die untere Grenzkurve ergibt sich das Minimum zu $m^* - 1 = 7$. Die optimale Stufenzahl ist $m^* = 8$, aber auch für $m = 7$ und $m = 9$ ergeben sich ähnliche Versuchszahlen.

Tabelle 8.2: Versuchszahlen für M/M/1/85 für $d_B = 0.01$ und $\eta = 0.8$

M/M/1/85, B=86		
	Stufen	Versuchsanzahl
Direkte Simulation	$m = 0$	2.80×10^{13}
RESTART/LRE: Einstufiges Verfahren $I = I_0 = 42$	$m = 1$	88×10^8
RESTART/LRE: m-stufiges Verfahren berechnete approx. Stützstellen 9, 19, 28, 38, 47, 57, 66, 75	$m^* = 8$	1.10×10^8
Minimumabschätzung	$m^* = 8$	0.42×10^8
Maximumabschätzung und Approximation	$m = 7$	1.158×10^8
Maximumabschätzung und Approximation	$m^* = 8$	1.1272×10^8
Maximumabschätzung und Approximation	$m = 9$	1.1277×10^8



Versuchszahl $n(m)$ zur Anzahl der Stützpunkte m
 Approximation und obere Grenze entspricht der Versuchszahl $n^*(m)$
 Verkehrslast $\eta = 0.8$ und maximaler Fehler $d_B = 0.01$

Abbildung 8.6: Wartesystem M/M/1/85: Versuchszahl

Beispiel: M/M/1/21 und M/M/1/50

Tabelle 8.3 vergleicht zwei Referenzsysteme mit Verlustwahrscheinlichkeiten in der Größenordnung 10^{-9} und 10^{-21} . Die Anzahl der benötigten Versuche für die mehrstufigen RESTART/LRE-Simulationen werden mit den Zahlen für die direkte und die einstufige Simulation verglichen, siehe [Fuß97]. Der Vergleich der gemessenen Werte der Verlustwahrscheinlichkeit $\tilde{P}_L = \tilde{G}_B$ und der lokalen Korrelation am Endpunkt $\tilde{\varrho}_B$ mit den analytisch berechneten Werten zeigt die sehr gute Übereinstimmung der simulierten und berechneten Werte.

Tabelle 8.3: Vergleich simulierter und theoretischer Werte für $\eta = 0.4$ und $d_{max} = 0.03$
Die Rechenzeit basiert auf 10^4 Versuche/sec.

	M/M/1/21			M/M/1/50		
	Simulation einstufig $m = 1$	Simulation $m_{opt} = 9$	Berechnung direkt $m = 0$	Simulation $m = 5$	Simulation $m_{opt} = 22$	Berechnung direkt $m = 0$
\tilde{G}_B bzw. G_B	1.091×10^{-9}	1.031×10^{-9}	1.056×10^{-9}	2.98×10^{-21}	3.02×10^{-21}	3.04×10^{-21}
$\tilde{\varrho}_B$ bzw. ϱ_B	0.291	0.280	0.286	0.292	0.256	0.286
\tilde{n} bzw. $n^*(0)$	3.42×10^8	2.70×10^6	18.94×10^{11}	5.14×10^8	1.39×10^7	6.57×10^{23}
$n^*(0)/\tilde{n}$	5.5×10^3	701×10^3	–	1.3×10^{15}	473×10^{15}	–
Rechenzeit	9.5 Std.	4.5 Min.	6 Jahre	14.28 Std.	23.17 Min.	2×10^{12} Jahre

$n^*(0)$ gibt die theoretisch benötigte Versuchsanzahl für eine direkte Simulation ohne RESTART an. Der Beschleunigungsfaktor wird mit $n^*(0)/\tilde{n}$ angegeben. Wie bereits am Anfang dieses Abschnitts erwähnt, ist eine direkte Simulation praktisch nicht durchführbar. Mit der Annahme von 10^4 Versuchen pro Sekunde ergeben sich Laufzeiten von 6 und 2×10^{12} Jahre für die Beispiele in Tabelle 8.3. Mit der Beschleunigung durch das RESTART/LRE-Verfahren wurden die Simulationen alle *innerhalb eines Tages* und durch Verwendung der optimalen Stufenzahl sogar *innerhalb einer halben Stunde* durchgeführt.

Wenn das Verhältnis von N/m klein ist, z. B. $N/m < 3$, dann zeigt sich bei der Simulation, daß der Endwert der lokalen Korrelation in den einzelnen Abschnitten $\tilde{\varrho}_i$ aufgrund der Unterbrechungen niedriger ist als der des ursprünglichen Prozesses, wodurch die Anzahl der Versuche \tilde{n} bei der Durchführung der Simulation reduziert wird. Man sollte also die Stufenzahl aus diesem Grund nicht zu groß wählen im Verhältnis zur maximalen Belegung. Da bei der Simulation in einer Stufe auch Werte der nächsten Stufe anfallen und ausgewertet werden können, könnte man diese Werte übernehmen oder zum Vergleich heranziehen und darüber die Auswahl der Stützstellen steuern. In ähnlicher Weise muß man vorgehen, wenn das Ziel der Simulation darin besteht, die Korrelation des ursprünglichen Systems zu messen. Hier wäre ggf. auch eine Überlappung von Stützstellenbereichen denkbar.

8.4 Einstufiges RESTART/LRE-Verfahren

Den einstufigen RESTART/LRE-Algorithmus erhält man für $m = 1$, d. h. es gibt nur einen Stützpunkt $I_0 = I$. Durch Einsetzen von $m = 1$ in die Formeln (8.8) und (8.9) erhält man die bereits in [SG94] angegebenen Formeln² für das optimale Niveau $H_0^* = G_{I_{opt}}$ und die minimale Versuchsanzahl $n_{min}(G_B)$:

$$H_0^*(H_1) = G_{I_{opt}}(G_B) = \sqrt{\frac{\gamma_0}{\gamma_1} H_1} = \sqrt{\frac{cf_I}{cf_B} G_B} \quad (8.36)$$

$$n_{min}(H_m) = n_{min}(G_B) = \frac{2}{d_S^2} \sqrt{\frac{\gamma_0 \gamma_1}{H_1}} = \frac{2}{d_S^2} \sqrt{\frac{cf_I cf_B}{G_B}} \approx \frac{4}{d_B^2} \sqrt{\frac{cf_I cf_B}{G_B}} \quad (8.37)$$

Für das Referenzsystem M/M/1/N ergibt sich nach (8.33) der optimale Stützpunkt zu:

$$I_{opt} = \frac{\ln(G_{I_{opt}}(1 - \eta^{B+1}) + \eta^{B+1})}{\ln(\eta)} \quad (8.38)$$

8.5 Beschreibung des RESTART/LRE-Algorithmus

Es folgt eine vereinfachte algorithmische Beschreibung des Verfahrens. Zur Durchführung des Algorithmus müssen zusätzlich zur maximalen Belegung $B = N + 1$ und dem maximalen Fehler d_B alternativ folgende Größen vorgegeben werden:

- die Stufenzahl m und die Stützstellen I_j für $j = 0 \dots m - 1$
- G_{min} als Approximation von G_B , so daß die Stützstellen approximativ vom Programm bestimmt werden können

²Man beachte, daß im Vergleich zu [SG94] die Indexzuordnung um eine Position verschoben wurde.

RESTART/LRE-Algorithmus mit m Stufen:**Initialisierung:**

Vorgabe der Parameter des Algorithmus: $B, m, I_0, \dots, I_{m-1}, d_B$;

Vorgabe weiterer Parameter für Eingabe- und Bedienprozeß;

Bestimme max. Schrittfehler d_S aus max. Gesamtfehler d_B nach Gl. (8.20);

Generiere Objekte für Verkehrsquellen, Warteschlangen und Bedienstationen;

Generiere LRE-Auswertungsobjekt mit Bereich $0, \dots, B$, und max. Fehler d_S ;

Setze Umrechnungsfaktor $U := 1$;

Starte Simulation;

Schritt $i = 0$:

while not (Fehler $\leq d_S$ für G_1 bis G_{I_0})

do begin

Setze Simulation fort und bestimme neuen Wert;

Übergebe Wert an LRE-Auswertung;

if Alter Wert = $I_0 - 1$ und neuer Wert = I_0

then Speichere Zustand für RESTART bei I_0 ;

end

Schritt $i = 1, \dots, m$:

for $i = 1$ **to** m

do begin

Initialisierung Stufe i :

Vorbereitung LRE für Wertebereich von I_{i-1} bis B ;

Laden eines zufälligen Zustands " I_{i-1} ";

Stufe i :

while not (Fehler $\leq d_S$ für $G_{I_{i-1}|\alpha \geq I_{i-1}}$ to $G_{I_i|\alpha \geq I_{i-1}}$) **do begin**

Setze Simulation fort und bestimme neuen Wert;

if Alter Wert = $I_i - 1$ und neuer Wert = I_i

then Speichere Zustand für RESTART bei I_i ;

if neuer Wert $< I_{i-1}$

then Laden eines zufälligen Zustands " I_{i-1} "; Setze Simulation fort;

else Übergebe Wert an LRE-Auswertung;

end

Terminierung Stufe i :

Multiplikation der G-Werte mit U für I_{i-1} bis I_i ;

Bestimme neuen Umrechnungsfaktor $U := G_{I_i|\alpha \geq I_{i-1}}$;

end

Terminierung:

Ausgabe der LRE-Auswertung:

kompl. VF $\tilde{G}_i, \tilde{q}_i, d_i$ für $i = 0, \dots, B$;

8.6 RESTART/LRE für Warternetze

Die Erweiterung des RESTART/LRE-Algorithmus auf Warteschlangennetze, siehe Abschnitt 3.5, ist im Prinzip eine direkte Erweiterung seiner Anwendung auf elementare Wartesysteme, wobei jeweils ein Knoten r zur Untersuchung herausgegriffen wird und der Rest des Systems als sog. Restsystem zusammengefaßt wird, siehe Abbildung 8.7. Das Restsystem enthält auch die Quellen des Systems. Es müssen während der Simulation von Stufe i die Zustände beim Übergang von $I_i - 1$ nach I_i gespeichert werden, dazu gehören die Zustände aller Quellen des Systems, aller Warteschlangen des Systems und aller Bedienstationen des Systems inklusive der Restbedienzeiten der gerade in Bearbeitung befindlichen Aufträge.

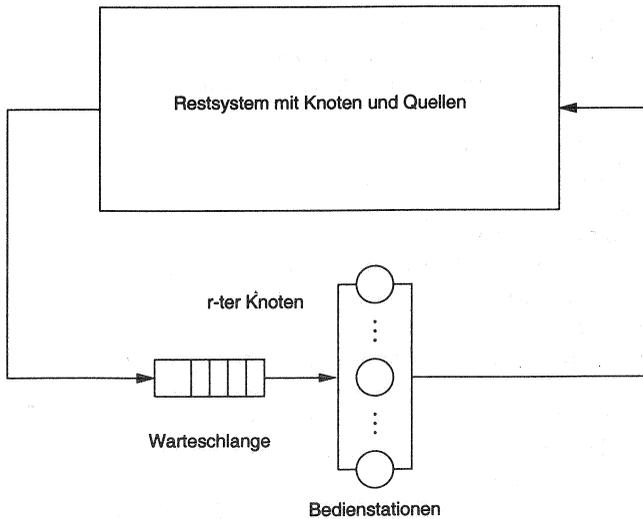


Abbildung 8.7: Untersuchung des r -ten Knotens in einem Warternetz

Das betrachtete Wartesystem mit Index r , welches untersucht wird, steuert die RESTART/LRE-Simulation, d. h. die Einteilung des RESTART-Algorithmus wird in Bezug auf dieses Wartesystem vorgenommen, und es wird die Ankunftsbelegung α_r dieses Wartesystems und damit die Verlustwahrscheinlichkeit P_{B_r} am Knoten r mit dem LRE-Algorithmus ausgewertet. Die Auswertung erfolgt wie in Abschnitt 8.5 beschrieben in den einzelnen Stufen des RESTART/LRE-Algorithmus unter der Bedingung $\alpha_r \geq I_i$.

Diese Vorgehensweise muß für jeden Knoten einzeln vorgenommen werden. Dies ist dann vorteilhaft, wenn sehr kleine Verlustwahrscheinlichkeiten simulativ ermittelt werden sollen. Eine getrennte Simulation für jeden Knoten lohnt sich, wenn die Zahl der Knoten klein ist gegenüber der Beschleunigung, die durch das RESTART/LRE-Verfahren erreicht wird.

Betrachtet man beispielsweise das Tandemnetz gemäß Abbildung 8.8, so wird die Simulation getrennt für den ersten und zweiten Knoten durchgeführt. Als Ergebnis erhält man die Verlustwahrscheinlichkeiten der einzelnen Knoten und für das gesamte Netz läßt sich die Verlustwahrscheinlichkeit nach folgender Formel bestimmen:

$$P_B = P_{B_1} + P_{B_2}(1 - P_{B_1}) \quad (8.39)$$

In Abschnitt 8.7 werden Tandemnetze als Modelle für ATM-Referenzverbindungen untersucht.

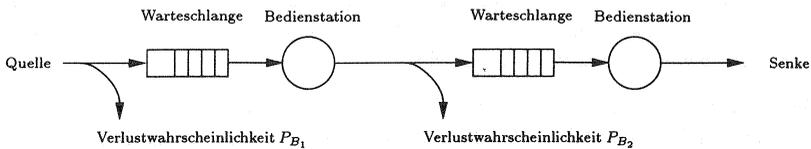


Abbildung 8.8: Tandemnetz

8.7 Verlustwahrscheinlichkeiten in ATM-Netzen

ATM (ATM – *Asynchronous Transfer Mode*) wurde 1989 vom CCITT (jetzt ITU-T) als Verfahren zur Übertragung und Vermittlung im zukünftigen Breitband-ISDN (Bitraten bis 622 Mbit/s) standardisiert. ATM stellt ein modifiziertes Paketvermittlungsverfahren dar mit folgenden Besonderheiten: Die Pakete aller Verbindungen, die hier Zellen genannt werden, sind gleich lang, so daß auf eine Paketlängenangabe im Zellkopf verzichtet werden kann. Sie sind, verglichen mit herkömmlichen Paketvermittlungsverfahren, relativ kurz (Informationsfeld: 48 *byte*, Zellkopf: 5 *byte*). Dadurch bleibt eine Störung durch Zellverlust oder Fehlleitung einer Zelle in einem Vermittlungsknoten gering, und die Signalverzögerung infolge der Paketierung beim Sender wird nicht zu groß.

Die Übermittlung erfolgt verbindungsorientiert über virtuelle Kanäle und virtuelle Pfade. In einem Verbindungsaufbauvorgang werden Betriebsmittel des Netzes für eine Verbindung reserviert. Die Zellen einer Verbindung nehmen alle denselben Weg durch das Netz. Der Kopf jeder Zelle trägt u. a. eine Nummer, die den virtuellen Kanal und den virtuellen Pfad kennzeichnet.

Der Zellkopf wird fehlerkorrigierend codiert. Auf die Quittierung und wiederholte Übertragung im Fehlerfall nach Übertragung einer Zelle über einen Teilabschnitt wird verzichtet, damit eine geringe Signalverzögerung für Sprachübertragung und Echtzeitanwendungen gewährleistet werden kann. Dies wird durch hochwertige Übertragungsmedien mit sehr niedriger Bitfehlerwahrscheinlichkeit, wie z. B. Lichtwellenleiter, ermöglicht.

Da ATM-Netze keine Kanäle mit festen Bitraten bereitstellen, können sie ohne Zusatzeinrichtungen Dienste mit beliebigen Bitraten-Anforderungen und mit zeitlich schwankenden Bitraten übertragen. Dies ist z. B. für Videoquellen von Vorteil, weil diese aufgrund von Komprimierungsverfahren als *natürliche* Bitrate eine zeitvariable Bitrate liefern. Da nicht für die Dauer einer Verbindung ein Kanal reserviert wird, sondern Kapazität des Übertragungsmediums nur belegt wird, solange Information übertragen wird, ist eine bessere Ausnutzung der vorhandenen Kapazität als bei herkömmlichen kanalvermittelten Netzen möglich, die mit synchronem Zeitvielfach und Kanalreservierung arbeiten (STM – *Synchronous Transfer Mode*).

In Abbildung 8.9 ist ein Vermittlungsknoten für eine Übertragungsrichtung schematisch dargestellt. Zellen von Verbindungen aus verschiedenen Richtungen werden auf eine gemeinsame Ausgangsleitung vermittelt. Die Übertragung einer Zelle auf einer Ausgangsleitung erfordert Zeit (Länge der Zelle dividiert durch Bitrate), so daß später ankommende Zellen ggf. in einem Zwischenspeicher warten müssen. Hier ist die ausgangsseitige Speicherung gezeigt, da Untersuchungen ergeben haben, daß diese Form der Speicherung den Durchsatz nicht beschränkt [KHM87].

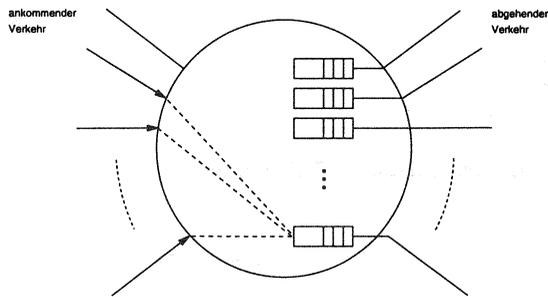


Abbildung 8.9: ATM-Vermittlungsknoten mit mehreren Multiplexern

Somit kann jede Ausgangsleitung eines Vermittlungsknotens als ein Wartesystem aufgefaßt werden, dessen Bedienprozeß durch die konstante Dauer des Auslesevorgangs repräsentiert wird. Der Eingabeprozess besteht aus einer Überlagerung von Zellströmen verschiedener Verkehrsquellen, siehe Abbildung 8.10. Untersuchungen [RW89] ergaben, daß diese Überlagerungen auf Zwischenankunftszeit-Prozesse führen, deren Korrelationen nicht vernachlässigbar sind.

In Abbildung 8.10 wird das als *statistisches Multiplexen* bezeichnete Verfahren dargestellt. Zur Sicherstellung einer ausreichenden Übertragungsqualität sind sehr niedrige Wahrscheinlichkeiten für Zellverluste in einem Vermittlungsknoten vorgeschrieben. Experten [LB91] nennen Größenordnungen von 10^{-9} bei Videoquellen. Im ITU-T Standard I.356 (Stand Mai 1996) [Rec96] wird für die anspruchsvollste Dienstgüteklasse (*class 1: stringent class*) ein Wert von 3×10^{-7} genannt. Diese Werte sollen nach den ersten Betriebserfahrungen angepaßt werden auf z. B. 10^{-8} . Das Aufspüren von so seltenen Verlustereignissen für die Pufferdimensionierung nach herkömmlichen Simulationsverfahren würde Simulationslaufzeiten in der Größenordnung Monate bis Jahre erfordern. Daher sind hier Verfahren für die Simulation seltener Ereignisse notwendig, um akzeptable Simulationslaufzeiten zu erreichen.

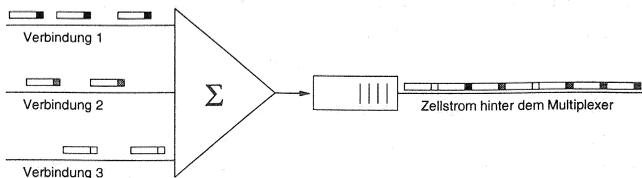


Abbildung 8.10: ATM Multiplexer

Im folgenden werden Ergebnisse für die RESTART/LRE-Simulation von ein- und mehrstufigen ATM-Netzen vorgestellt, die die Vorteile des Verfahrens eindrucksvoll demonstrieren.

8.7.1 Modelle für einzelne ATM-Knoten

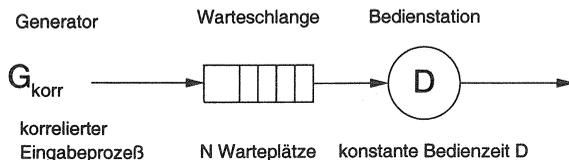


Abbildung 8.11: Modell eines einzelnen ATM-Knotens als Wartesystem

Das RESTART/LRE-Verfahren wurde insbesondere für die Untersuchung von Modellen für ATM-Vermittlungsknoten konzipiert. Ergebnisse für ausgewählte Systeme zur Validierung der Simulationsmethode wurden in Abschnitt 8.3.2 beschrieben. Das allgemeine Modell eines ATM-Vermittlungsknotens mit Pufferung am Ausgang wird in Abbildung 8.11 dargestellt. Realitätsnahe Untersuchungen müssen Korrelationen der von den Quellen erzeugten Zellströme berücksichtigen, wie sie z. B. durch das SSMP(Geo,Geo)/D/1/N-Modell beschrieben werden. Dieses Modell erlaubt durch Vergleich mit mathematisch berechenbaren Ergebnissen eine Validierung der Simulationsmethode für ein komplexes Modell, das einen korrelierten Ankunftsprozeß und nur diskrete Verteilungen enthält, wobei bei gleichzeitigen Ankünften und Abgängen eine Reihenfolge (z. B. AF – *Arrival First*, DF – *Departure First*) vereinbart werden muß, da die Reihenfolge der Bearbeitung Einfluß auf die Verlustwahrscheinlichkeit hat. Im folgenden wird die für die Verlustwahrscheinlichkeit günstigere Variante DF gewählt.

Abbildung 8.12 zeigt Simulationsergebnisse für den Verlauf der komplementären Verteilungsfunktion $\tilde{G}(x)$ für verschiedene Verkehrsangebote η für das Wartesystem SSMP(Geo,Geo)/D/1/36. In Tabelle 8.4 wird anhand dieses Beispiels die sehr gute Übereinstimmung der analytischen und simulativen Werte der Verlustwahrscheinlichkeit G_B bzw. \tilde{G}_B gezeigt. Die Berechnung der Parameter des SSMP folgt Gleichung (3.49) mit Vorgabe von η bzw. λ und κ , q_0 und q_1 .

Abbildung 8.13 zeigt ein typisches Ergebnis für eine Verkehrslast von $\eta = 0.4$ für den Warteraum SSMP(Geo,Geo)/D/1/36, die zu einer Verlustwahrscheinlichkeit der Größenordnung 10^{-9} führt. Die komplementäre Verteilungsfunktion $G(x)$ zeigt klar die Übereinstimmung der mathematisch berechneten mit den simulierten Werten. Diese Simulationsergebnisse wurden zur Validierung der analytischen Berechnungen

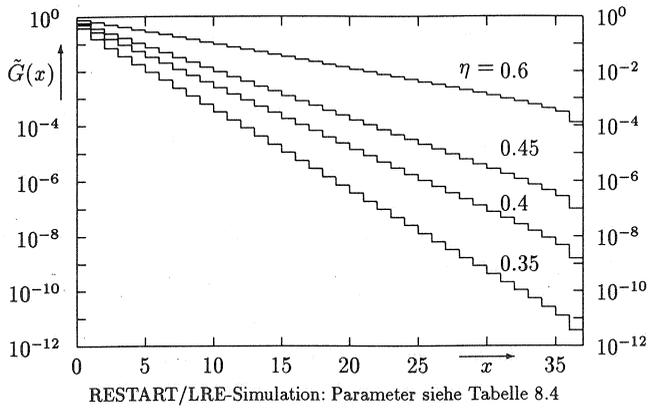
in [Her95] erzeugt, ein Beispiel für die Notwendigkeit von analytischen und simulativen Methoden.

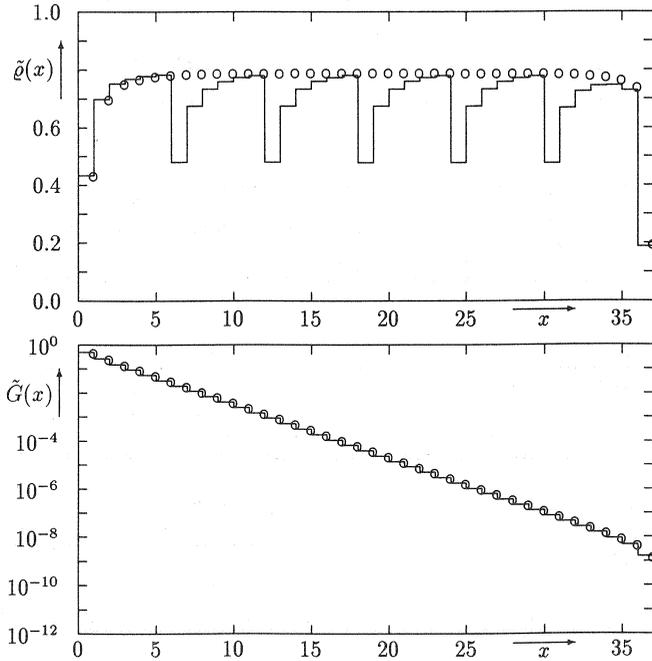
Die charakteristischen Stufen des RESTART/LRE-Verfahrens können am Verlauf des lokalen Korrelationskoeffizienten beobachtet werden, der bei jeder Stufe fällt, um dann wieder auf den Wert des ursprünglichen Systems anzusteigen. Unter Nutzung der Kenntnis des Verlaufs von $\varrho(x)$ als Glockenkurve kann man den Einfluß der Stufen entsprechend korrigieren, indem man die Maximalwerte verbindet.

Die Rechenzeit und der Speicherbedarf der analytischen Lösung hängt von der Zeilen- bzw. Spaltenzahl der quadratischen Matrix ab, die bei der Berechnung benötigt wird. Sie beträgt $(1 + B \cdot D) \cdot k$, wobei k die Anzahl der Zustände des SSMP, B die maximale Anzahl von Aufträgen im Wartesystem und D die Größe der diskreten Bediendauer bezeichnet. Im vorliegenden Fall erhält man eine (224×224) -Matrix, für die die Berechnung der Ergebnisse ca. 10 Sekunden auf einem Pentium PC unter Linux benötigt. Die Tabelle für $\eta = 0.4$ und $m = 5$ entspricht dem Beispiel in Abbildung 8.13.

Für $\eta = 0.65$ ist die Beschleunigung noch gering, aber mit fallendem η nimmt die Beschleunigung zu, so daß bereits bei $\eta = 0.45$ die direkte Simulation für $d_{max} = 0.01$ mit einer Rechenzeit von ca. einem halben Jahr nicht mehr praktisch durchführbar ist. Die Übereinstimmung der analytischen und simulativen Werte ist generell sehr gut. Die Endwerte für den lokalen Korrelationskoeffizienten $\hat{\varrho}_B$ sind teilweise aufgrund des zur Simulation benutzten Stufen-Verfahrens etwas zu niedrig. Bei diesen Auswertungen stand die effiziente Bestimmung der Verlustwahrscheinlichkeit im Vordergrund. Ist das Ziel der Untersuchung eine genaue Bestimmung der lokalen Korrelationskoeffizienten, so muß eine kleinere Stufenzahl oder eine andere Aufteilung gewählt werden, wie dies in den Beispielen mit $m = 4$ zu sehen ist. Die Tabelle gibt auch die Beschleunigung $n^*(0)/\tilde{n}$ durch Relation der Versuchszahlen von direkter und mehrstufiger Simulation an. Man sieht, daß die Beschleunigung vor allem für kleine angestrebte Verlustwahrscheinlichkeiten dramatisch steigt.

Bei der vergleichenden Untersuchung von Verbindungsannahmestrategien (CAC – *Connection Admission Control*) [Kr95, Vor96] muß zur Validierung der berechneten Leistungskenngrößen fast immer die Simulation herangezogen werden. Die entsprechenden bisherigen Arbeiten haben die Modellparameter so gewählt, daß die zu untersuchende Verlustwahrscheinlichkeit in der Größenordnung 10^{-5} liegt, um sie mit direkter Simulation validieren zu können. Das mehrstufige RESTART/LRE-Verfahren wird in Zukunft die Validierung auch für Dienste gestatten, bei denen viel kleinere Verlustwahrscheinlichkeiten von z. B. 10^{-9} gefordert werden.

Abbildung 8.12: SSMP(Geo,Geo)/D/1/36: verschiedene Verkehrsangebote η



Kompl. Verteilungsfunktion $\tilde{G}(x)$	Lokaler Korrelationskoeffizient $\tilde{g}(x)$
Simulation: Treppenkurve für $m = 5$	Analytische Werte für $G(x)$ und $g(x)$: \circ
Stufenzahl $m = 5$	Stützstellen $(I_0, \dots, I_4) = (6, 12, 18, 24, 30)$
Warteplätze: $N = 36$	Maximale Belegung: $B = 37$
Verkehrslast: $\eta = 0.4$	Verluste: $P_B = G_B \approx 1.57 \times 10^{-9}$
Max. Fehler: $d_{max} = 0.01$	Benötigte Versuchszahl: $n \approx 17.4 \times 10^7$

Abbildung 8.13: RESTART/LRE-Simulation: SSMP(Geo,Geo)/D/1/36

Tabelle 8.4: Simulationsergebnisse: SSMP(Geo,Geo)/D/1/36 für verschiedene Lasten

q_0	0.9	Parameter der ersten Geometrischen Verteilung des SSMP
q_1	0.3	Parameter der zweiten Geometrischen Verteilung des SSMP
κ	0.6	Lok. Korrelationskoeffizient der eingebetteten Markoff-Kette
D	3	Deterministische Bedienzeit des Servers
m	5	Stützstellen: $(I_0, \dots, I_4) = (6, 12, 18, 24, 30)$
m	4	Stützstellen: $(I_0, \dots, I_3) = (5, 10, 15, 20)$

SSMP(Geo/Geo)/D/1/36 $\eta = 0.65$					
	Verlustwahrsch.	Korrel.-Koeff.	Anzahl Versuche	$n^*(0)/\bar{n}$	Rechenzeit
analytisch	$\hat{G}_B = 5.602 \times 10^{-4}$	$\hat{\rho}_B = 0.246$	-	-	≈ 10 Sek.
RESTART/LRE-Berechnung $m = 0, d_{max} = 0.01$	-	-	$n^*(0) = 29.48 \times 10^6$	-	49 Min.
RESTART/LRE-Simulation $m = 5, d_{max} = 0.01$	$\hat{G}_B = 5.698 \times 10^{-4}$	$\hat{\rho}_B = 0.157$	$\bar{n} = 22.3 \times 10^6$	1.32	37 Min.
RESTART/LRE-Berechnung $m = 0, d_{max} = 0.005$	-	-	$n^*(0) = 117.93 \times 10^6$	-	196 Min.
RESTART/LRE-Simulation $m = 4, d_{max} = 0.005$	$\hat{G}_B = 5.631 \times 10^{-4}$	$\hat{\rho}_B = 0.233$	$\bar{n} = 91.44 \times 10^6$	1.29	152 Min.

SSMP(Geo/Geo)/D/1/36 $\eta = 0.6$					
	Verlustwahrsch.	Korrel.-Koeff.	Anzahl Versuche	$n^*(0)/\bar{n}$	Rechenzeit
analytisch	$\hat{G}_B = 1.287 \times 10^{-4}$	$\hat{\rho}_B = 0.234$	-	-	≈ 10 Sek.
RESTART/LRE-Berechnung $m = 0, d_{max} = 0.01$	-	-	$n^*(0) = 125.15 \times 10^6$	-	3.5 Std.
RESTART/LRE-Simulation $m = 5, d_{max} = 0.01$	$\hat{G}_B = 1.302 \times 10^{-4}$	$\hat{\rho}_B = 0.173$	$\bar{n} = 22.76 \times 10^6$	5.5	38 Min.
RESTART/LRE-Berechnung $m = 0, d_{max} = 0.005$	-	-	$n^*(0) = 500.63 \times 10^6$	-	14 Std.
RESTART/LRE-Simulation $m = 4, d_{max} = 0.005$	$\hat{G}_B = 1.301 \times 10^{-4}$	$\hat{\rho}_B = 0.235$	$\bar{n} = 152.52 \times 10^6$	3.3	4.24 Std.

		SSMP(Geo/Geo)/D/1/36 $\eta = 0.45$			
	Verlustwahrsch.	Korrel.-Koeff.	Anzahl Versuche	$n^*(0)/\bar{n}$	Rechenzeit
analytisch	$G_B = 9.756 \times 10^{-8}$	$\ell_B = 0.210$	-	-	≈ 10 Sek.
RESTART/LRE-Berechnung $m = 0, d_{max} = 0.01$	-	-	$n^*(0) = 1.57 \times 10^{11}$	-	0.5 Jahre
RESTART/LRE-Simulation $m = 5, d_{max} = 0.01$	$\tilde{G}_B = 9.966 \times 10^{-8}$	$\tilde{\ell}_B = 0.191$	$\bar{n} = 87.79 \times 10^6$	1.79×10^3	2.4 Std.

		SSMP(Geo/Geo)/D/1/36 $\eta = 0.4$			
	Verlustwahrsch.	Korrel.-Koeff.	Anzahl Versuche	$n^*(0)/\bar{n}$	Rechenzeit
analytisch	$G_B = 1.567 \times 10^{-9}$	$\ell_B = 0.195$	-	-	≈ 10 Sek.
RESTART/LRE-Berechnung $m = 0, d_{max} = 0.01$	-	-	$n^*(0) = 9.47 \times 10^{12}$	-	30 Jahre
RESTART/LRE-Simulation $m = 5, d_{max} = 0.01$	$\tilde{G}_B = 1.569 \times 10^{-9}$	$\tilde{\ell}_B = 0.187$	$\bar{n} = 17.4 \times 10^7$	54.0×10^3	5 Std.

		SSMP(Geo/Geo)/D/1/36 $\eta = 0.35$			
	Verlustwahrsch.	Korrel.-Koeff.	Anzahl Versuche	$n^*(0)/\bar{n}$	Rechenzeit
analytisch	$G_B = 3.62 \times 10^{-12}$	$\ell_B = 0.175$	-	-	≈ 10 Sek.
RESTART/LRE-Berechnung $m = 0, d_{max} = 0.05$	-	-	$n^*(0) = 1.57 \times 10^{14}$	-	500 Jahre
RESTART/LRE-Simulation $m = 5, d_{max} = 0.05$	$\tilde{G}_B = 3.57 \times 10^{-12}$	$\tilde{\ell}_B = 0.162$	$\bar{n} = 8.35 \times 10^6$	1.88×10^7	14 Min.

Die angegebene Rechenzeit für die Simulationen bezieht sich auf einen mittleren Wert von 10^4 Versuchen pro Sekunde.

Die Rechenzeit der analytischen Ergebnisse beziehen sich auf die Ausführungszeit auf einem Pentium 90 Mhz unter Linux.

8.7.2 Modelle für sequentielle ATM-Knoten (Netzmodelle)

Um weiterhin einen Vergleich mit analytisch berechenbaren Ergebnissen zu ermöglichen, wird hier ein sog. Tandemnetz als Modell zur Berechnung der Verlustwahrscheinlichkeit einer virtuellen ATM-Verbindung untersucht, siehe Abbildung 8.14. Die Bediendauerverteilung im ersten Knoten wird deterministisch gewählt. Im zweiten Knoten muß eine andere Verteilung (z. B. negativ-exponentiell) gewählt werden,

da das Analyseverfahren nur in einem Knoten eine deterministische Verteilung erlaubt.

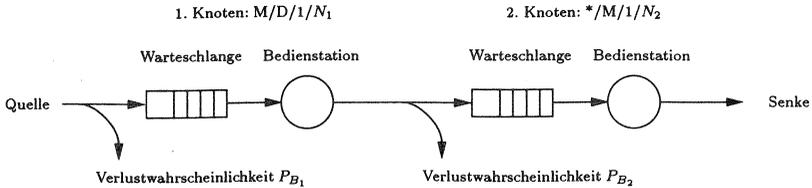


Abbildung 8.14: Tandemnetz: $M/D/1/N_1 \rightarrow */M/1/N_2$

Der erste Warteraum ist vom Typ $M/D/1/N_1$. Der Abgangsprozeß des ersten Warteraums ist der Ankunftsprozeß des zweiten Warteraums. Er wird in der Typangabe $*/M/1/N_2$ des zweiten Warteraums mit "*" bezeichnet. Im Beispiel wurde $N_1 = N_2 = 11$ gewählt. Dies führt mit $\lambda = 0.4$ und $E[\tau_{b1}] = E[\tau_{b2}] = 1$, d. h. $\eta = 0.4$, für den ersten Warteraum zu einer analytisch berechneten Belegungswahrscheinlichkeit, die identisch zur Verlustwahrscheinlichkeit³ ist, von $P_{B1} = 6.53 \cdot 10^{-9}$.

Für den Vergleich werden mit TimeNET (TimeNET – *Timed Net Evaluation Tool*) [GKZH94, KGZH95] berechnete Ergebnisse⁴ herangezogen. TimeNET basiert auf zeiterweiterten Petri-Netzen. Als Ergebnis erhält man Belegungswahrscheinlichkeiten und Durchsätze. Am zweiten Knoten unterscheiden sich die Wahrscheinlichkeiten für die Belegung und die Ankunftsbelegung, so daß die Bestimmung der Verlustwahrscheinlichkeit P_{B2} über folgende Durchsatzäquivalenz erfolgen muß, wobei λ_{eff} den effektiven Durchsatz am Ausgang des zweiten Knotens bezeichnet:

$$\lambda(1 - P_{B1})(1 - P_{B2}) = \lambda_{eff} \quad (8.40)$$

$$P_{B2} = 1 - \frac{\lambda_{eff}}{\lambda(1 - P_{B1})} = 3.352 \cdot 10^{-6} \quad (8.41)$$

Die RESTART/LRE-Simulation ergab mit $m = 3$ Stufen und einer LRE-Fehlervorgabe von $d_{max} = 0.01$ eine Verlustwahrscheinlichkeit von $\tilde{P}_{B2} = 3.353 \cdot 10^{-6}$ für das nachgeschaltete Wartesystem, siehe Tabelle 8.5. Im zweiten Beispiel in Tabelle 8.6 wird die Bediendauerverteilung im zweiten Knoten als Annäherung an eine deterministische Verteilung durch eine Erlang-2-Verteilung dargestellt. Die Ergebnisse für den ersten Knoten bleiben erhalten. Für den zweiten Knoten erhält man eine um 2 Größenordnungen niedrigere Verlustwahrscheinlichkeit, d. h. die Verteilung der Bediendauer hat einen starken Einfluß auf die Verlustwahrscheinlichkeit.

³Dies gilt aufgrund des Poisson-Ankunftsprozesses mit der sog. PASTA-Eigenschaft: *Poisson Arrivals See Time Averages*.

⁴Diese analytisch berechneten Ergebnisse wurden mit TimeNET von Prof. B. Müller-Clostermann der GH-Essen ermittelt. TimeNET erlaubt nur *eine* deterministische Bediendauer im Netz.

Abbildung 8.15 zeigt das äquivalente zeiterweiterte Petri-Netz, wie es für TimeNET definiert wurde.

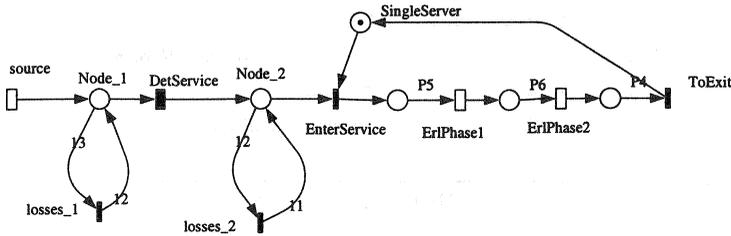


Abbildung 8.15: Tandemnetz als zeiterweitertes Petri-Netz

Die Übereinstimmung zwischen den simulativ und analytisch ermittelten Werten ist sehr gut. Bei allen durchgeführten Simulationen war die prozentuale Abweichung unter 1 %.

Andere Ansätze zur Beschleunigung von ATM-Netz-Simulationen [Leh93, SKMLK95] basieren auf der verteilten Simulation mit Modellaufteilung, siehe Abschnitt 7.1, sowie der Ausnutzung der speziellen Eigenschaft eines ATM-Netzes, bei dem alle Ereigniszeitpunkte durch die konstante Länge der Zellen vorgegeben sind. Damit kann die Laufzeit einer Netz-Simulation bei entsprechender Anzahl Prozessoren in etwa auf die Laufzeit einer Simulation eines ATM-Knotens reduziert werden. Damit lassen sich Simulationen mit einem Verhältnis von Modellzeit:Rechnerzeit von ungefähr 40:1 [SKMLK95] durchführen, d. h. zur Untersuchung seltener Ereignisse sind weiterhin zusätzliche Techniken wie das RESTART/LRE-Verfahren notwendig.

Die hier vorgestellten, vielversprechenden Ergebnisse motivieren zur Untersuchung umfangreicherer Netzmodelle für ATM-Referenzverbindungen. Ziel ist die Weiterentwicklung des vorliegenden RESTART/LRE-Simulators, so daß Untersuchungen von Netzmodellen durch eine entsprechende Parametrisierung möglich werden, um als Werkzeug einem größeren Spezialistenkreis verfügbar gemacht zu werden.

Tabelle 8.5: Simulation ATM-Verbindung: Tandemnetz

Knoten 1	M/D/1/N ₁ = 11	Warteraumtyp
E[τ _{b1}]	1	mittlere Bediendauer 1. Knoten
Knoten 2	*M/1/N ₂ = 11	Warteraumtyp
E[τ _{b2}]	1	mittlere Bediendauer 2. Knoten
λ	0.4	Ankunftsrate
d _{max}	0.01	maximaler LRE-Fehler

	M/D/1/N ₁ = 11 → *M/1/N ₂ = 11	
	sim.: m = 3	analytisch
\tilde{P}_{B1} bzw. P_{B1}	$6.587 \cdot 10^{-9}$	$6.530 \cdot 10^{-9}$
\tilde{P}_{B2} bzw. P_{B2}	$3.353 \cdot 10^{-6}$	$3.352 \cdot 10^{-6}$
\tilde{P}_B bzw. P_B	$3.353 \cdot 10^{-6}$	$3.358 \cdot 10^{-6}$

Tabelle 8.6: Simulation ATM-Verbindung: Tandemnetz

Knoten 1	M/D/1/N ₁ = 11	Warteraumtyp
E[τ _{b1}]	1	mittlere Bediendauer 1. Knoten
Knoten 2	*E ₂ /1/N ₂ = 11	Warteraumtyp
E[τ _{b2}]	1	mittlere Bediendauer 2. Knoten
λ	0.4	Ankunftsrate
d _{max}	0.01	maximaler LRE-Fehler

	M/D/1/N ₁ = 11 → *E ₂ /1/N ₂ = 11	
	sim.: m = 3	analytisch
\tilde{P}_{B1} bzw. P_{B1}	$6.587 \cdot 10^{-9}$	$6.530 \cdot 10^{-9}$
\tilde{P}_{B2} bzw. P_{B2}	$3.550 \cdot 10^{-8}$	$3.550 \cdot 10^{-8}$
\tilde{P}_B bzw. P_B	$4.209 \cdot 10^{-8}$	$4.203 \cdot 10^{-8}$

Zusammenfassung und Ausblick

Diese Habilitationsschrift gibt eine Übersicht über wichtige verkehrstheoretische Modelle sowie bekannte und neue stochastische Simulationstechniken zur Bestimmung charakteristischer Leistungskenngrößen von Kommunikationsnetzen. Die Modellierung und Anwendung der Techniken wurde exemplarisch an sehr unterschiedlichen Modellen von Teilsystemen von Kommunikationsnetzen gezeigt, um zu demonstrieren, welche Rolle diese Techniken bei der Entwicklung neuer Kommunikationsnetze spielen bzw. spielen können.

Ein Schwerpunkt der Arbeit liegt bei der simulativen Untersuchung seltener Ereignisse, so daß dieses Thema besonders ausführlich behandelt wird und das hier entwickelte RESTART/LRE-Verfahren mathematisch und experimentell untersucht wird. Diese Arbeit leistet auch Beiträge zur Entwicklung eines RESTART/LRE-Simulators, der in Zukunft einem größeren Anwenderkreis als abgerundetes Programmpaket basierend auf der CNCL-Klassenbibliothek zur Verfügung stehen soll.

Es zeigt sich, daß mit dem RESTART/LRE-Verfahren ATM-Zellverlustwahrscheinlichkeiten der Größenordnung 10^{-9} simulativ mit vertretbaren Laufzeiten von einigen Minuten bis zu einigen Stunden – abhängig von der geforderten Genauigkeit – durchgeführt werden können, so daß Untersuchungen dieser Art zukünftig mit Parametern arbeiten können, wie sie in der Praxis benötigt werden. Zur weiteren Beschleunigung kann das RESTART/LRE-Verfahren in einfacher Weise parallelisiert werden.

Verfahren zur Simulationsbeschleunigung sind aufgrund der Dienstgüteanforderungen neuer Netze, wie z. B. ATM, oder auch allgemein im Bereich der Zuverlässigkeitsanalyse, von aktuellem Interesse [Kel96, GHSZ96]. Weitere Arbeiten sind notwendig, die die Bestimmung der optimalen Verfahrensparameter betreffen. Dies betrifft beispielweise den Aufsplittungsgrad des RESTART-Verfahrens bzw. die benötigte Anzahl gespeicherter Zustände. Ferner sind Arbeiten notwendig, die die Anwendbarkeit des Verfahrens und die zugehörigen Zielgrößen genauer untersuchen, d. h. die Klassen von Wartemodellen definieren, für die das Verfahren anwendbar bzw. nicht anwendbar ist.

Die Entwicklung der Simulationstechnik zu Werkzeugen mit graphischer Ein- und Ausgabe, die ohne Programmierkenntnisse eingesetzt werden können, führt dazu, daß auch Nichtfachleute der Simulationstechnik diese einsetzen. Dies ist zu befürworten solange dies dem Ziel dient, das Verhalten eines Systems zu verstehen. Der Pro-

duktentwickler sieht, welche Möglichkeiten der Beschreibung es gibt und kann damit die Eigenschaften seines Produkts besser beschreiben und mit Hilfe animierter Abläufe besser verstehen und ausnutzen. Eine quantitative Leistungsbewertung sollte nur von einem Experten durchgeführt werden, der die Simulationstechnik, analytische Techniken zur Validierung, und Spezialmethoden beherrscht, wozu auch das RESTART/LRE-Verfahren gehört, und der die erzielten Ergebnisse interpretieren kann.

Für die in dieser Arbeit vorgestellten Techniken gilt dasselbe wie für jede andere Technik auch und deshalb möchte zum Abschluß Shannon [Sha81] zitieren, der die Simulationstechnik mit dem Skalpell eines Chirurgen vergleicht: *“in the right hand [it] can accomplish tremendous good, but it must be used with great care and by someone who knows what they are doing.”*

ANHANG A

Wichtige Verteilungen

Gleichverteilung $U(a, b)$

Anwendungsbereich:	Verteilung einer Größe über die außer dem Wertebereich $[a, b]$ nichts bekannt ist. Die Verteilung $U(0,1)$ bildet die Grundlage bei der Generierung von Zufallszahlen, siehe Abschnitt 4.
Dichtefunktion:	$f(x) = \begin{cases} \frac{1}{b-a} & \text{für } a \leq x \leq b \\ 0 & \text{sonst} \end{cases} \quad (\text{A.1})$
Verteilungsfunktion:	$F(x) = \begin{cases} 0 & \text{für } x < a \\ \frac{x-a}{b-a} & \text{für } a \leq x \leq b \\ 0 & \text{sonst} \end{cases} \quad (\text{A.2})$
Parameter:	a und b reelle Zahlen mit $a < b$
Wertebereich:	$[a, b]$
Mittelwert:	$E[X] = \frac{a+b}{2}$
Varianz:	$\text{Var}[X] = \frac{(b-a)^2}{12}$
Variationskoeffizient:	$c = \frac{b-a}{\sqrt{3} \cdot (a+b)}$
Bemerkungen:	Die $U(0,1)$ -Verteilung ist ein Spezialfall der β -Verteilung für $\alpha_1 = \alpha_2 = 1$.

Negativ-Exponentielle Verteilung $M(\lambda)$

Anwendungsbereich:	Zwischenankunftszeiten von Aufträgen, Nachrichten usw. in einem System
Dichtefunktion:	$f(x) = \begin{cases} \lambda \cdot e^{-\lambda x} & x \geq 0 \\ 0 & \text{sonst} \end{cases} \quad (\text{A.3})$
Verteilungsfunktion:	$F(x) = \begin{cases} 1 - e^{-\lambda x} & x \geq 0 \\ 0 & \text{sonst} \end{cases} \quad (\text{A.4})$
Parameter:	Intensität $\lambda > 0$
Wertebereich:	$[0, \infty)$
Mittelwert:	$E[X] = 1/\lambda$
Varianz: Variationskoeffizient:	$\text{Var}[X] = E[X]^2 = 1/\lambda^2$ $c = 1$
Bemerkungen:	<p>Die negativ-exponentielle Verteilung ist ein Spezialfall der Gamma- und der Weibull-Verteilung.</p> <p>Die Exponentialverteilung ist die einzige kontinuierliche Verteilung mit der Markoff-Eigenschaft (Gedächtnislosigkeit), siehe Abschnitt 3.2.</p> <p>Sind X_1, X_2, \dots, X_k unabhängige negativ-exponentiell verteilte Zufallsvariablen, so ist die Summe $X_1 + X_2 + \dots + X_k$ Erlang-k verteilt.</p>

Erlang-k-Verteilung $E_k(\mu)$

Anwendungsbereich:	schwach streuende Zwischenankunftszeiten oder Bediendauern von Aufträgen, Nachrichten usw. in einem System
Dichtefunktion:	$f(x) = \begin{cases} k\mu \frac{(k\mu x)^{k-1} e^{-k\mu x}}{(k-1)!} & x \geq 0 \\ 0 & \text{sonst} \end{cases} \quad (\text{A.5})$
Verteilungsfunktion:	$F(x) = \begin{cases} 1 - e^{-k\mu x} \sum_{i=0}^{k-1} \frac{(k\mu x)^i}{i!} & x \geq 0 \\ 0 & \text{sonst} \end{cases} \quad (\text{A.6})$
Parameter:	Intensität $\mu > 0$
Wertebereich:	$[0, \infty)$
Mittelwert:	$E[X] = 1/\mu$
Varianz:	$\text{Var}[X] = \frac{1}{k\mu^2}$
Variationskoeffizient:	$c^2 = 1/k$
Bemerkungen:	<p>Die Erlang-k-Verteilung entspricht einer Summe $X_1 + X_2 + \dots + X_k$ unabhängiger negativ-exponentiell verteilter Zufallsvariablen X_i mit Parameter $\lambda = k\mu$.</p> <p>Die Erlang-k-Verteilung ist ein Spezialfall der Gamma-Verteilung.</p>

Hyperexponentielle Verteilung $H_k(a_1, \mu_1, a_2, \mu_2, \dots, a_n, \mu_n)$

Anwendungsbereich:	stark streuende Zwischenankunftszeiten oder Bediendauern von Aufträgen, Nachrichten usw. in einem System
Dichtefunktion:	$f(x) = \begin{cases} \sum_{i=1}^n p_i \mu_i e^{-\mu_i x} & x \geq 0 \\ 0 & \text{sonst} \end{cases} \quad (\text{A.7})$
Verteilungsfunktion:	$F(x) = \begin{cases} 1 - \sum_{i=1}^n p_i e^{-\mu_i x} & x \geq 0 \\ 0 & \text{sonst} \end{cases} \quad (\text{A.8})$
Parameter:	Intensitäten $\mu_i > 0$ und Wahrscheinlichkeiten $p_i > 0$ mit $\sum_{i=1}^n p_i = 1$
Wertebereich:	$[0, \infty)$
Mittelwert:	$E[X] = \sum_{i=1}^n p_i / \mu_i$
Varianz: Variationskoeffizient:	$\text{Var}[X] = \sum_{i=1}^n \frac{p_i}{\mu_i} (2/\mu_i - 1)$ $c^2 = \text{Var}[X] / E[X]^2$
Bemerkungen:	Die Hyperexponentielle Verteilung entspricht einer Mischverteilung unabhängiger negativ-exponentiell verteilter Zufallsvariablen X_i mit Parameter μ_i .

Gamma-Verteilung $\gamma(\alpha, \beta)$

Anwendungsbereich:	Zeit bis zur Beendigung eines Vorgangs, z.B. Bedienung eines Auftrags
Dichtefunktion:	$f(x) = \begin{cases} \frac{\beta^{-\alpha} x^{\alpha-1} e^{-x/\beta}}{\Gamma(\alpha)} & \text{für } x > 0 \\ 0 & \text{sonst} \end{cases} \quad (\text{A.9})$
Verteilungsfunktion:	<p>Wenn α nicht ganzzahlig ist, gibt es keine geschlossene Form. Wenn α ganzzahlig ist, dann gilt</p> $F(x) = \begin{cases} 1 - e^{-x/\beta} \sum_{j=0}^{\alpha-1} (x/\beta)^j j! & \text{für } x > 0 \\ 0 & \text{sonst} \end{cases} \quad (\text{A.10})$
Parameter:	Formparameter α , Skalierungsparameter β
Wertebereich:	$[0, \infty)$
Mittelwert:	$E[X] = \alpha\beta$
Varianz:	$\text{Var}[X] = \alpha\beta^2$
Variationskoeffizient:	$c^2 = \frac{1}{\alpha}$
Bemerkungen:	Die χ^2 -Verteilung mit k Freiheitsgraden ergibt sich für $\alpha = k/2$ und $\beta = 2$.

Normal-Verteilung $N(\mu, \sigma^2)$

Anwendungsbereich:	Berechnung von Fehlermaßen und Konfidenzintervallen für Summen von Zufallsvariablen aufgrund des Zentralen Grenzwertsatzes
Dichtefunktion:	$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2} \text{ für alle reellen } x \quad (\text{A.11})$
Verteilungsfunktion:	keine geschlossene Form
Parameter:	Ortsparameter $\mu \in (-\infty, \infty)$, Skalierungsparameter $\sigma > 0$
Wertebereich:	$(-\infty, \infty)$
Mittelwert:	$E[X] = \mu$
Varianz:	$\text{Var}[X] = \sigma^2$
Variationskoeffizient:	$c = \sigma/\mu$
Bemerkungen:	$N(0,1)$ wird als Standardnormalverteilung bezeichnet. Sind zwei normalverteilte Zufallsvariable unkorreliert, so sind sie auch unabhängig. Für andere Verteilungen gilt dies im allgemeinen nicht.

Beta-Verteilung $\beta(\alpha_1, \alpha_2)$

Anwendungsbereich:	Wird häufig als Modell benutzt, wenn keine Daten vorhanden sind, da diese Verteilung die Beschreibung einer Vielzahl von Verteilungen mit endlichem Wertebereich über einfache Variation der Parameter α_1 und α_2 erlaubt.
Dichtefunktion:	$f(x) = \begin{cases} \frac{x^{\alpha_1-1}(1-x)^{\alpha_2-1}}{B(\alpha_1, \alpha_2)} & \text{für } 0 < x < 1 \\ 0 & \text{sonst} \end{cases} \quad (\text{A.12})$ <p>mit der Betafunktion $B(\alpha_1, \alpha_2)$</p>
Verteilungsfunktion:	keine geschlossene Form
Parameter:	Formparameter $\alpha_1 > 0, \alpha_2 > 0$
Wertebereich:	$[0, 1]$
Mittelwert:	$E[X] = \frac{\alpha_1}{\alpha_1 + \alpha_2}$
Varianz: Variationskoeffizient:	$\text{Var}[X] = \frac{\alpha_1 \alpha_2}{(\alpha_1 + \alpha_2)^2 (\alpha_1 + \alpha_2 + 1)}$ $c^2 = \frac{\alpha_2}{\alpha_1 (\alpha_1 + \alpha_2 + 1)}$
Bemerkungen:	$U(0,1)$ stimmt mit $\text{beta}(1, 1)$ überein.

Bernoulli-Verteilung $\text{Ber}(p)$

Anwendungsbereich:	Zufallsvariable mit zwei möglichen Realisierungen, die zur Erzeugung anderer diskreter Verteilungen, z.B. binomial, geometrisch und negativ binomial benutzt wird
Wahrscheinlichkeitsfunktion:	$P(x) = \begin{cases} 1-p & \text{für } x=0 \\ p & \text{für } x=1 \\ 0 & \text{sonst} \end{cases} \quad (\text{A.13})$
Verteilungsfunktion:	$F(x) = \begin{cases} 0 & \text{für } x < 0 \\ 1-p & \text{für } 0 \leq x < 1 \\ 1 & \text{für } 1 \leq x \end{cases} \quad (\text{A.14})$
Parameter:	$p \in (0, 1)$
Wertebereich:	$\{0, 1\}$
Mittelwert:	$E[X] = p$
Varianz:	$\text{Var}[X] = p(1-p)$
Variationskoeffizient:	$c^2 = (1-p)/p$
Bemerkungen:	<p>Die Bernoulli-Verteilung(p) ist ein Spezialfall der Binomialverteilung mit $n = 1$ und p.</p> <p>Diese Verteilung wird auch als <i>binäre</i> Verteilung bezeichnet.</p> $p(x) = p^x(1-p)^{1-x} \quad \text{für } x = 0, 1 \quad (\text{A.15})$ <p>ist eine andere Form der Wahrscheinlichkeitsfunktion.</p>

Binomial-Verteilung Binom(n, p)

Anwendungsbereich:	Anzahl der Erfolge bei n unabhängigen Bernoulli-Versuchen mit Wahrscheinlichkeit p für einen Erfolg bei einem Versuch
Wahrscheinlichkeitsfunktion:	$P(x) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x} & \text{für } x \in \{0, 1, \dots, n\} \\ 0 & \text{sonst} \end{cases} \quad (\text{A.16})$
Verteilungsfunktion:	$F(x) = \begin{cases} 0 & \text{für } x < 0 \\ \sum_{i=0}^{\lfloor x \rfloor} \binom{n}{i} p^i (1-p)^{n-i} & \text{für } 0 \leq x \leq n \\ 1 & \text{für } n < x \end{cases} \quad (\text{A.17})$
Parameter:	n positive ganze Zahl, $p \in (0, 1)$
Wertebereich:	$\{0, 1, \dots, n\}$
Mittelwert:	$E[X] = n \cdot p$
Varianz:	$\text{Var}[X] = n \cdot p(1-p)$
Variationskoeffizient:	$c^2 = (1-p)/(n \cdot p)$
Bemerkungen:	Für $n = 1$ stimmt die Verteilung mit der Bernoulli-Verteilung überein.

Geometrische Verteilung $\text{Geo}(p)$

Anwendungsbereich:	Anzahl von Versuchen bis zu einem ersten Erfolg, diskrete Zwischenankunftszeiten
Wahrscheinlichkeitsfunktion:	$P(x) = \begin{cases} p(1-p)^{x-1} & \text{für } x = 1, 2, \dots \\ 0 & \text{sonst} \end{cases} \quad (\text{A.18})$
Verteilungsfunktion:	$F(x) = \begin{cases} 1 - (1-p)^{\lfloor x \rfloor} & \text{für } x \geq 0 \\ 0 & \text{sonst} \end{cases} \quad (\text{A.19})$
Parameter:	$p \in (0, 1), q = 1 - p$
Wertebereich:	$\{1, 2, \dots\}$
Mittelwert:	$E[X] = 1/p$
Varianz:	$\text{Var}[X] = q/p^2$
Variationskoeffizient:	$c^2 = q$
Bemerkungen:	<p>Die geometrische Verteilung entspricht innerhalb der diskreten Verteilungen der negativ-exponentiellen Verteilung, da sie die einzige diskrete Verteilung mit der Eigenschaft der Gedächtnislosigkeit ist.</p> <p>Teilweise wird abhängig von der Anwendung die geometrische Verteilung auch mit $p(1-p)^x$ und dem Wertebereich $\{0, 1, \dots\}$, d. h. um einen Wert verschoben angegeben. Damit ändern sich auch die Kennwerte.</p>

Poisson Verteilung $\text{Pois}(\lambda)$

Anwendungsbereich:	Anzahl von Ereignissen, die in einem Zeitintervall stattfinden, wenn die Ereignisse mit einer konstanten Rate auftreten
Wahrscheinlichkeitsfunktion:	$P(x) = \begin{cases} \frac{e^{-\lambda} \lambda^x}{x!} & \text{für } x \in \{0, 1, \dots\} \\ 0 & \text{sonst} \end{cases} \quad (\text{A.20})$
Verteilungsfunktion:	$F(x) = \begin{cases} 0 & \text{für } x < 0 \\ e^{-\lambda} \sum_{i=0}^{\lfloor x \rfloor} \frac{\lambda^i}{i!} & \text{für } 0 \leq x \end{cases} \quad (\text{A.21})$
Parameter:	$\lambda > 0$
Wertebereich:	$\{0, 1, \dots\}$
Mittelwert:	$E[X] = \lambda$
Varianz:	$\text{Var}[X] = \lambda$
Variationskoeffizient:	$c = 1$
Bemerkungen:	Die Poisson Verteilung ist eine andere Form der Beschreibung für die negativ-exponentielle Verteilung, die die Verteilung der Zeiten zwischen zwei Ereignissen beschreibt.

Ergänzungen zum RESTART/LRE-Verfahren

In diesem Anhang werden die Herleitungen zu den Ergebnisformeln in Kapitel 8 ausgeführt.

B.1 Minimale Versuchszahl n_{min}

Zur Herleitung von Gleichung (8.8) und (8.9) wird ausgehend von Gleichung (8.7) für

$$n(H_0, H_1, H_2, \dots, H_m) = \frac{1}{d_S^2} \sum_{i=0}^m \frac{\gamma_i H_{i-1}}{H_i} \quad (\text{B.1})$$

durch vollständige Induktion mit Hilfe der notwendigen Bedingung für ein Minimum gezeigt, daß die folgende Gleichung für $n(H_0^*, \dots, H_{i-1}^*, H_i, \dots, H_m)$ gilt, wobei H_j^* für $j = 0 \dots i - 1$ die bereits bestimmten optimalen Niveaus bezeichnen:

$$n(H_0^*, \dots, H_{i-1}^*, H_i, \dots, H_m) = \frac{1}{d_S^2} \left\{ \frac{(i+1)(\gamma_0 \dots \gamma_i)^{\frac{1}{i+1}}}{H_i^{\frac{1}{i+1}}} + \frac{\gamma_{i+1} H_i}{H_{i+1}} + \frac{\gamma_{i+2} H_{i+1}}{H_{i+2}} + \dots + \frac{\gamma_m H_{m-1}}{H_m} \right\} \quad (\text{B.2})$$

Die Behauptung gilt für $i = 0$, da man dafür die Ursprungsformel nach Gleichung (8.7) bzw. (B.1) erhält.

Für die vollständige Induktion ist zu zeigen, daß Gleichung (B.2) für $i + 1$ gilt. Dazu wird entsprechend der notwendigen Bedingung für ein Minimum die Ableitung nach H_i gebildet und gleich Null gesetzt:

$$\frac{\partial n(H_0^*, \dots, H_{i-1}^*, H_i, \dots, H_m)}{\partial H_i} = 0 \quad (\text{B.3})$$

Dies führt auf die Gleichung für das optimale Niveau H_i^* :

$$H_i^* = (\gamma_0 \dots \gamma_i)^{\frac{1}{i+2}} \left(\frac{H_{i+1}}{\gamma_{i+1}} \right)^{\frac{i+1}{i+2}} \quad (\text{B.4})$$

Setzt man dies in Gleichung (B.2) ein, so erhält man die entsprechende Gleichung zu (B.2) für $i + 1$, womit die Induktion und damit Gleichung (B.2) bewiesen ist.

Gleichung (B.2) für $i = m$ führt direkt zu Gleichung (8.9).

Als weitere Bedingung für ein Minimum muß die zweite Ableitung der Anzahl der Versuche $n(H_0^*, \dots, H_{i-1}^*, H_i, \dots, H_m)$ nach H_i positiv sein, siehe Gleichung (B.2):

$$\frac{\partial^2 n(H_0^*, \dots, H_{i-1}^*, H_i, \dots, H_m)}{\partial H_i \partial H_i} > 0 \quad (\text{B.5})$$

Für die vorliegende Anwendung gilt $\gamma_i > 0$ und $H_i > 0$ und damit kann man zeigen, daß auch Bedingung (B.5) gilt.

B.2 Optimale Niveaus des RESTART/LRE-Verfahrens

Zum Beweis von Gleichung (8.8) beginnt man eine Induktion in absteigender Richtung beginnend mit $m - 1$. Den Induktionsanfang erhält man mit Gleichung (B.4) für $i = m - 1$. Für die Induktion ist nun zu zeigen, daß mit $H_i^*(H_m)$ nach Gleichung (8.8) auch die entsprechende Gleichung für $H_{i-1}^*(H_m)$ gilt. Dies erreicht man durch Einsetzen von $H_i^*(H_m)$ nach Gleichung (8.8) in $H_{i-1}^*(H_i)$ nach Gleichung (B.4).

B.3 Nachweis des Minimums

Zum vollständigen Nachweis des Minimums von Gleichung (8.7) muß gezeigt werden, daß die folgende Matrix $[u_{ij}]$ positiv definit ist [BS71, Zur64] für $i = 0, \dots, m - 1$ und $j = 0, \dots, m - 1$:

$$u_{ij} = \frac{\partial^2 n}{\partial H_i \partial H_j} \Big|_{H_0=H_0^*, H_1=H_1^*, \dots, H_{m-1}=H_{m-1}^*} \quad (\text{B.6})$$

Da der positive konstante Faktor $1/d_5^2$ in Gleichung (8.7) keinen Einfluß auf den Ort des Minimums hat, wird er im folgenden nicht weiter betrachtet. Damit ergibt sich für die Elemente der Matrix:

$$\begin{aligned} u_{ii-1} &= -\gamma_i (H_i^*)^{-2} \\ u_{ii} &= 2\gamma_i H_{i-1}^* (H_i^*)^{-3} \\ u_{ii+1} &= -\gamma_{i+1} (H_{i+1}^*)^{-2} \\ u_{ij} &= 0 \quad \text{sonst} \end{aligned} \quad (\text{B.7})$$

Es handelt sich um eine Tridiagonalmatrix, d. h. eine Bandmatrix, in der nur die Diagonale sowie zwei Nebendiagonalen besetzt sind.

Wie in Abschnitt 8.3 wird angenommen, daß γ_i für $i = 0, \dots, m-1$ durch γ approximiert wird. Damit ergibt sich H_i^* mit $w = (\frac{\gamma}{\gamma_m} H_m)^{1/(m+1)} > 0$ nach Gleichung (8.16) für $i = 0, \dots, m-1$ zu:

$$H_i^* = w^{i+1} \quad (\text{B.8})$$

Einsetzen in Gleichung (B.7) mit $r := 1/w > 0$ ergibt:

$$\begin{aligned} u_{ii-1} &= -\gamma r^{2i+2} \\ u_{ii} &= 2\gamma r^{2i+3} \\ u_{ii+1} &= -\gamma r^{2i+4} \\ u_{ij} &= 0 \quad \text{sonst} \end{aligned} \quad (\text{B.9})$$

Da ein positiver Faktor in jedem Element der Matrix keinen Einfluß auf das Vorzeichen der Determinante einer Matrix hat, wird im folgenden der Faktor γr^2 als Faktor $(\gamma r^2)^m$ herausgezogen:

$$\text{Det}[u_{ij}] = (\gamma r^2)^m \text{Det}[v_{ij}] \quad (\text{B.10})$$

$$\begin{aligned} v_{ii-1} &= -r^{2i} \\ v_{ii} &= 2r^{2i+1} \\ v_{ii+1} &= -r^{2i+2} \\ v_{ij} &= 0 \quad \text{sonst} \end{aligned} \quad (\text{B.11})$$

Bezeichnet man die Determinante der $m \times m$ -Matrix mit D_m , so erhält man für $m = 1$ und $m = 2$ folgende Werte der Determinante:

$$\begin{aligned} D_m &:= \text{Det}[v_{ij}] \\ D_1 &= 2r \\ D_2 &= 3r^4 \end{aligned} \quad (\text{B.12})$$

Nach [Zur64] gilt, daß eine Matrix dann und nur dann eigentlich positiv definit ist, wenn alle Hauptabschnittsdeterminanten positiv sind. Die Hauptabschnittsdeterminanten entsprechen im vorliegenden Fall den Determinanten D_m . Per Induktion läßt sich zeigen, daß für D_m mit $r > 0$ gilt:

$$D_m = (m+1)r^{m^2} > 0 \quad \text{für } m = 1, 2, \dots \quad (\text{B.13})$$

Die Induktionsverankerung gilt mit Gleichung (B.12). Für die Induktion ist zu zeigen, daß $D_{m+1} = (m+2)r^{(m+1)^2}$ gilt. Dies erhält man, indem man D_{m+1} nach der letzten Zeile entwickelt:

$$D_{m+1} = 2r^{2(m+1)-1}D_m - (-r^{2m}(-r^{2m}D_{m-1})) = (m+2)r^{(m+1)^2} \quad (\text{B.14})$$

B.4 Maximaler Gesamtfehler d_{max}

Die Herleitung des maximalen Gesamtfehlers für G_B geht von Gleichung (8.2) aus, in der G_B als Produkt definiert ist. Mit einer abkürzenden Schreibweise für die Faktoren des Produkts:

$$H_{i|i-1} = P\{\alpha \geq i | \alpha \geq i-1\} \quad (\text{B.15})$$

ergibt sich:

$$G_B = H_{m|m-1}H_{m-1|m-2} \dots H_{1|0}H_0 \quad (\text{B.16})$$

Bei der Simulation sind die Faktoren dieses Produkts als Zufallsvariablen zu betrachten, wobei der Meßwert dem Mittelwert entspricht und der Fehler dem Variationskoeffizienten. Für das Produkt $Z = X \cdot Y$ unabhängiger Zufallsvariablen gilt:

$$Z^{(k)} = X^{(k)}Y^{(k)} \quad (\text{B.17})$$

Mit Gleichung (3.11) gilt

$$E[Z](c_Z^2 + 1) = E[X](c_X^2 + 1)E[Y](c_Y^2 + 1) \quad (\text{B.18})$$

Die Faktoren in Gleichung (B.16) werden als Messungen in getrennten Simulationsläufen bestimmt, so daß sie als unabhängig betrachtet werden können, wenn unabhängige Basisgeneratoren benutzt werden. Wendet man dies auf das Produkt nach Gleichung (B.15) an, so erhält man

$$\tilde{G}_B(\tilde{c}_B^2 + 1) = \quad (\text{B.19})$$

$$\tilde{H}_{m|m-1}(\tilde{c}_{m|m-1}^2 + 1)\tilde{H}_{m-1|m-2}(\tilde{c}_{m-1|m-2}^2 + 1) \dots \tilde{H}_{1|0}(\tilde{c}_{1|0}^2 + 1)\tilde{H}_0(\tilde{c}_0^2 + 1)$$

Für $\tilde{c}_B = d_{max}$ und $\tilde{c}_{j|j-1} = d_S$ ergibt sich der gesuchte Zusammenhang zwischen dem Gesamtfehler d_{max} und dem Fehler in den einzelnen Schritten d_S , siehe Gleichung (8.12):

$$\begin{aligned}
 (d_{max}^2 + 1) &= \\
 (\tilde{c}_{m|m-1}^2 + 1)(\tilde{c}_{m-1|m-2}^2 + 1) \cdots (\tilde{c}_{1|0}^2 + 1)(\tilde{c}_0^2 + 1) &= \quad (B.20) \\
 (d_S^2 + 1)^{m+1} &
 \end{aligned}$$

Bezeichnungen

<i>Bereich</i>	<i>Bedeutung</i>
\sim Messung	Kennzeichnung von Größen einer stat. Messung, z. B. $\tilde{G}(x)$
α Wartesysteme	Zufallsvariable: Ankunftsbelegung eines Warteraums (<i>arrival occupancy</i>)
β Wartesysteme	Zufallsvariable: Belegung eines Warteraums (<i>occupancy</i>)
γ_j RESTART/LRE	Korrelationsfaktor RESTART-Stützstellen $j = 0..m - 1$
δ Stochastik	Dirac-Stoß
ε Stochastik	relatives Konfidenz-Fehlermaß
η Wartesysteme	relatives Verkehrsangebot $\eta = \frac{E[\tau_b]}{S \cdot E[\tau_a]}$
κ SSMP	Korrelationsparameter
λ Wartesysteme	Intensität, Rate (meist bzgl. Zwischenankunftszeit)
μ Wartesysteme	Intensität, Rate (meist bzgl. Bediendauer)
μ Stochastik	$\mu = E[X]$ Mittelwert bzw. Erwartungswert von X
ν Stochastik	$\nu = E[Y]$ Mittelwert bzw. Erwartungswert von Y
ϱ LRE	lokaler Korrelationskoeffizient, $-1 \leq \varrho \leq 1$
σ Stochastik	Streuung, Standardabweichung, z. B. σ_X
σ^2 Stochastik	Varianz
τ Stochastik	Zeitzufallsvariable: allgemein
τ_a Stochastik	Zeitzufallsvariable: Zwischenankunftszeit
τ_b Stochastik	Zeitzufallsvariable: Bediendauer
τ_d Stochastik	Zeitzufallsvariable: Durchlaufzeit
τ_g Stochastik	Zeitzufallsvariable: busy period
τ_w Stochastik	Zeitzufallsvariable: Wartezeit
χ^2 Stochastik	χ^2 -Verteilung
ω Wartesysteme	Parameter

	<i>Bereich</i>	<i>Bedeutung</i>
A	Wartesysteme	Verkehrsangebot $A = \frac{E[\tau_b]}{E[\tau_a]}$
A_0, A_1, A_2	RESTART	Ereignisse
A_0, \dots, A_n, \dots	Wartesysteme	Folge von Ereignisabständen
a	Wartesysteme	Index des Ankunftsabstandes τ_a
$a_i \approx c_i$	LRE	Übergangshäufigkeit (<i>transition frequency</i>)
B	LRE	maximale Belegung
b	Wartesysteme	Index der Bediendauer τ_b
c_X	Stochastik	Variationskoeffizient von X
c_i	LRE	Übergangshäufigkeit (<i>transition frequency</i>)
cf_i	LRE	Korrelationsfaktor (<i>correlation factor</i>)
D	Wartesysteme	Kendall-Bez. für deterministische Verteilung
d	Wartesysteme	Index der Durchlaufzeit τ_d
d_{max}, d_B, d_S	LRE	LRE-Fehler (<i>limited relative error</i>), Fehlerschranke
$E[X]$	Stochastik	Erwartungswert von X
$F(x), F_X(x)$	Stochastik	Verteilungsfunktion (VF), (<i>cumulative distribution function</i>)
$f(x), f_X(x)$	Stochastik	Dichtefunktion (DF), Verteilungsdichtefunktion (<i>probability density function</i>)
G	Wartesysteme	Kendall-Bezeichnung für allg. Verteilung (General)
$G(x), G_X(x)$	Stochastik	komplementäre Verteilungsfunktion (KVF), $G(x) = 1 - F(x)$, (<i>compl. distribution function</i>)
G_i	RESTART/LRE	diskrete komplementäre Verteilungsfunktion
$G_{I_m} = G_B$	RESTART/LRE	letzte Stufe der kompl. VF, entspricht der Verlustwahrscheinlichkeit $P_B = G_B$
g	Wartesysteme	Index der <i>busy period</i> τ_g
H_j	RESTART/LRE	diskrete KVF an den RESTART-Stützstellen $H_j := G_{I_j}$, Niveau (<i>level</i>), $j = 0..m - 1$
h_i	RESTART/LRE	Zustandshäufigkeit (<i>state frequency</i>)

	<i>Bereich</i>	<i>Bedeutung</i>
I	RESTART/LRE	Stützpunkt für $m = 1$, $I_0 = I$
I_j	RESTART/LRE	Index der RESTART-Stützstellen (<i>intermediate points</i>) $j = 0..m - 1$
i	Allgemein	allgemeine Laufvariable, kontextabhängig
j	Allgemein	allgemeine Laufvariable, kontextabhängig
K	Simulation	Anzahl Prozessoren bei verteilter Simulation
K	Wartesysteme	Anzahl Aufträge in einem geschlossenen Wartenetz
k	LRE	k Level, LRE II
L	Wartesysteme	Zufallsvariable: Anzahl der Wartenden
M	Wartesysteme	Kendall-Bezeichnung für die negativ-exponentielle bzw. Poisson-Verteilung (M: Markoff)
m	RESTART/LRE	RESTART mit m Stufen und $m + 1$ Läufen (RESTART with m stages and $m + 1$ runs) $m = 0$ ohne RESTART, $m = 1$ einfacher RESTART
m	Markoff-Ketten	m Zeitschritte in einer diskreten Markoff-Kette
N	Wartesysteme	Anzahl Warteplätze G/G/S/N
N	Wartesysteme	Anzahl Knoten in einem Wartenetz
n	Simulation	Anzahl Meßwerte, Stichprobengröße (<i>sample size, number of trials</i>)
n_j	RESTART/LRE	Anzahl Versuche im Lauf j
n^*	RESTART/LRE	optimale Versuchsanzahl
$P\{\cdot\}$	Stochastik	Wahrscheinlichkeitsoperator
$P_\beta(j)$	Wartesysteme	Wahrscheinlichkeitsfunktion der Belegung β
P_B	Wartesysteme	Verlustwahrscheinlichkeit, auch P_L (<i>loss probability</i>)
$p_i = P(i)$	Wartesysteme	Wahrscheinlichkeit für Ereignis i Wahrscheinlichkeitsfunktion (WF) (<i>probability function, frequency function</i>)
p	Markoff-Ketten	2-gliedrige Markoff-Kette

	<i>Bereich</i>	<i>Bedeutung</i>
$p_{ji} = P(j i)$	Markoff-Ketten	Übergangswahrscheinlichkeit (<i>transition probability</i>)
$p_0(x), p_1(x)$	Markoff-Ketten	Übergangswahrscheinlichkeit 2-gliedrige Markoff-Kette
$Q(x)$	Markoff-Ketten	2-gliedrige Markoff-Kette, Zustandswahrscheinlichkeit linker Zustand $S_0(x)$ (<i>state probability</i>)
q	Wartesysteme	Quellenzahl
r	Allgemein	Indexvariable
$r_i = n - v_i$	LRE	Zustandshäufigkeit (<i>sum frequency</i>)
S	Wartesysteme	Anzahl Bedienstationen Kendall-Klassifikation
S	RESTART/LRE	Index für den Fehler in einem Schritt d_S
$S_0(x)$	Markoff-Ketten	2-gliedrige Markoff-Kette, linker Zustand
$S_1(x)$	Markoff-Ketten	2-gliedrige Markoff-Kette, rechter Zustand
s	Stochastik	Variable der Laplace-Transformierten $L(s)$
T	Allgemein	Zeitabschnitte, z. B. Rechenzeit
t	Allgemein	allg. Variable, z. B. bei der Integration (dt)
U	RESTART/LRE	Umrechnungsfaktor RESTART/LRE-Algorithmus
v_i	LRE	Zustandshäufigkeit (<i>sum frequency</i>)
w	Wartesysteme	Index der Wartezeit τ_w
X, Y, Z	Stochastik	allgemeine Zufallsvariable
$X^{(k)}$	Stochastik	k -tes Nullmoment von X
x, y, z	Stochastik	allgemeine Variablen, Werte bzw. Realisierungen der Zufallsvariablen X, Y, Z
x_i	Stochastik	Wertefolge einer Zufallsvariablen X
z	Allgemein	Exponent der Zehnerpotenz 10^{-z}

Abkürzungsverzeichnis

ACG	Additive Congruent Generator
ADSL	Asymmetric Digital Subscriber Line
AF	Arrival First
ARQ	Automatic Repeat Request
ATM	Asynchronous Transfer Mode
B-ISDN	Breitband-ISDN
BVSt	Bereichsvermittlungsstelle
CAC	Connection Admission Control
CCF	Call Control Function
CLIP	Calling Line Identification Presentation
CNCL	Communication Networks Class Library
COST	European COoperation in the field of Scientific and Technical research
DCA	Dynamic Channel Assignment
DECT	Digital Enhanced Cordless Telecommunication
DECT	Digital European Cordless Telecommunication (alte Bezeichnung)
DESI	DEct Simulator
DF	Departure First
DF	Dichtefunktion
DQDB	Distributed Queue Dual Bus
DSRC	Dedicated Short-Range Communication systems
DTMF	Dual Tone Multiple Frequency
ETSI	European Telecommunications Standards Institute
FCA	Fixed Channel Assignment
FDDI	Fiber Distributed Data Interface

FDMA	Frequency Division Multiple Access
FEC	Forward Error Correction
FIFO	First In First Out
GIST	Graphic Interactive Simulation result Tool
GOS	Grade Of Service
GSM	Global System for Mobile Communication
HDSL	High Speed Digital Subscriber Line
HIPERLAN	High PERformance Radio LAN
IKM	Intelligenter KommunikationsManager
INAP	Intelligent Network Application Protocol
IN	Intelligent Network
ISDN	Integrated Services Digital Network
IS	Infinite Server
ITU-T	International Telecommunication Union – Telecommunications standardization sector
ITU	International Telecommunication Union
KVF	Komplementäre Verteilungsfunktion
LAN	Local Area Network
LCFS	Last Come First Served
LCG	Linear Congruential Generator
LFG	Lagged-Fibonacci Generator
LJF	Longest Job First
LPT	Longest Processing Time first
LRE	Limited Relative Error
MAC	Medium Access Control
MAP	Markovian Arrival Process
MBS	Mobile Broadband System
MCG	Multiplicative Congruential Generator
MIMD	Multiple Instruction Multiple Data
MLFG	Mixed Lagged-Fibonacci Generator

MMPP	Markov Modulated Poisson Process
MTP	Message Transfer Part
NFS	Network File System
OBU	On Board Unit
PCS	Personal Communication System
PDA	Personal Digital Assistant
PDU	Protocol Data Unit
PSCS	Personal Services Communication Space
PS	Processor Sharing
PUI	Personal User Identification
RACE	Research and technology development in Advanced Communications technologies in Europe
RESTART	REpetitive Simulation Trials After Reaching Threshold
RFP	Radio Fixed Part
RLL	Radio in the Local Loop
RR	Round-Robin
RSE	Road Side Equipment
SCF	Service Control Function
SCP	Service Control Point
SDF	Service Data Function
SDL	Specification and Description Language
SDP	Service Data Point
SIB	Service Independent Building Block
SIMCO3++	Simulation of Mobile COmmunication systems, Version 3 in C++
SIMD	Single Instruction Multiple Data
SIRO	Service In Random Order
SJF	Shortest Job First
SPT	Shortest Processing Time first
SRF	Specialized Resource Function
SRPN	Shortest Rest Packet Number

SRPT	Shortest Remaining Processing Time first
SS.7	Signalling System Number 7
SSF	Service Switching Function
SSMP	Special Semi-Markov Process
STM	Synchronous Transfer Mode
TDD	Time Division Duplexing
TDMA	Time Division Multiple Access
TETRA	TransEuropean Trunked RAdio
THT	Token Holding Timer
TRT	Target Rotation Timer
TTRT	Target Token Rotation Time
UPT	Universal Personal Telecommunication
VF	Verteilungsfunktion
WVST	Weitvermittlungsstelle

PERSÖNLICHES LITERATURVERZEICHNIS – QUELLENVERZEICHNIS

- [1] C. Görg. Neuere Entwicklungen und Projekte auf dem Gebiet der Betriebssysteme. *Informatik Spektrum*, 4:68–81, 1981.
- [2] C. Görg. *Warteraum M/G/1: Die SRPT-Optimalstrategie im Vergleich mit der Zeitscheibenstrategie unter Berücksichtigung von Verwaltungszeiten*. Dissertation, RWTH Aachen, 1983.
- [3] C. Görg. Evaluation of Preemptive Queueing Strategies with Overhead and their Applicability in Data Networks. *Proceedings 10th International Teletraffic Congress (ITC), Montreal, Canada*, 3.3.7(1–7), 1983.
- [4] C. Görg, X.H. Pham. Improving Mean Delay in Data Communication Networks by New Combined Strategies based on the SRPT-Principle. *Proceedings 11th International Teletraffic Congress (ITC), Kyoto, Japan*, 5.2A–4(1–7), 1985.
- [5] C. Görg. Evaluation of the Optimal SRPT Strategy with Overhead. *IEEE Trans. Commun.*, COM-34(4):338–344, 1986.
- [6] C. Görg. HYPERNET – Funktions- und Protokollhandbuch. Technischer Bericht für das Forschungszentrum Jülich, 1987.
- [7] C. Görg. HYPERNET – Leistungsstudie. Technischer Bericht für das Forschungszentrum Jülich, 1988.
- [8] B. Kluth, C. Görg. Performance Analysis of a Cost Effective Multiprocessor for Teletraffic Simulation. In *12th Int. Teletraffic Congr. (ITC), Torino, Italy*, 4.3B.4(1–7), 1988.
- [9] L. Schmickler, C. Görg. Performance Evaluation of a New CSMA/CD Protocol based on the SRPT Principle. In *Proc. of GLOBECOM'89*, 924–929, Dallas, 1989.
- [10] C. Görg. Further Results on a New Combined Strategy Based on the SRPT-Principle. *IEEE Trans. Commun.*, COM-38(5):568–570, 1990.
- [11] C. Görg, F. Schreiber. Der LRE-Algorithmus für die statistische Auswertung korrelierter Zufallsdaten. *6. ASIM-Symposium, Wien 1990; Breitenecker, F. et al., Simulationstechnik, Vieweg, Wiesbaden*, 170–174, 1990.
- [12] C. Görg, B. Kluth, U. Salewski. SIC – a System for Stochastic Simulation in C++. In E. Mosekilde, editor, *Modelling and Simulation 1991, Proceedings of the 1991 European Simulation Multiconference, Copenhagen, 1991, The Society of Computer Simulation*, 413–414, 1991.

- [13] B. Kluth, C. Görg. Efficient simulation of network models in C++. In A. Jensen, V.B. Iversen, editors, *Teletraffic and Datatraffic, 13th Int. Teletraffic Congr. (ITC), Copenhagen, 1991, Elsevier Science Publ.*, 601–607, Amsterdam, 1991.
- [14] C. Görg, R. Lehnert. QLD – Queueing Literature Database. *MMB'93: Messung, Modellierung und Bewertung von Rechen- und Kommunikationssystemen, Kurzberichte und Werkzeugvorstellungen*, 1993.
- [15] M. Junius, C. Görg. Ein Programmpaket zur ereignisgesteuerten Simulation in C++. *MMB'93: Messung, Modellierung und Bewertung von Rechen- und Kommunikationssystemen, Kurzberichte und Werkzeugvorstellungen*, 1993.
- [16] M. Fröhlich, C. Görg, M. Guntermann, S. Kleier, G. Ratayczak, J. Sauer mann, P. Hourmann, M. Abramovici. Integration of Advanced Communication Services in the Personal Services Communication Space – A Realization Study. In *Proceedings 1st International Conference on Intelligence in Broadband Services and Networks (IS&N'93)*, II/1/1–12, Paris, November 1993.
- [17] M. Guntermann, C. Görg, M. Fröhlich, S. Kleier. Integration of Advanced Communication Services in the Personal Services Communication Space – A Realization Study. In *Mobile Kommunikationssysteme, 8. Aachener Kolloquium Signaltheorie, VDE-Verlag, Berlin*, 127–131, Aachen, März 1994.
- [18] F. Schreiber, C. Görg. Rare Event Simulation: a Modified RESTART-Method using the LRE-Algorithm. *Teletraffic and Datatraffic, Proceedings 14th ITC, Antibes, Juan-Les-Pins, France, June 6-10, 1994*, 787–796, 1994.
- [19] K. Keil, N. Niebert, C. Görg. Personal Services Communication Space – Ein Konzept zum Management mobiler Kommunikation. In *Congress I: Wireless Networks, Mobile Communications, 17th Online Congress*, Hamburg, Februar 1994.
- [20] M. Guntermann, C. Görg, M. Fröhlich, S. Kleier. IN based End-User Service Management for Advanced UPT. In *Broadband Island '94 - Connecting with the End-User, 3rd International Conference, Hamburg*, 193–201. North-Holland Elsevier Science B.V., Amsterdam, London, New York, Tokyo, June 1994.
- [21] S. Kleier, C. Görg, M. Fröhlich, M. Guntermann. IN Based Personal Message Service for Advanced UPT. In *Intelligence in Networks*, 377–378. 3rd International Conference (ICIN'94), Bordeaux, October 1994.
- [22] A. Folkestad, S. Kleier, C. Görg. Impact of UPT Services on Network Performance and Call Set-Up Times: Distribution of UPT Service Logic in SS7. In *Proceedings International Switching Symposium 1995*, 97–101. ISS'95 Conference, Berlin, April 1995.
- [23] C. Görg, S. Kleier, M. Guntermann, M. Fröhlich, H. Bisseling. A European Solution for Advanced UPT: Integration of Services for Personal Communication.

- In *Integrating Telecommunications and Distributed Computing – from Concept to Reality, Proceedings TINA'95 Conference, February 1995, Melbourne, Australia*, 603–617, February 1995.
- [24] S. Kleier, S. Tirney, C. Görg. A Fully Distributed Approach of Personalized Mobile Communication – Implementation and Performance Evaluation. In *Mobile Kommunikation, ITG-Fachtagung, ITG-Fachbericht 135, VDE-Verlag, Berlin*, 149–156, Neu-Ulm, September 1995.
 - [25] C. Görg. Kommunikationssysteme II. Vorlesungsskript, Universität Gesamthochschule Essen, 1995.
 - [26] C. Görg. Betriebliche Netze und Verbundsysteme. Vorlesungsskript, Universität Gesamthochschule Essen, 1994-1996.
 - [27] S. Kleier, C. Görg. Personal Communication System Realisations: Performance and Quality of Service Aspects on SS-No.7. In *Bringing Services to People, Proceedings 3rd International Conference on Intelligence in Services and Networks (IS&N '95)*, 213–224, Crete, October 1995.
 - [28] C. Görg (Editor). Grundlegende Probleme der Simulationstechnik. Workshop 17.-18. Januar 1996, Lehrstuhl Informatik III (Universität Würzburg), Lehrstuhl Kommunikationsnetze (RWTH Aachen), 1996.
 - [29] F. Schreiber, C. Görg. Stochastische Simulationstechnik mit Anwendungen auf Kommunikationsnetze. Vorlesungsskript, Lehrstuhl Kommunikationsnetze, RWTH Aachen, 1992-1996.
 - [30] F. Schreiber, C. Görg. Stochastic Simulation: a Simplified LRE-Algorithm for Discrete Random Sequences. *AEÜ*, 50:233–239, 1996.
 - [31] C. Görg, P. Decker, M. Junius, M. Lott. Praktikum: Stochastische Simulation von Datennetzen und ihren Protokollen. Praktikumsskript, Lehrstuhl Kommunikationsnetze, RWTH Aachen, 1991-1996.
 - [32] C. Görg, F. Schreiber. The RESTART/LRE Method for Rare Event Simulation. In *1996 Winter Simulation Conference*, 390–397, Coronado, California, USA, December 1996.
 - [33] C. Görg. How is Simulation used in Practical Experiments? In *1996 Winter Simulation Conference, Poster Session*, Coronado, California, USA, December 1996.
 - [34] C. Görg, M. Guntermann, S. Kleier. Future Systems for Personal Mobility Services: Design, Performance Evaluation, and Implementation. *IEEE JSAC on Personal Communications*, October, 1997.
 - [35] C. Görg. Simulation seltener Ereignisse mit dem mehrstufigen RESTART/LRE-Verfahren – oder – Wie man die Laufzeit von ATM-Simulationen von Jahren

auf Minuten verkürzen kann. *MMB'97: Messung, Modellierung und Bewertung von Rechen- und Kommunikationssystemen, Freiberg, Sept. 97, Informatik in Freiberg, Band 4, 4–11, 1997.*

- [36] B. Walke, C. Görg. Verbundprojekt: Schnelle Netze – Nutzung und Entwicklung der Datenübertragung mit Hoch- und Höchstgeschwindigkeit. *Informatik'97 – Informatik als Innovationsmotor, Aachen, Sept. 97, Informatik Aktuell, Springer Verlag, 1997.*

LITERATURVERZEICHNIS – QUELLENVERZEICHNIS

- [AE91] A. Altehage (Editor). *Digitale Vermittlungssysteme für Fernsprechen und ISDN (Digital Switching Systems for Telephony and ISDN)*. R. v. Decker's Verlag, G. Schenk, Heidelberg, 1991.
- [AG94] Deutsche Telekom AG. Geschäftsbericht Telekom 94 (Annual Report). DTAG, 1994.
- [AHH96] R. Andreassen, P.E. Heegaard, B.E. Helvik. Importance Sampling for speed-up simulation of heterogeneous MPEG sources. In *The 13th Nordic Teletraffic Seminar (NTS-13), August, 1996*, 190–203, Trondheim, Norway, 1996.
- [ATT90] ATT. *UNIX System V, AT&T C++ Language System, Release 2.1*. UNIX Software Operation, 1990.
- [BA82] G. Bolch, I.F. Akyildiz. *Analyse von Rechensystemen*. Teubner Studienbücher Informatik, 1982.
- [Baf95] M. Bafutto. *Modellierung, Verkehrsanalyse und Planung von Zentralkanal-Signalisierung im ISDN mit besonderer Berücksichtigung neuer Dienste der Mobilkommunikation und des Intelligenten Netzes*. Dissertation, Universität Stuttgart, Institut für Nachrichtenvermittlung und Datenverarbeitung, 63. Bericht über verkehrstheoretische Arbeiten, 1995.
- [Bar83] M. Barel. *Ein Multiprozessorsystem für die stochastische, ereignisorientierte Simulation*. Dissertation, RWTH Aachen, Lehrstuhl Datenfernverarbeitung, 1983.
- [Bay70] A.J. Bayes. Statistical Techniques for Simulation Models. *The Australian Computer Journal*, 2(4):180–184, 1970.
- [BCMP75] F. Baskett, K.M. Chandy, R.R. Muntz, F.G. Palacios. Open, Closed, and Mixed Networks of Queues with Different Classes of Customers. *Journal of the ACM*, 22(2):248–260, 1975.
- [BDM+89] R.F.B. Belanger, B. Donovan, K.L. Morse, S.V. Rice, D.B. Rockower. *MODSIM II Reference Manual*. CACI Products Company, La Jolla, California, 1989.
- [BDMN73] G.M. Birtwhistle, O.-J. Dahl, B. Myrhaug, K. Nygaard. *SIMULA BEGIN*. Studentlitteratur, Lund, Sweden (Chartwell-Bratt Ltd. Bromley, England), 1973.

- [Bei90] H. Beilner. Simulation (Folienskopen zur Vorlesung). Interne Berichte und Skripten an der Abteilung Informatik der Universität Dortmund (Universität Dortmund, Informatik IV, D-44221 Dortmund), 1990.
- [Bei91] H. Beilner. The Hierarchical Evaluation Tool HIT. Kurzvorträge und Werkzeugausstellung, Messung, Modellierung und Bewertung von Rechensystemen, 6. GI/ITG - Fachtagung MMB'91, 1991.
- [Bes91] T. Beshah. Simulationssystem SIC: Modellierung und Simulation eines Hochgeschwindigkeitsnetzes mit dem CRMA-Medienzugriffsverfahren. Diplomarbeit, RWTH Aachen, Lehrstuhl Datenfernverarbeitung, 1991.
- [Bey93] W. Beyers. Lastmodelle für multimediale Anwendungen und deren Realisierung in C++. Diplomarbeit, RWTH Aachen, Lehrstuhl Kommunikationsnetze, 1993.
- [BFS87] P. Bratley, B.L. Fox, L.E. Schrage. *A guide to simulation*. Springer, New York, 1987.
- [Böh96] S. Böhmer. *Entwurf eines ATM-basierten Funknetzes und Softwareentwurfsmethodik zur Implementierung*. Dissertation, RWTH Aachen, Aachener Beiträge zur Mobil- und Telekommunikation, Berichte des Lehrstuhls Kommunikationsnetze, Verlag der Augustinusbuchhandlung, A&BMT Band 5, 1996.
- [Bir79] G.M. Birtwhistle. *Demos - Discrete Event Modelling on Simula*. Macmillan: London, 1979.
- [Blo67] N. Blomqvist. The Covariance Function of the M/G/1 Queueing System. *Skand. Aktuarietidskr.*, 50:157–174, 1967.
- [Bol89] G. Bolch. *Leistungsbewertung von Rechensystemen*. Teubner, 1989.
- [Boo94] G. Booch. *Object-Oriented Analysis and Design with Applications*. The Benjamin/Cummings Publishing Company, Inc., Redwood City, California, 1994.
- [Bru72] S.L. Brumelle. A Generalization of $L = \lambda W$ to Moments of Queue Length and Waiting Times. *Operations Research*, 20(6):1127–1136, 1972.
- [BS71] I.N. Bronstein, K.A. Semendjajew. *Taschenbuch der Mathematik*. Verlag Harri Deutsch, Zürich, 1971.
- [Cou77] P.J. Courtois. *Decomposability, Queueing and Computer System Applications*. Academic Press, New York, 1977.

- [DB78] P.J. Denning, J.P. Buzen. The Operational Analysis of Queuing Network Models. *Computing Surveys*, 10(3):225–261, 1978.
- [DD90] W. Ding, P. Decker. Waiting Time Distribution of a Discrete SSMP/G/1 Queue and its Implications in ATM Systems. In *7th ITC Seminar, Morristown, Morristown, New Jersey, USA, October, 9-11, 1990*.
- [Dev86] L. Devroye. *Non-Uniform Random Variate Generation*. Springer, New York, 1986.
- [Din91a] W. Ding. *Korrelierte Zufallsprozesse in Wartesystemen von Kommunikationsnetzen*. Dissertation, RWTH Aachen, Lehrstuhl Datenfernverarbeitung, 1991.
- [Din91b] W. Ding. A Unified Correlated Input Process Model for Telecommunication Networks. *Teletraffic and Datatraffic, Proceedings 13th ITC, Copenhagen, June 19-26, 1991; Elsevier, North Holland*, 539–544, 1991.
- [DS90] W. Ding, F. Schreiber. Local Correlation Properties of Random Sequences Generated by Queuing System M/M/1. *AEÜ*, 44:384–389, 1990.
- [ETCRE92] Systems TC-RES ETSI Technical Committee Radio Equipment. *A guide to the DECT features that influence the traffic capacity and the maintenance of a high radio link quality, including the results of simulations*. ETSI Technical Report. European Telecommunication Standards Institute, Sophia Antipolis, August 1992.
- [ETCRE94] Systems TC-RES ETSI Technical Committee Radio Equipment. *Radio in the Local Loop (RLL)*. ETSI Technical Report. European Telecommunication Standards Institute, Sophia Antipolis, November 1994.
- [ETS91a] ETSI. *Digital European Cordless Telecommunications, Part 2: Physical Layer (DE/RES 3001-2)*. ETSI Technical Report (Draft Standard). European Telecommunication Standards Institute, Sophia Antipolis, August 1991.
- [ETS91b] ETSI. *Digital European Cordless Telecommunications, System Description Document (DR/RES 3004)*. ETSI Technical Report (Draft Standard). European Telecommunication Standards Institute, Sophia Antipolis, Juni 1991.
- [ETS91c] ETSI. *Radio Equipment and Systems, Digital European Cordless Telecommunications, Draft prETS 300 175*. ETSI Technical

- Report (Draft Standard), Part 1-6. European Telecommunication Standards Institute, Sophia Antipolis, June 1991.
- [ETS95] ETSI. *Wireless Base Station, Radio Equipment and Systems, Digital European Cordless Telecommunications, Draft prETS 300 175-10*. ETSI Technical Report (prStandard). European Telecommunication Standards Institute, Sophia Antipolis, January 1995.
- [Fis78] G.S. Fishman. *Principles of Discrete Event Simulation*. J. Wiley & Sons, New York, 1978.
- [FKG95] A. Folkestad, S. Kleier, C. Görg. Impact of UPT Services on Network Performance and Call Set-Up Times: Distribution of UPT Service Logic in SS7. In *Proceedings International Switching Symposium 1995*, 97–101. ISS'95 Conference, Berlin, April 1995.
- [FMH93] W. Fischer, K.S. Meier-Hellstern. The MMPP Cookbook. *Performance Evaluation*, 18(2):149–171, 1993.
- [Fra90] M.R. Frater. *Estimation of the Statistics of Rare Events in Data Communications Systems*. PhD thesis, The Australian National University, 1990.
- [Fuß97] O. Fuß. Validierung von Dienstgüteparametern von B-ISDN/ATM-Referenzverbindungen durch das RESTART/LRE-Verfahren zur Simulation seltener Ereignisse. Diplomarbeit, RWTH Aachen, Lehrstuhl Kommunikationsnetze, 1997.
- [Fuj90] R.M. Fujimoto. Parallel Discrete Event Simulation. *Communications of the ACM*, 33, 10:30–53, 1990.
- [GDJL96] C. Görg, P. Decker, M. Junius, M. Lott. Praktikum: Stochastische Simulation von Datennetzen und ihren Protokollen. Praktikumskript, Lehrstuhl Kommunikationsnetze, RWTH Aachen, 1991-1996.
- [GE96] C. Görg (Editor). Grundlegende Probleme der Simulationstechnik. Workshop 17.-18. Januar 1996, Lehrstuhl Informatik III (Universität Würzburg), Lehrstuhl Kommunikationsnetze (RWTH Aachen), 1996.
- [GE97] C. Görg (Editor). Rare Event Simulation. Workshop 28.-29. August 1997, Lehrstuhl Kommunikationsnetze (RWTH Aachen), 1997.
- [GGK97] C. Görg, M. Guntermann, S. Kleier. Future Systems for Personal Mobility Services: Design, Performance Evaluation, and Implementation. *IEEE JSAC on Personal Communications*, October, 1997.
- [GHSZ96] P. Glasserman, P. Heidelberger, P. Shahabuddin, T. Zajic. Splitting for Rare Event Simulation: Analysis of Simple Cases. In *1996 Win-*

- ter Simulation Conference*, 302–308, Coronado, California, USA, December 1996.
- [GKG⁺95] C. Görg, S. Kleier, M. Guntermann, M. Fröhlich, H. Bisseling. A European Solution for Advanced UPT: Integration of Services for Personal Communication. In *Integrating Telecommunications and Distributed Computing – from Concept to Reality, Proceedings TIN A'95 Conference, February 1995, Melbourne, Australia*, 603–617, February 1995.
- [GKS91] C. Görg, B. Kluth, U. Salewski. SIC – a System for Stochastic Simulation in C++. In E. Mosekilde, (Ed.), *Modelling and Simulation 1991, Proceedings of the 1991 European Simulation Multiconference, Copenhagen, 1991, The Society of Computer Simulation*, 413–414, 1991.
- [GKZH94] R. German, C. Kelling, A. Zimmermann, G. Hommel. TimeNET – A Toolkit for Evaluating Non-Markovian Stochastic Petri Nets. Technischer Bericht 1994-19, Technische Universität Berlin, 1994. (erscheint in *Performance Evaluation*).
- [Gly96] P. Glynn. Some New Results on the Initial Transient Problem. Department of Operations Research, Stanford University, Stanford, CA 94305-4022, USA, 1996.
- [GN67] W.J. Gordon, G.F. Newell. Closed Queueing Systems with Exponential Servers. *Operations Research*, 15(2):254–265, 1967.
- [GNU97] GNU Project. <http://www.gnu.ai.mit.edu>, 1997.
- [Gör83] C. Görg. *Warteraum M/G/1: Die SRPT-Optimalstrategie im Vergleich mit der Zeitscheibenstrategie unter Berücksichtigung von Verwaltungszeiten*. Dissertation, RWTH Aachen, 1983.
- [Gör87] C. Görg. HYPERNET – Funktions- und Protokollhandbuch. Technischer Bericht für das Forschungszentrum Jülich, 1987.
- [Gör88] C. Görg. HYPERNET – Leistungsstudie. Technischer Bericht für das Forschungszentrum Jülich, 1988.
- [Gör90] C. Görg. Further Results on a New Combined Strategy Based on the SRPT-Principle. *IEEE Trans. Commun.*, COM-38(5):568–570, 1990.
- [Gör96] C. Görg. How is Simulation used in Practical Experiments? In *1996 Winter Simulation Conference, Poster Session*, Coronado, California, USA, December 1996.
- [GS90] C. Görg, F. Schreiber. Der LRE-Algorithmus für die statistische Auswertung korrelierter Zufallsdaten. *6. ASIM-Symposium, Wien*

- 1990; Breitenecker, F. et al., *Simulationstechnik*, Vieweg, Wiesbaden, 170–174, 1990.
- [GS96] C. Görg, F. Schreiber. The RESTART/LRE Method for Rare Event Simulation. In *1996 Winter Simulation Conference*, 390–397, Colorado, California, USA, December 1996.
- [Gud85] M. Gude. Concept for a High Performance Random Number Generator based on Physical Random Phenomena. *Frequenz*, 39:187–190, 1985.
- [Gud87] M. Gude. *Ein quasi-idealer Gleichverteilungsgenerator basierend auf physikalischen Zufallsphänomenen*. Dissertation, RWTH Aachen, Lehrstuhl Datenfernverarbeitung, 1987.
- [Gun96] M. Guntermann. *Universelle Benutzermobilität auf der Basis des Intelligenten Netzes – Entwurf, Bewertung und Implementierung*. Dissertation, RWTH Aachen, Aachener Beiträge zur Mobil- und Telekommunikation, Berichte des Lehrstuhls Kommunikationsnetze, Verlag der Augustinusbuchhandlung, ABMT Band 6, 1996.
- [Hau90] A. Haugen. Simulation des DQDB-Kanalzugriffsverfahrens mit dem Simulationssystem SIC. Diplomarbeit, RWTH Aachen, Lehrstuhl Datenfernverarbeitung, 1990.
- [Hau91] D. Haupt. Erneuerung der zentralen Rechnerversorgung mit Installation eines zweiten Landesvektorrechners und Aufbau eines Datenkommunikationsnetzes an der RWTH Aachen. Bericht 89/90, 1. März 1991, Rechenzentrum RWTH Aachen, 1991.
- [Hee95] P.E. Heegaard. Speed-up Techniques for Simulation. *Elektronikk ISSN 0085-7130*, 91(2-3):195–207, 1995.
- [Hel95] P. Hellekalek. Inversive Pseudorandom Number Generators: Concept, Results and Links. In *Proceedings of the 1995 Winter Simulation Conference*, 255–262, Arlington, Virginia, USA, 1995.
- [Her94] C. Herrmann. The Local Correlation Properties of some Discrete-Time Finite Queues with Semi-Markovian Input. *AEÜ*, 48:244–255, 1994.
- [Her95] C. Herrmann. *Stochastische Modelle für ATM-Konzepte*. Dissertation, RWTH Aachen, Aachener Beiträge zur Mobil- und Telekommunikation, Berichte des Lehrstuhls Kommunikationsnetze, Verlag der Augustinusbuchhandlung, ABMT Band 1, 1995.
- [HHK+91] G. Hasler, S. Heppner, B. Kluth, U. Krüger, C. Görg, D. Pesch. *Das Simulationssystem SIC, Benutzeranleitung*. Lehrstuhl Kommunikationsnetze, RWTH Aachen, 1991.

- [HK79] A.C.M. Hopmans, J.P.C. Kleijnen. Importance Sampling in Systems Simulation: a Practical Failure? In *Mathematics and Computers in Simulation XXI*, 209–220, North-Holland, 1979.
- [HO86] J.L. Hammond, P.J.P. O'Reilly. *Performance Analysis of Local Computer Networks*. Addison-Wesley Publishing Company, Reading, Mass., 1986.
- [HP75] N.A.J. Hastings, J.B. Peacock. *Statistical Distributions*. Butterworth, London, 1975.
- [HW83] F. Hoßfeld, P. Weidner. Parallele Algorithmen. *Informatik-Spektrum*, 6:142–154, 1983.
- [Jab92] B. Jabbari. Intelligent Network Concepts in Mobile Communications. *IEEE Communication Magazine*, 64–35, Feb. 1992.
- [Jac63] J.R. Jackson. Jobshop-Like Queueing Systems. *Management Science*, 10(1):131–142, 1963.
- [Jai68] N.K. Jaiswal. *Priority Queues*. Mathematics in Science and Engineering, Volume 50, Academic Press, New York, London, 1968.
- [Jai91] R. Jain. *The Art of Computer Systems Performance Analysis – Techniques for Experimental Design, Measurement, Simulation, and Modeling*. John Wiley & Sons, New York, 1991.
- [Jam90] F. James. A Review of Pseudorandom Number Generators. *Computer Physics Communications*, 60:329–344, 1990.
- [JG93] M. Junius, C. Görg. Ein Programmpaket zur ereignisgesteuerten Simulation in C++. *MMB'93: Messung, Modellierung und Bewertung von Rechen- und Kommunikationssystemen, Kurzberichte und Werkzeugvorstellungen*, 1993.
- [JSBP96] M. Junius, M. Stepler, M. Büter, D. Pesch. *CNCL – Communication Networks Class Library*. Communication Networks, Aachen University of Technology, Januar 1996.
- [K+97] P.J. Kühn et al. ITG-Fachgruppe 5.2.1 Verkehrstheorie, Begriffe der Nachrichtenverkehrstheorie 5.2-03. *NTZ*, 1997. (in Vorbereitung).
- [Kel96] C. Kelling. A Framework for Rare Event Simulation of Stochastic Petri Nets using “RESTART”. In *1996 Winter Simulation Conference*, 317–324, Coronado, California, USA, December 1996.
- [KG91] Z.V. Karian, R. Goyal. Random Number Generation and Testing. Technical report, Department of Mathematical Sciences, Denison University, Granville Ohio 43023, Email: karian@denison.edu, goyan_r@denison.edu, 1991.

- [KG95] S. Kleier, C. Görg. Personal Communication System Realisations: Performance and Quality of Service Aspects on SS-No.7. In *Bringing Services to People, Proceedings 3rd International Conference on Intelligence in Services and Networks (IS&N '95)*, 213–224, Crete, October 1995.
- [KGZH95] C. Kelling, R. German, A. Zimmermann, G. Hommel. TimeNET – ein Werkzeug zur Modellierung mit zeiterweiterten Petri-Netzen. *Informationstechnik und Technische Informatik (it+ti)*, 37(3):21–28, 1995.
- [Küh79] P.J. Kühn. Approximate Analysis of General Queueing Networks by Decomposition. *IEEE Transactions on Communications*, COM-27, 1:113–126, 1979.
- [KHM87] M.J. Karol, M.G. Hluchyj, S.P. Morgan. Input Versus Output Queueing on a Space Division Packet Switch. *IEEE Trans. Com.*, 35(12):1347–1356, 1987.
- [Kle74] J.P.C. Kleijnen. *Statistical Techniques in Simulation*, volume I. Marcel Dekker, New York, 1974.
- [Kle75] L. Kleinrock. *Queueing Systems (Theory)*, volume 1. John Wiley, New York, 1975.
- [Kle76] L. Kleinrock. *Queueing Systems (Applications)*, volume 2. John Wiley, New York, 1976.
- [Kle96] S. Kleier. *Neue Konzepte zur Unterstützung von Mobilität in Telekommunikationsnetzen*. Dissertation, RWTH Aachen, Aachener Beiträge zur Mobil- und Telekommunikation, Berichte des Lehrstuhls Kommunikationsnetze, Verlag der Augustinusbuchhandlung, ABMT Band 7, 1996.
- [Klu90] B. Kluth. *Multiprozessorarchitekturen mit funktionsorientierter Parallelisierung für die stochastische Simulation*. Dissertation, RWTH Aachen, Lehrstuhl Datenfernverarbeitung, 1990.
- [KM53] H. Kahn, A.W. Marshall. Methods of Reducing Sample Size in Monte Carlo Computations. *J. Oper. Res. Society of America*, 1:263–278, 1953.
- [KMN94] H.-J. Kugler, A. Mullery, N. Niebert, (Eds.). *2nd International Conference on Intelligence in Services and Networks (IS&N '94): Towards a Pan-European Telecommunication Service Infrastructure*, Lecture Notes in Computer Science 851, Aachen, Germany, September 1994. Springer Verlag, Berlin, Heidelberg, New York.

- [Knu81] D.E. Knuth. *Seminumerical Algorithms*, volume 2 of *The Art of Computer Programming*. Addison-Wesley, Reading, Mass., second edition, 1981.
- [Koc82] M. Kochan. Entwicklung eines Pascal-Simulationssystems zur Untersuchung der F-LIFO II Warteprozedur und des unterbrechenden Prioritätensystems mit Verwaltungszeit. Diplomarbeit, RWTH Aachen, Lehrstuhl Datenfernverarbeitung, 1982.
- [Krö95] H. Kröner. *Verkehrssteuerung in ATM-Netzen – Verfahren und verkehrstheoretische Analysen zur Zellpriorisierung und Verbindungsannahme*. Dissertation, Universität Stuttgart, Institut für Nachrichtenvermittlung und Datenverarbeitung, 1995.
- [Kre73] E. Kreyszig. *Statistische Methoden und ihre Anwendungen*. Vandenhoeck & Rupprecht, Göttingen, 1973.
- [Kre86] W. Kreutzer. *System Simulation Programming Styles and Languages*. Addison-Wesley Publishing Co., 1986.
- [Lan92] H. Langendörfer. *Leistungsanalyse von Rechensystemen – Messen, Modellieren, Simulation*. Carl Hanser Verlag, München, 1992.
- [LB91] J. Le Boudec. An Efficient Solution Method for Markov Models of ATM Links with Loss Priorities. *IEEE J. Select. Areas Commun.*, 9:408–417, 1991.
- [L'E96] P. L'Ecuyer. Random Number Generators. *Encyclopedia of Operations Research and Management Science*, to appear:1–11, 1996.
- [Leh79] R. Lehnert. *Ein modularer Prozessor zur Simulation von Zufallsnetzwerken*. Dissertation, Lehrstuhl Datenfernverarbeitung, RWTH Aachen, 1979.
- [Leh93] R. Lehnert. Distributed Simulation for Performance Evaluation of Broadband Communication Networks. *AEÜ Hirzel Verlag*, 47, 5/6:420–425, 1993.
- [Lie82] O. Lie. Simulation des Datenverkehrs in Datennetzen: Simulation einer bestehenden Netzkonfiguration mit gemessenen Eingabedaten. Diplomarbeit, RWTH Aachen, Lehrstuhl Datenfernverarbeitung, 1982.
- [Lin96] Y.-B. Lin. A Cache Approach for Supporting Life-Time UPT Numbers. *Wireless Networks*, J.C. Baltzer AG, Science Publishers, 2, 2:155–160, 1996.
- [LK91] A.M. Law, W.D. Kelton. *Simulation Modeling & Analysis*. McGraw-Hill, 1991.

- [LO89] P.A.W. Lewis, E.J. Orav. *Simulation Methodology for Statisticians, Operations Analysts, and Engineers*. Wadsworth & Brooks/Cole Advanced Books and Software, Belmont, California, 1989.
- [Mar85] G. Marsaglia. A Current View of Random Number Generators. In *Computer Science and Statistics, 16th Symposium on the Interface*, Elsevier, Amsterdam, 1985.
- [Meh92] S. Mehrholz. Simulationssystem SIC: Vergleich von Mittelwertanalyse und Simulation für das FDDI – Zugriffsprotokoll. Studienarbeit, RWTH Aachen, Lehrstuhl Kommunikationsnetze, 1992.
- [Meh96] S. Mehrholz. Intelligente Netze: Analyse und Optimierung eines OSI-Anwendungsschicht-Protokolls für die Benutzer-Netzchnittstelle: Teil I+II. Diplomarbeit, RWTH Aachen, Lehrstuhl Kommunikationsnetze, 1996.
- [Mei84] K.S. Meier. *A Statistical Procedure for Fitting Markov-Modulated Poisson Processes*. PhD thesis, University of Delaware, Newark, Delaware, 1984.
- [MIL97] MIL3. *OPNET Simulation Tool*. MIL3, Washington D.C., USA, 1997. <http://www.mil3.com>.
- [Mor58] P.M. Morse. *Queues, Inventories, and Maintenance*. Wiley, New York, 1958.
- [Mor68] D. Morgenstern. *Einführung in die Wahrscheinlichkeitsrechnung und mathematische Statistik*. Springer, Berlin, 1968.
- [MS90] A. Modarressi, R. Skoog. Signaling System No. 7: A Tutorial. *IEEE Communication Magazine*, 19–35, July 1990.
- [NA-92a] NA-71101. *Universal Personal Telecommunications (UPT) Phase 1 (Restricted UPT Service Scenario) Service Aspects: Guidelines*. ETSI Technical Report (Draft). European Telecommunication Standards Institute, Networks Aspects (NA), Sophia Antipolis, July 1992.
- [NA-92b] NA-71207. *Universal Personal Telecommunications (UPT) Phase 1: Man-Machine Interface*. ETSI Technical Report (Draft). European Telecommunication Standards Institute, Networks Aspects (NA), Sophia Antipolis, July 1992.
- [Neu81] M.F. Neuts. *Matrix-Geometric Solutions in Stochastic Models, An Algorithmic Approach*. The Johns Hopkins University Press, Baltimore and London, 1981.
- [Pag91] B. Page. *Diskrete Simulation, Eine Einführung mit Modula-2*. Springer, Berlin, 1991.

- [Pap91] A. Papoulis. *Probability, Random Variables, and Stochastic Processes*. McGraw-Hill, New York, 1991.
- [Paw90] K. Pawlikowski. Steady-State Simulation of Queueing Processes: A Survey of Problems and Solutions. *ACM Computing Surveys*, 22, 2:123–170, 1990.
- [PFJ95] R. Popp, M. Fröhlich, N. Jefferies. Security Services for Telecommunications Users. In *Bringing Services to People*, 28–39. 3rd International Conference on Intelligence in Broadband Services and Networks (IS&N'95), Crete, October 1995.
- [PH⁺96] D. Plaßmann, M. Horneffer et al. *GIST Manual - Graphical Interactive Simulation result Tool*. RWTH Aachen, Lehrstuhl für Kommunikationsnetze, 1996. <http://www.comnets.rwth-aachen.de/doc/cncl.html>.
- [Ple96] C. Plenge. *Leistungsbewertung öffentlicher DECT-Systeme*. Dissertation, RWTH Aachen, Aachener Beiträge zur Mobil- und Telekommunikation, Berichte des Lehrstuhls Kommunikationsnetze, Verlag der Augustinusbuchhandlung, ABMT Band 10, 1996.
- [PM88] S.K. Park, K.W. Miller. Random Number Generators: Good Ones are Hard to Find. *Communications of the ACM*, 31:1192–1201, 1988.
- [Por93] M. Portz. *Verbindungsnetzwerke als Permutationsgeneratoren im kryptologischen Kontext*. Dissertation, RWTH Aachen, 1993.
- [PS93] R. Perera, F. Schreiber. Distribution function and local correlation of the delay time in queueing system M/G/1-SRPT. *AEÜ*, 47:108–110, 1993.
- [Rü895] M. Rübmann. Corporate Networks: Analysis of New Mobile Communication Services and Interfaces to the Public Network based on Intelligent Network Technology. Diplomarbeit, RWTH Aachen, Lehrstuhl Kommunikationsnetze, 1995.
- [Rap96] T.S. Rappaport. *Wireless Communications – Principles & Practice*. Prentice Hall PTR, Upper Saddle River, New Jersey, 1996.
- [RBGW94] C.-H. Rokitansky, B. Bjelajac, A. Guntsch, C. Wietfeld. SIMCO 3++: SIMulation of Mobile COMMunications for Performance Evaluation of Cellular Networks, Short Range and Satellite Communications. In *Proceedings RACE Mobile Workshop '94*, 578–584, Amsterdam, Netherlands, 1994.
- [Rec93a] Q.1200 ITU-T Recommendation. *Q-Series Intelligent Network Recommendation Structure*. Study Group 11: Telephone Switching

- and Signalling. International Telecommunication Union, Geneva, March 1993.
- [Rec93b] Q.706 ITU-T Recommendation. *Signalling System No. 7 - Message Transfer Part Signalling Performance*. Study Group XI: Switching and Signalling. International Telecommunication Union, Geneva, March 1993.
- [Rec93c] Q.766 ITU-T Recommendation. *Specifications of Signalling System No. 7 ISDN User Part - Performance Objectives in the Integrated Services Digital Network Application*. Study Group XI: Switching and Signalling. International Telecommunication Union, Geneva, March 1993.
- [Rec96] I.356 ITU-T Recommendation. *Integrated Services Digital Network (ISDN), Overall Network Aspects and Functions, B-ISDN ATM Layer Cell Transfer Performance*. Working Party 4/13, Temporary Document 53-E, 05/06 Draft Rev. International Telecommunication Union, Geneva, May 1996.
- [Rey75] J.F. Reynolds. The Covariance Structure of Queues and Related Processes – a Survey of Recent Work. *Adv. Appl. Prob.*, 7:383–415, 1975.
- [Ric92] M. Richter. *Ein Rauschgenerator zur Gewinnung von quasi-idealen Zufallszahlen für die stochastische Simulation*. Dissertation, RWTH Aachen, Lehrstuhl Datenfernverarbeitung, 1992.
- [Ric93] M. Richter. Puran2: Ein Zufallszahlengenerator zur Erzeugung von quasi-idealen Zufallszahlen aus elektronischem Rauschen. In B. Walke, O. Spaniol, (Eds.), *Messung, Modellierung und Bewertung von Rechen- und Kommunikationssystemen, Kurzberichte und Werkzeuqvorstellungen*, 49–62, Aachen, September 1993. Verlag der Augustinus Buchhandlung.
- [RL80] M. Reiser, S.S. Lavenberg. Mean-Value Analysis of Closed Multichain Queueing Networks. *Journal of the ACM*, 27:313–322, 1980.
- [Ros92] Ortwin Rose. *SimPlusPlus, an Object-Oriented Language for Process-Oriented Discrete-Event Simulation*. Institute of Telematics, University of Karlsruhe, 1992.
- [RW89] V. Ramaswami, W. Willinger. Efficient Traffic Performance Strategies for Packet Multiplexers. In *ITC Specialist Seminar*, 401–407, Adelaide, Australia, Sept. 25–29, 1989.
- [Sac92] L. Sachs. *Angewandte Statistik*. Springer Verlag, Berlin, 1992.

- [Sch80] F. Schreiber. Improved Simulation by Application of the Objective Bayes-Statistics. *AEÜ*, 34:243–249, 1980.
- [Sch84] F. Schreiber. Time Efficient Simulation: the LRE-Algorithm for Producing Empirical Distribution Functions with Limited Relative Error. *AEÜ*, 38:93–98, 1984. (see also: Proc. 10th Int. Teletraffic Congr. (ITC), Montreal, 1983).
- [Sch86] F. Schreiber. Measurement of the Correlation Coefficient of the 2-Node Markov Chain. *AEÜ*, 40:402–405, 1986.
- [Sch87a] F. Schreiber. The Empirical Stationary Distribution Function of Markovian Correlated Random Sequences. *AEÜ*, 41:257–263, 1987.
- [Sch87b] F. Schreiber. Measurement of the Stationary State Probabilities of the 2-Node Markov Chain. *AEÜ*, 41:117–122, 1987.
- [Sch88a] F. Schreiber. Effective Control of Simulation Runs by a New Evaluation Algorithm for Correlated Random Sequences. *AEÜ*, 42:347–354, 1988. (see also: Proc. 12th Int. Teletraffic Congr. (ITC), Torino, 1988, p. 4.3B.1.1–9).
- [Sch88b] F. Schreiber. *Vorlesung Datenfernverarbeitung: Verkehrstheorie*. Lehrstuhl Kommunikationsnetze, RWTH, Aachen, 1988.
- [Sch90] O. Schlageter. Zurück in die Zukunft: PC-Simula Release 108. *c't*, 168–173, 1990.
- [Sch91a] H.-J. Schlebusch. *Importance sampling Techniken zur effizienten Simulation*. Dissertation, RWTH Aachen, Lehrstuhl für Elektrische Regelungstechnik, 1991.
- [Sch91b] L. Schmickler. *Einsatzmöglichkeiten der Optimalstrategie SRPT in Lokalen Rechnernetzen*. Dissertation, RWTH Aachen, Lehrstuhl Datenfernverarbeitung, 1991.
- [Sch92] L. Schmickler. MEDA: Mixed Erlang Distributions as Phase-Type Representations of Empirical Distribution Functions. *Commun. Statist. -Stochastic Models*, 131–156, 1992.
- [Sch94] F. Schreiber. Local Correlation Properties of the Finite Buffer Queueing System M/M/1/N. *AEÜ*, 48:63–67, 1994.
- [SG89] L. Schmickler, C. Görg. Performance Evaluation of a New CSMA/CD Protocol based on the SRPT Principle. In *Proc. of GLOBECOM'89*, 924–929, Dallas, 1989.
- [SG94] F. Schreiber, C. Görg. Rare Event Simulation: a Modified RESTART-Method using the LRE-Algorithm. *Teletraffic and Data-*

- traffic, Proceedings 14th ITC, Antibes, Juan-Les-Pins, France, June 6-10, 1994, 787-796, 1994.*
- [SG96a] F. Schreiber, C. Görg. Stochastische Simulationstechnik mit Anwendungen auf Kommunikationsnetze. Vorlesungsskript, Lehrstuhl Kommunikationsnetze, RWTH Aachen, 1992-1996.
- [SG96b] F. Schreiber, C. Görg. Stochastic Simulation: a Simplified LRE-Algorithm for Discrete Random Sequences. *AEÜ*, 50:233-239, 1996.
- [Sha81] R.E. Shannon. Test for the Verification and Validation of Computer Simulation Models. In *1981 Winter Simulation Conference (Atlanta, Ga., USA), ACM/IEEE*, 573-577, New York, 1981.
- [SI95] M.N.O. Sadiku, M. Ilyas. *Simulation of Local Area Networks*. CRC press, Boca Raton, Florida, 1995.
- [Sie91] H.-J. Siebert. *Simulation zeitdiskreter Systeme*. Oldenbourg, 1991.
- [SIM87] SIMULA. SIMULA Standard. Svensk Standard SS 63 61 14 Data Processing-Programming Languages-SIMULA, 1987.
- [SKMLK95] K. Schmidt, U. Killat, F. Mayer-Lindenberg, P. Kraft. A Parallel ATM-Network Simulator. In *Proceedings International Switching Symposium 1995, ISS'95*, Vol.1:304-308, Berlin, Germany, April 1995.
- [Str89] B. Stroustrup. *Die C++-Programmiersprache*. Addison-Wesley (Deutschland), 1989.
- [Sug81] O. Sugiri. Simulation von Paketvermittlungsnetzen mit austauschbaren Protokollmoduln. Diplomarbeit, RWTH Aachen, Lehrstuhl Datenfernverarbeitung, 1981.
- [Tan91] M. Tangemann. A Mean Value Analysis for Throughputs and Waiting Times of the FDDI Timed Token Protocol. In A. Jensen, V.B. Iversen, (Eds.), *Teletraffic and Datatraffic*, 173-179, Elsevier Science Publ., Amsterdam, 1991. 13th Int. Teletraffic Congr. (ITC), Copenhagen, 1991.
- [Tau65] R.C. Tausworthe. Random Numbers Generated by Linear Recurrence Mod Two. *Mathematics of Computation*, 19:201-209, 1965.
- [TG89] H. Tuchel, C. Görg. SIMPAS+: Simulation in PASCAL. Technical report, Lehrstuhl DFV, RWTH Aachen, 1989.
- [TG96] P. Tran-Gia. *Analytische Leistungsbewertung verteilter Systeme*. Springer, Berlin, 1996.
- [Thö94] J. Thörner. *Intelligent Networks*. Artech House, Boston, London, 1994.

- [vA95] H. van As. Photonische Kommunikationsnetze – Universelle Infrastruktur für die Informationsgesellschaft. Vortrag Lehrstuhl Kommunikationsnetze 1. Dezember 1995, 1995.
- [VAMMGFC94] M. Villén-Altamirano, A. Martínez-Marrón, J. Gamo, F. Fernández-Cuesta. Enhancement of the Accelerated Simulation Method RESTART by considering Multiple Thresholds. In *Proceedings 14th International Teletraffic Congress*, 797–810. North-Holland, 1994.
- [VAVA91] M. Villén-Altamirano, J. Villén-Altamirano. Accelerated simulation of rare events using RESTART method with hysteresis. In J. Filipiak, (Ed.), *Telecommunication Services for Developing Economies.*, 675–686. Proceedings of the ITC Specialist Seminar, Cracow, Poland, April 1991, Elsevier, Amsterdam, Netherlands, 1991.
- [VAVA94] M. Villén-Altamirano, J. Villén-Altamirano. RESTART: A Straightforward Method for Fast Simulation of Rare Events. In *Proceedings of the 1994 Winter Simulation Conference*, 282–289, 1994.
- [Vor96] U. Vornefeld. Analytische und simulative Bewertung eines adaptiven Verfahrens zur Verbindungsannahmesteuerung in ATM-Netzen. Studienarbeit, RWTH Aachen, Lehrstuhl Kommunikationsnetze, 1996.
- [Wal88] J. Walrand. *An Introduction to Queueing Networks*. Prentice Hall, Englewood Cliffs, New Jersey, 1988.
- [Wal92] B. Walke. *Datenfernverarbeitung II: Verkehrstheoretische Modelle von Echtzeitsystemen und Rechnernetzen*. Lehrstuhl Kommunikationsnetze, RWTH, Aachen, 1992.
- [Wal96a] B. Walke. *Kommunikationsnetze und Verkehrstheorie II: Verkehrstheoretische Modelle von Echtzeitsystemen und Kommunikationsnetzen*. Lehrstuhl Kommunikationsnetze, RWTH, Aachen, 1996.
- [Wal96b] B. Walke. *Mobilfunknetze und ihre Protokolle*. Lehrstuhl Kommunikationsnetze, RWTH, Aachen, 1996.
- [Wen91] C. Wendelin. *Objektorientierte Simulation mit Amadeus*. Dissertation, Sozial- und Wirtschaftswissenschaftliche Fakultät der Universität Wien, 1991.
- [Wie97] C. Wietfeld. *Mobilfunksysteme für die europäische Verkehrsleittechnik: – Leistungsanalyse des CEN-DSRC-Standards*. Dissertation, RWTH Aachen, Aachener Beiträge zur Mobil- und Telekommuni-

- kation, Berichte des Lehrstuhls Kommunikationsnetze, Verlag der Augustinusbuchhandlung, ABMT Band 12, 1997.
- [WK90] G. Willmann, P.J. Kühn. Performance Modeling of Signaling System No.7. *IEEE Commun. Magazine*, 28, 7:44–56, 1990.
- [WPS95] B. Walke, C. Plenge, M. Scheibenbogen. Technische Realisierbarkeit öffentlicher DECT-Anwendungen im Frequenzband 1880 -1900 MHz. Studie, Lehrstuhl für Kommunikationsnetze, RWTH Aachen, August 1995.
- [WS93] B. Walke, O. Spaniol, (Eds.). *Messung, Modellierung und Bewertung von Rechen- und Kommunikationssystemen*. Tagungsband 7. ITG/GI-Fachtagung, Informatik Aktuell, Springer Verlag, Berlin, September 1993.
- [WW93] M. Wippenbeck, G. Willmann. Some Recipes for the Statistical Analysis of Stochastic Simulation Output. *MMB'93: Messung, Modellierung und Bewertung von Rechen- und Kommunikationssystemen, Kurzberichte und Werkzeugvorstellungen*, 1993.
- [YD94] J. Yan, MacDonald D. Teletraffic Performance in Intelligent Network Services. In J. Labetoulle, J.W. Roberts, (Eds.), *The Fundamental Role of Teletraffic in the Evolution of Telecommunications Networks*, volume 1 a+b, 357–366. 14th International Teletraffic Conference, Antibes, North-Holland Elsevier Science B.V., 1994.
- [Zur64] R. Zurmühl. *Matrizen und ihre technischen Anwendungen*. Springer Verlag, Berlin, 1964.

NACHWORT

Die vorliegende Arbeit entstand am Lehrstuhl für Kommunikationsnetze der RWTH Aachen unter Leitung von Prof. Dr.-Ing. B. Walke. Ihm gilt mein besonderer Dank für die kontinuierliche Unterstützung meiner Arbeit, für seine wertvollen Anregungen und Diskussionen, sowie für die Möglichkeit zur Entfaltung selbständiger wissenschaftlicher Arbeit. Durch die Übertragung der Verantwortung bei der Planung und Durchführung von Projekten als Gruppenleiterin hatte ich die Möglichkeit in der Zusammenarbeit mit Industriepartnern zur praktischen Umsetzung meiner wissenschaftlichen Arbeit an aktuellen Anwendungsbeispielen.

Prof. (em) Dr.-Ing. F. Schreiber danke ich für die wissenschaftliche Basis, die er durch seine Arbeit gelegt hat, und die die RESTART/LRE-Untersuchungen zur Simulation seltener Ereignisse erst möglich gemacht haben. Die Zusammenarbeit mit ihm bei wissenschaftlichen Veröffentlichungen und bei der gemeinsamen Durchführung der Vorlesung *Stochastische Simulationstechnik* war für mich stets Vorbild und Antrieb.

Prof. Dr. F. Hoffeld und Prof. Dr.-Ing. P. Kühn (Universität Stuttgart) danke ich für die Bereitschaft und den Zeitaufwand als Gutachter dieser Arbeit sowie für die wertvollen Hinweise zum Manuskript.

Dem *Mobilise Projektteam* (Dr. M. Fröhlich, Dr.-Ing. M. Guntermann, Dr.-Ing. S. Kleier, Dipl.-Ing. R. Popp) danke ich für die anregenden und vorantreibenden Diskussionen und den Teamgeist, der die Zusammenarbeit so erfolgreich und gleichzeitig so erfreulich gemacht hat. M. Guntermann und S. Kleier sei auch für das Korrekturlesen einiger Kapitel dieser Arbeit gedankt.

Dr.-Ing. Christoph Herrmann danke ich für die Zusammenarbeit bei der Validierung der Ergebnisse dieser Arbeit, Dr.-Ing. M. Junius für die Kooperation bei der Konzeption und Implementierung der Simulationsbibliothek CNCL, Dr.-Ing. C. Plenge und Dr.-Ing. C. Wietfeld für die in dieser Arbeit verwendeten Beispiele zur Bewertung von Mobilfunknetzen.

Außerdem danke ich allen Diplomanden und Studienarbeitern, die meine Arbeit unterstützt haben. Namentlich möchte ich hier Dipl.-Ing. O. Fuß, der den RESTART/LRE-Simulator implementiert hat, und Dipl.-Ing. U. Vornefeld nennen, der Algorithmen zur Verbindungsannahmesteuerung in ATM-Netzen untersucht hat, und die meine Arbeit mit großer Sorgfalt Korrektur gelesen haben.

Der ITG-Fachgruppe 5.2.1 *Verkehrstheorie* unter Leitung von Prof. Dr.-Ing. P. Kühn danke ich für die Gelegenheit zu vielen anregenden Diskussionen bei Fachgruppentreffen und Workshops.

Ferner danke ich Prof. Dr. B. Müller-Clostermann (Universität GH Essen) für die Unterstützung bei der Validierung der Ergebnisse der Tandem-Netze und Prof. Dr. H. Swann (San Jose State University) für die Diskussionen und Anregungen zum Beweis der Optimalität der Stufenzahl des RESTART/LRE-Verfahrens.

Für die Unterstützung, insbesondere durch Bereitstellung und Pflege einer hervorragenden Infrastruktur, möchte ich mich bei allen Mitarbeitern des Lehrstuhls bedanken.

Aachen, im November 1997

Carmelita Görg

AACHENER BEITRÄGE ZUR MOBIL- UND TELEKOMMUNIKATION

ABMT Band 1

Herrmann, C.
Stochastische Modelle für
ATM-Konzepte,
1. Auflage 1995, 138 Seiten;
ISBN 3-86073-380-X

ABMT Band 2

Lawniczak, D. R.
Modellierung und Bewertung
der Datenverwaltungskonzepte
in UMTS,
1. Auflage 1995, 230 Seiten;
ISBN 3-86073-381-8

ABMT Band 3

Junius, M.
Leistungsbewertung intelligenter
Handover-Verfahren für
zellulare Mobilfunksysteme,
1. Auflage 1995, 208 Seiten;
ISBN 3-86073-382-6

ABMT Band 4

Steffan, H.
Stochastische Modelle für den
Funkkanal und deren Anwendung,
1. Auflage 1996, 164 Seiten;
ISBN 3-86073-383-4

ABMT Band 5

Böhmer, S.
Entwurf eines ATM-basierten
Funknetzes und Software-
Entwurfsmethodik zur Implementierung,
1. Auflage 1996, 172 Seiten;
ISBN 3-86073-384-2

ABMT Band 6

Guntermann, M.
Universelle Benutzermobilität
auf der Basis des Intelligenten
Netzes - Entwurf, Bewertung
und Implementierung -,
1. Auflage 1996, 164 Seiten;
ISBN 3-86073-385-0

ABMT Band 7

Kleier, S.
Neue Konzepte zur Unterstützung
von Mobilität in
Telekommunikationsnetzen,
1. Auflage 1996, 204 Seiten;
ISBN 3-86073-386-9

ABMT Band 8

Decker, P.
Entwurf und Leistungsbewertung
hybrider Fehlersicherungs-
protokolle für paketierte
Sprach- und Datendienste im
GSM-Mobilfunksystem,
1. Auflage 1997, 232 Seiten;
ISBN 3-86073-387-7

ABMT Band 9

Hußmann, H.
Algorithmen zur Kapazitäts-
optimierung schnurloser
Mobilfunksysteme nach DECT-
Standard,
1. Auflage 1997, 180 Seiten;
ISBN 3-86073-388-5

ABMT Band 10

Plenge, C.
Leistungsbewertung öffentlicher
DECT-Systeme,
1. Auflage 1997, 258 Seiten;
ISBN 3-86073-389-3

ABMT Band 11

Kennemann, O.
Lokalisierung von Mobilstationen
anhand ihrer Funkmeßdaten,
1. Auflage 1997, 162 Seiten;
ISBN 3-86073-620-5

ABMT Band 12

Wiefeld, C. M.
Mobilfunksysteme für die
europäische Verkehrsleittechnik -
Leistungsanalyse des
CEN-DSRC-Standards -,
1. Auflage 1997, 210 Seiten;
ISBN 3-86073-621-3

ABMT Band 13

Görg, C.
Verkehrstheoretische Modelle
und stochastische Simulationstechniken
zur Leistungsanalyse von
Kommunikationsnetzen,
1. Auflage 1997, 220 Seiten;
ISBN 3-86073-622-1

